Semantic representation for copyright metadata of user-generated content in folksonomies

Haklae Kim
Service Platform Group, Samsung Electronics, Suwon-city, Korea

John Breslin
Electrical and Electronic Engineering, School of Engineering and Informatics, National University of Ireland, Galway, Ireland, and

Jae Hwa Choi
College of Business, Dankook University, Choongnam, Korea

Abstract

Purpose – The purpose of this research is to investigate some general features of folksonomies and user-generated content with copyright issues, and to present semantic representation for folksonomies using a tag ontology that can be used to represent tagging data at a semantic level using Semantic Web technologies.

Design/methodology/approach – An exploratory study is described that features current social tagging methods and copyright metadata. In particular, a tag ontology is extended for representing copyright metadata across different platforms.

Findings – The main finding is that Social Semantic Cloud of Tags can improve the expressive knowledge representation of folksonomies and that this ontology can aid in describing copyright metadata using some extended properties.

Originality/value – The paper gives a valuable insight into representing folksonomies with Semantic Web technologies that enable the representation, exchange, and reuse of tagging data, and provides a way to reduce the risk of copyright infringements in the process of tag sharing in folksonomies.

Keywords Semantics, Worldwide web, Data handling, Intellectual property

Paper type Research paper

Introduction

Recently the exponential growth of user-generated content (UGC) and social software has been reshaping the landscape of the web. Content creation on the web was previously dominated by professional media services or domain experts. With social software such as blogs, wikis, and other content sharing services, non-experts can participate in content creation as well as sharing. In particular they have carried out tagging activities in a number of social software applications, and the aggregation of these activities creates emergent social knowledge, known as folksonomy (Mathes, 2004; Kroski, 2005). User-centred activities have led to a new phenomenon whereby people can communicate, connect and collaborate with each other via the web. The

term "Social Web" is often used to describe this phenomenon and includes relevant services and technologies (Gruber, 2008).

Most people participating in content creation and dissemination, however, are not trained as lawyers and may not consider the copyright and legal issues involved in user-generated content. Although O'Reilly (2005) argued that Web 2.0 is characterised by the principles of participation, sharing, and openness, it does not mean that this content has no intellectual property rights or is license-free. As a growing number of users participate in content sharing, the risk of infringements may increase. There have been several efforts to deal with copyright issues for UGC. The Digital Millennium Copyright Act (DMCA), which was passed in 1998, aimed to protect Internet Service Providers from liability for copyright infringement. The DMCA states that ISPs do not need to monitor their sites for infringing material. Instead it established a notice-and-take-down procedure (US Copyright Office, 1998). The DMCA, however, has been criticised for making it too easy for copyright owners to encourage web site owners to take down allegedly infringing content and links which may in fact not be infringing. In contrast the UGC Principles have different approaches to prevent infringement. While the Principles are not legally binding, they do set out meaningful obligations for web site companies to proactively block infringing content (Sawyer, 2009). In a slightly different way, there are some representational languages for expressing copyrighted information such as the Creative Commons Metadata Specification, Open Digital Rights Language, or RDF Site Summary 1.0 Modules, etc. These languages can describe copyrighted information of a certain resource with semantics.

Despite these efforts, existing copyright laws do not always resolve the copyright issues of UGC. UGC differs from traditional content not only in the manner in which it is created but also in the way rights to and liabilities arising from such content are allocated. Moreover we also need to take into account diverse aspects associated with UGC and copyright issues, since the content can be linked, embedded, copied, emailed, or mixed. In this paper we will focus on expressing copyrighted information in the process of tag sharing in folksonomies. Although copyright and intellectual property are related to content rather than folksonomies, tags in folksonomies are essentially assigned to a particular resource or content. It means that the aggregated tags will be shared with the associated content. It is the reason why copyright issues need to be considered in the process of tag sharing. Presently there is no uniform way to describe copyrighted information for this process. This paper aims to propose a semantic model for representing tagged social content with copyrights using Semantic Web technologies. The proposed model is extended by interlinking several Resource Description Framework (RDF) vocabularies. Since this model describes whole properties with appropriate semantics, the relationships among tagged content, users, and copyright for content can be explicitly specified.

The remainder of the paper is structured as follows. We start by providing an overview of social tagging and folksonomies, and introducing the Semantic Web. Then a brief overview of SCOT is presented. The next section describes the linking mechanisms between existing ontologies and SCOT, and the representation of copyright-protected content in SCOT. The following section discusses implementations in order to generate SCOT instances, and in the penultimate

section the experiments in terms of this implementation are presented. Finally we present some conclusions and explore future research directions.

Motivation

Various issues related to interoperability arise in the process of tag sharing. Current tagging systems do not allow the reuse of tags between different platforms (TagCommons Project, 2007); tagging activity via users' participation is locked into host sites and does not allow users on heterogeneous platforms to share meaningful information (Breslin and Decker, 2007). Thus the tagging behaviour of users cannot be navigated across different platforms with their tags. This may block the opportunity to connect people who may have common interests.

As illustrated in Figure 1, users' interests are represented not only by bookmarks or photos as a specific object in the sites, but also by tags as a common object. In this example if both users have common tags on Delicious and Flickr, they might have similar interests and easily integrated similar tagging behaviours. Through the tag sharing, they can share their resources, even though the resources are of different types and located on different service platforms. A set of tags (i.e. a tag cloud) acts as a social medium to mediate and interlink users across heterogeneous sites.

However, there is no way to reuse and transfer one's personal set of tags across platforms (Kim *et al.*, 2008a). Although RSS syndication and public APIs can be an alternative solution to publish and share data from Web 2.0 sites, the copyright information cannot be specified in the resources or is not described in an explicit manner. Thus creating new knowledge via tag sharing could expose users to potential copyright infringement. In addition legal issues regarding content arise throughout the process of sharing tag metadata, although these have attracted little attention in both the academic and business worlds to date. The copyrights can be separated from the original content in the process of tag sharing. To solve those problems we need to conceptualise tagging activities and to provide explicit methods for describing the conceptualisation. Moreover a consistent way of exposing and accessing copyrighted
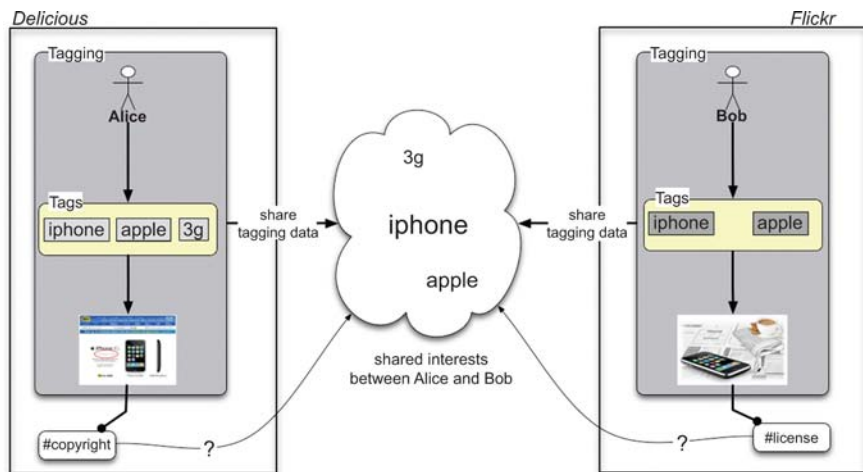

Figure 1.
Tag-sharing process across independent, heterogeneous tagging platforms

resources is required. Alternative technologies such as the Semantic Web may help in solving these problems.

Related work

UGC and intellectual property rights

UGC, also known as user-created content or consumer-generated media, refers to media sources such as blog entries, video clips, music, or photos that are publicly available and produced by individual users (Biederman and Andrews, 2008). Yet "user-generated" does not mean that the content was actually created by the current user. A majority of UGC service providers, which rely solely on their users for content, allow users to easily upload and share content regardless of whether or not they were the ones to produce it. UGC service providers often allow everyday consumers/users to access their data through official application programming interfaces (APIs) and content syndication. These technologies may enhance the quality of content by bringing it to the public in easy and inexpensive ways. Once content is uploaded to UGC service providers, it is readily available to the public. The content consisting of a combination of data from different sources is republished on the web by active user participation. This encourages and facilitates third parties to use the data (Blezquez, 2008).

However, the collaborative nature of Web 2.0 is likely to expose users to risks of intellectual property violations, regardless of their intention. Intellectual property covers any form of knowledge or expression and comprises a range of legal rights for things created via human intellect. Intellectual property law consists of several separate and overlapping legal disciplines, each with their own characteristics and terminology as illustrated in Table I.

Most UGC service sites tend to state clearly within their terms of conditions or services who owns copyright and who is responsible for dealing with infringements. However, mixing, republishing, and distributing content may lead to blurring the copyright issues, because users may not know the specific conditions regarding mixed content. One cannot control and organise the overall process in a consistent way, since this indirect collaboration is often decentralised and distributed. Moreover, there is no explicit way to represent completed copyright infringements of service providers. As the users tend to participate in the process of sharing and exchanging their own or other copyright protected content, they will inevitably encounter legal problems.

There are common languages for digital copyright representation in the open and global framework such as the Creative Commons Metadata Specification (Abelson *et al.*, 2008), RDF Site Summary 1.0 Modules (Hammersley, 2002), MPEG-21 REL (ISO, 2004), Open Digital Rights Language (ODRL) (Iannella, 2002), IPROnto (Delgado *et al.*, 2003) and Copyright Ontology (Garcia *et al.*, 2007). The mod_cc of RDF 1.0 modules provided by Creative Commons allows describing metadata in terms of the copyright license of RSS feeds. Creative Commons provide copyright licenses for creative works, with a focus on the development of innovative and balanced approaches to copyright and intellectual property law. Instead of the usual "all rights reserved" attributed to a copyright, users can choose to reserve some customised rights. The CC license describes works, licenses, and license features including permissions, prohibitions, and requirements in machine-readable RDF/XML. However CCL is not sufficient to cover a number of problems occurring in UGC (Lee *et al.*, 2008). IPROnto aims to offer

630

	Copyright	Patent	Trade mark
What does it protect?	Original works in fixed, perceptible form	Products, processes	Business identities: words, symbols, devices, logos or slogans
Symbols	©	Types of patents: utility, plant and design "Patented . . ."	™ ®
Tests	Original	Unique, novel, non-obvious	Does not duplicate existing mark
How to obtain?	Publication (registration)	Examination by patent	Registration with PTO, common law
How long does it last?	Life of author + 70 years for works created by a single individual	Generally 20 years from date of filing application	Must be renewed every ten years

Table I.
Comparison of copyright, patent, and trade mark

interoperability of IPR frameworks, integrating both the Rights Expression Language (Wang *et al.*, 2005) and the Rights Data Dictionary (Garcia *et al.*, 2005). Since Web Ontology Language (OWL) represents this model, it can be modulated to other web ontologies. These languages aim to provide interoperability between digital copyright systems and to build "a complete framework for representing copyright value chains and the associated flow of rights" (Garcia *et al.*, 2007, p. 9).

Tagging on the Semantic Web
A key feature of user-contributed content in Web 2.0 sites is that the content may be tagged and can be shared by others. These sites allow users to upload content items, submit comments, and categorise them using topics or tags. The overall task of tagging is a simpler process tapping into an existing cognitive process without adding much cognitive cost (Mathes, 2004). Tagging does not aim to create a rigid classification of objects, but to categorise an object according to users' interests with their own keywords (Golder and Hubermann, 2006). Although a few words cannot identify all users' interests, a culture of mass participation leads to social interaction among users and influences the use of terms in a community (Kroski, 2005). This is how social tagging works. People can in general use any term as a tag without exactly understanding the meaning of the terms they choose. The tags produced by a number of users can be aggregated to form a non-hierarchical taxonomy – the folksonomy created by Van der Wal (2005). For example tagging systems or folksonomies, such as Flickr or Delicious, enable a number of users to categorise information sources in an unstructured way and to visually represent popular tag usages via the use of "tag clouds". A tag is not just a keyword, but also acts like a subject or category for the associated content (Mathes, 2004). Tagging is not only a common feature of social content, but also an important way of mediating common interests across independent, heterogeneous sources. In this perspective, tagging can be a common tool for sharing, exchanging, and integrating users' interests within social objects from various Web 2.0 sites. Tag sharing is an alternative method for creating new knowledge from heterogeneous platforms.

The ambiguity of tags, however, is one of the problems inherent in an uncontrolled vocabulary (Rosati *et al.*, 2006) and the lack of synonym control can lead to different tags being used for the same concept (Mathes, 2004; Quintarelli, 2005). In contrast to these inherent problems, tagging systems allow users a restricted functionality for reusing, sharing, and discovering tags (TagCommons Project, 2007). These limitations are due to the inability to express tag structures and relationships, and the lack of an explicit semantics of tags. Data mining, knowledge representation, or natural language processing techniques can be used to overcome these disadvantages. Moreover folksonomies on the Semantic Web are represented by a standard format to structure tagging data with precise semantics. The Semantic Web, a web of data, enables the web to understand the requests of people and machines to use the web content semantically and automatically (Berners-Lee *et al.*, 2001). It aims to provide a common framework that allows data to be shared and reused across applications, enterprises, and communities (Berners-Lee, 2006b). Semantic Web-based approaches can support a standardised metadata schema to represent both structures and semantics of tagging data. Furthermore this kind of schema does not just define certain tags, but can also robustly represent the relationships among the entities that

shape tagging activities, explicitly stating the knowledge structure of tagging data (Gruber, 2008).

SCOT: a semantic model for folksonomy

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 relationships among users (Kim *et al.*, 2008b). Thus tagging on the Semantic Web has to represent overall features of tagging entities in a given community or site, and simultaneously to allow for a continuous transformation from individual to social tagging with appropriate semantics. In general, users may have a number of tagging activities with arbitrary relationships between them. As the users continue their activities, the relationships between tagging entities should then be updated. The proposed model aims to allow for expressivity for collective features of social tagging and to support social interoperability for sharing and reusing tagging data across different platforms, users, or sources. Gruber (2007) and Spivack (2008) emphasise the need for folksonomies and ontologies to work together, aiming to identify and formalise a conceptualisation of tagging data at a semantic level. Typical social tagging systems do not provide explicit links amongst the entities that are made up of tagging activities, nor expose their data in a standard form. Tag ontology aims to provide a common conceptualisation of what tagging means through a standardised way to collect, interpret, or use tagging data (TagCommons Project, 2007). One of the advantages of this ontology is that isolated tagging data can be easily made mobile and integrated across applications. Tags, a user, and their relations in a particular application can be represented in a form of ontology such as RDF or OWL, and these data can be accessible and movable on the web as linked data (Berners-Lee, 2006a; Bergman, 2008). This can be considered as a starting point of sharing, exchanging for separate tagging activities on different platforms.

There is agreement on what the most elemental building blocks of a tag model should be (TagCommons Project, 2007; Kim *et al.*, 2008b,c). The most popular model is the tripartite model: Tagging (U, R, T). U refers to the set of taggers who participate in the tagging activity, T identifies the set of tags that is assigned in resources, and R depicts the set of items. Moreover for T to be "social" there has to be some kind of social interaction between the users in U. The ternary relationship Y (i.e. $Y \in U \times R \times T$) among the entities is defined when tagging has occurred and the results have been shared. Gruber discussed an ontology for folksonomy that is for the infrastructure to build an ecosystem of tag data sources, services, agents, and tools (Gruber, 2007). In his model the core concept is tagging, that is the act of associating tags with an object or item (Gruber, 2005, 2007). Newman's model (Newman *et al.*, 2005) describes relationships between an agent, an arbitrary resource, and one or more tags. In this model there are three core concepts – tagger, tagging, and tag – to represent a tagging activity; the concepts are serialised in RDF/OWL. Existing tag ontologies including Gruber's (2007) and Newman's model do not provide a way of fully representing the meaning of a tag and the relationships between each tag, since they focus on expressing individual tagging instances.

SCOT (http://scot-project.org) is an acronym for Social Semantic Cloud of Tags. The name was chosen to emphasise the goal of providing a consistent framework for expressing tagging activities in machine-understandable way (Kim *et al.*, 2009). 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. These features are a cornerstone to being able to identify, formalise, and interoperate a common conceptualisation of tagging activity at a semantic level. SCOT offers various properties for representing tag semantics and collective characteristics of tagging entities (Kim *et al.*, 2008c). Both Tagcloud and Tag class play a key role in the representation of social tagging. scot:TagCloud has properties that describe a certain user, tag spaces, number of tags, posts and co-occurrences and their frequencies, as well as updated information. scot:Tag, as a subclass of tag:Tag from Newman's model, describes a tag that is aggregated from individual tagging activities. The property scot:contains links scot:TagCloud to a set of scot:Tag instances. It is important to note that SCOT uses concepts and properties from Newman's model. As shown in Figure 2 the Tagging class represents tags themselves (tag:associatedTag), the resources that are being tagged (tag:taggedResource), and the users that create these tags (tag:taggedBy). The scot:TagCloud class connects tag:Tagging instances via the property scot:taggingActivity. The property scot:taggingAccount represents an account of users in online services. Individual tags in tag:Tagging are mapped to a resource with scot:Tag instance and then these tags are represented by a collection of tags underlying a scot:TagCloud. The instances of scot:Tag are linked to individual tags defined in tag:Tag, using the property scot:aggregatedTag.

Tags can be used with many different conventions in the real world. Suppose a tag "iPhone", "IPHONE" in upper case, "iphone" in lower case, or "i-phone" in a compound term. All terms have the same meaning, if not the same intended purpose. The Tag class introduces some properties for describing these features. For example, scot:spellingVariant represents a variation in the way in which a word is spelt, scot:delimited describes a multiple-word tag name where each word is separated by a certain character, and scot:synonym describes a term which means the same as another word. These properties can reduce tag ambiguity from different conventions and even recommend more common patterns of tag name. Furthermore, in order to represent
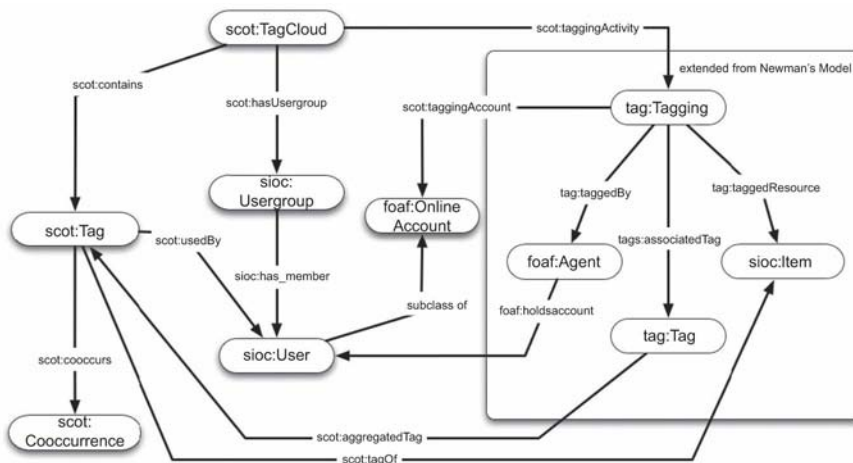
Figure 2.
Simplified folksonomy
model in SCOT

both formats of tag frequencies, SCOT introduces two properties: scot:ownAFrequency and scot:ownRFrequency. The former is intended to describe the absolute format of popularity for a specific tag and the purpose of the latter is to represent the relative format to identify the significance of the tag proportional to total tags. A single tag can have both frequency formats. The popularity of the tag plays a key role in distinguishing its significance in folksonomies.

Linking to existing ontologies

SCOT aims to incorporate and reuse existing RDF vocabularies as much as possible in order to avoid redundancies and to enable the use of richer metadata descriptions for specific domains (Kim *et al.*, 2009). We will now describe relations to other vocabularies, including a way to represent copyrighted information.

RDF vocabularies

Dublin Core (DC) provides a basic set of properties and types for annotating documents (Weibel, 1998). In SCOT we use the properties dc:title for the title of a TagCloud, dc:description to give a summary of the TagCloud, dc:publisher to define what system is generating the TagCloud, and dc:creator to link to the person who created this set of tags. dcterms:created, from the Dublin Core refinements vocabulary, is used to define when a TagCloud was first created. Friend of a Friend (FOAF) specifies the most important features related to people acting in online communities (Graves *et al.*, 2007). The vocabulary allows us to specify properties about people commonly appearing on personal homepages, and to describe links between people who know each other. foaf:Person is used to define the creator of a particular TagCloud. foaf:Group can be used to define a group of people who have created a group TagCloud. SIOC (Semantically-Interlinked Online Communities) provides the main concepts and properties required to describe information from online communities (e.g. message boards, wikis, blogs, etc.) on the Semantic Web (Bojars *et al.*, 2008). In the context of SCOT, sioc:Usergroup can be used to represent a set of users who have created the tags contained within a particular group TagCloud. A TagCloud is also a type of sioc:Container, in that it contains a set of Tags (subclass of sioc:Item). Simple Knowledge Organisation Systems (SKOS) provides specifications and standards to support the use of knowledge organisation systems such as thesauri, classification schemes, subject heading lists, taxonomies, and other types of controlled vocabulary, as well as terminologies and glossaries (Miles and Bechhofer, 2008). Tag is a subclass of skos:Concept, and a number of SKOS properties are used to define the relationships between Tags: broader, narrower, etc.

Intellectual property rights

Current research efforts have been directed toward the introduction of the scot:hasCopyright and scot:hasIPR properties rather than the task of establishing new terms in SCOT. The scot:hasCopyright property identifies copyrights, access restrictions, manipulation and republishing of a given resource. The property scot:hasIPR identifies and groups into a set of copyright such as DRM (Digital Rights Management), terms and conditions, services, usage restrictions, etc. More importantly both properties are designed to cater for the inclusion of any existing and/or future copyrights that contain the URI of the license applied, through

rdfs:Resource range value for their properties. If digital copyright is associated with certain content it could be easily linked to using a URI mechanism, e.g. copyright-protected content can be pointed to an associated copyright using the scot:hasCopyright property. When copyright is described by an ontology, it is possible to link a specific item, whereas copyright such as "terms of conditions" may be linked to an associated URL, which includes the detailed description. Copyright for a set of resources such as that within a TagCloud can be described with the scot:hasIPR property. As illustrated in Figure 3 Alice and Bob share their tagging data that they assign the tag "beatles" and "yesterday" into the resources (i.e. Music#id, Video#id, and Photo#id, respectively) with each resource having specific linked copyright represented by URI (#DRM, #CC, and #conditions, respectively). When it comes to content mixing or republishing the resources, the copyright will be explicitly associated with the tagging entities.

Creating and consuming semantic tag metadata

The majority of social sites now offer their APIs based on most popular formats (e.g. REST, SOAP and XML RPC). These APIs allow community users and applications easy and intuitive access to data associated with the sites. The SCOT Exporters allow for extracting tagging data from multiple sources and for generating semantic tag metadata with the full expressivity of SCOT. The export tools are available for download at http://scot-project.org/applications/. The WordPress Exporter, which is a type of plug-in, creates SCOT RDF/XML from an individual blog. Once this plug-in is activated in the WordPress administration panel, the instance metadata is created in the location http://yourhost/scot.rdf, and the metadata are updated when any information from the blog is changed. This generates semantic tag metadata from legacy tagging data without users' additional efforts. The Exporter for relational databases aims to create SCOT instances from a large number of RSS feeds stored in a relational database.

Both exporters rely upon the same algorithms (Kim *et al.*, 2007) to generate SCOT metadata, which may in turn be used as the basis for creating further exporters. A URI
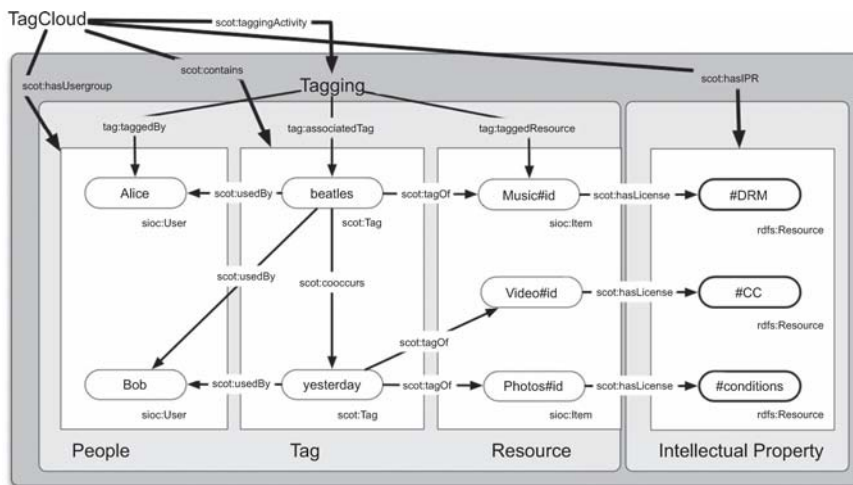
Figure 3.
Representing tagging activities and intellectual property regarding the resources in SCOT

for tagging entities is defined for compliance with the Linked Data principles (Berners-Lee, 2006a; Bizer *et al.*, 2007), enabling SCOT included data to be used for community content navigation. Figure 4 shows the result of using the WordPress Exporter in Tabulator (Berners-Lee *et al.*, 2007) which is a generic data browser and editor.

The Exporter, however, just provides a simple method for exposing SCOT metadata rather than supporting a method for managing and retrieving the metadata. The int.ere.st web site (http://int.ere.st) aims at publishing the open Semantic Web database for tagging data, including a large number of interlinks to several data sets on tagging applications (Kim *et al.*, 2008a). Tagging data of individual users, communities, or corporations existing in distributed environments can be imported to int.ere.st via mash-up services, then the data can be transformed into SCOT. 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 application across social tagging sites. In order to allow users and developers to support the social capabilities underlying tagging data, the platforms also provide some open APIs. All information in int.ere.st is published as linked data using the D2R Server (Bizer and Cyganiak, 2006), a tool that maps relational databases to RDF and is accessible through SPARQL. SCOT facilitates the linking of tagging entities and information queries can be made using SPARQL. int.ere.st implements a similar approach by using a HTTP content negotiation mechanism to provide users with either HTML or RDF data representation and a SPARQL (Seaborne and Prud'hommeaux, 2008) interface that allows users to query semantic tag metadata directly.

Figure 4.
The WordPress Exporter and SCOT metadata instance in Tabulator

Conclusion
This paper proposed an interlinked method for representing copyrighted information explicitly in the process of resource sharing. As more people participate in tagging activities on social web sites, there are growing demands to reuse or exchange a variety of tagged resources including videos, bookmarks, and photos. There are two main issues in terms of sharing tagged resources: a consistent representation of folksonomies and an explicit method for describing copyrights. The representation of folksonomies is comparatively limited, and tagging practices cannot currently be described in an explicit structure and are not easy to re-use and update. Representing copyrights for social content is very complicated, since the content can be easily shared and disseminated across different web sites. Although there have been some efforts to represent copyrights or intellectual property rights, there is little consideration in terms of tagged resources. These limitations in terms of representation can be corrected via Semantic Web technologies, by providing more specific ontological terms to represent tagged resources, including people, resources (social content), and their relationships.

We presented the SCOT ontology which begins to address these limitations in relation to the domain of tagged social content, the lack of an expressive format for describing the structure, tag semantics and the lack of tag exchange. The SCOT ontology provides not only a standardised format to share, exchange, and reuse tag data among users or communities, but it can also be interlinked with other ontologies, which aim to describe copyrights or intellectual property rights. In particular we proposed a way to interlink between SCOT and copyright-protected content, by extending a number of advanced tagging properties (e.g. scot:hasLicense). A SCOT instance includes a resource and its relevant copyright information as a URI. Thus users can recognise a relevant copyright for a certain resource when they use, share, or reproduce it. In decentralised and distributed environments, this approach can be an effective way to keep copyright information for relevant content in the process of content sharing.

The major drawbacks of folksonomies, however, are their lack of semantics and keyword ambiguity. A tag can be represented by a number of variations such as capitalisation (e.g. "Apple" and "apple"), singular vs. plural (e.g. "blog" and "blogs"), or delimited words (e.g. "iPhone" and "i-phone"), while the relationships between tags are not expressed in an explicit manner. Some ontologies have been created for the purpose of leveraging tagging data to a semantic level. For example MOAT (Meaning Of A Tag) aims to represent the relationships between tags and their meanings, using the URIs of existing Semantic Web vocabularies. By combining both ontologies, there is the possibility of not only assigning meanings to tags, but of reusing ontological concepts and instances of these concepts directly as tags. These "tags" can be considered "semantic tags" since they will have a clear pre-defined meaning as defined in ontologies.

A considerable amount of the available social content has been made and shared by end-users on the Social Web. From a social network perspective, folksonomies can be an alternative way to create new connections among people who participate in tagging activities from heterogeneous platforms. There may be a relationship among individuals based on users' behaviours in different online communities. For example if individuals use similar sets of tags in different web sites, they can be considered as a social network based on an object: people may be connected when they use a similar set

of tags. Once all this information is exposed by RDF vocabularies such as FOAF, intelligent technology can be employed to constantly elicit new knowledge by observing this network. Creating new knowledge from diverse social objects remains a big challenge in terms of interoperation during the process of data sharing. Thus this approach would benefit the online community with the provision of a seamless framework for interlinking social content across different applications and their automatic access.

References

Abelson, H., Adida, B., Linksvayer, M. and Yergler, N. (2008), "ccREL: the Creative Commons Rights Expression Language", available at: http://wiki.creativecommons.org/images/d/d6/Ccrel-1.0.pdf (accessed 23 February 2010).

Bergman, M. (2008), "What is linked data?", available at: www.mkbergman.com/?p=447 (accessed 23 February 2010).

Berners-Lee, T. (2006a), "Linked data, world wide web design issues", available at: www.w3.org/DesignIssues/LinkedData.html (accessed 23 February 2010).

Berners-Lee, T. (2006b), "Semantic Web roadmap, world wide web design", available at: www.w3.org/DesignIssues/Semantic.html (accessed 23 February 2010).

Berners-Lee, T., Hendler, J. and Lassila, O. (2001), "The Semantic Web", *Scientific American*, Vol. 5 No. 284, pp. 34-43, available at: www.sciam.com/article.cfm?id=the-semantic-web (accessed 23 February 2010).

Berners-Lee, T., Hollenbach, J., Lu, K., Presbrey, J., Prud'hommeaux, E. and Schraefel, M. (2007), "Tabulator redux: writing into the Semantic Web", available at: http://eprints.ecs.soton.ac.uk/14773/ (accessed 23 February 2010).

Biederman, C.J. and Andrews, D. (2008), "Applying copyright law to user-generated content", *Los Angeles Lawyer*, May, pp. 12-18, available at: www.lacba.org/Files/LAL/Vol31No3/2483.pdf (accessed 3 May 2010).

Bizer, C. and Cyganiak, R. (2006), "D2R Server: publishing relational databases on the semantic web", paper presented at the 5th International Semantic Web Conference, Athens, GA, 5-9 November.

Bizer, C., Cyganiak, R. and Heath, T. (2007), "How to publish linked data on the web", available at: www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial (accessed 23 February 2010).

Blezquez, F.J. (2008), "User-generated content services and copyright", *IRIS Plus*, May, pp. 408-17, available at: www.obs.coe.int/oea_publ/iris/iris_plus/iplus5_2008.pdf.en (accessed 3 May 2010).

Bojars, U., Breslin, J.G., Peristeras, V., Tummarello, G. and Decker, S. (2008), "Interlinking the Social Web with semantics", *IEEE Intelligent Systems*, Vol. 23 No. 3, pp. 29-40.

Breslin, J.G. and Decker, S. (2007), "The future of social networks on the internet: the need for semantics", *IEEE Internet Computing*, Vol. 11 No. 6, pp. 86-90.

Delgado, J., Gallego, I., Llorente, S. and García, R. (2003), "IPROnto: an ontology for digital rights management", *16th Annual Conference on Legal Knowledge and Information Systems, JURIX 2003, Frontiers in Artificial Intelligence and Applications*, Vol. 106, IOS Press, Amsterdam, pp. 111-20.

Garcia, R., Delgado, J. and Rodriguez, E. (2005), "An ontological approach for the management of rights data dictionaries", in Moens, M. and Spyns, P. (Eds), *Proceedings of the 2005 Conference on Legal Knowledge and Information Systems: Frontiers in Artificial Intelligence and Applications*, Vol. 134, IOS Press, Amsterdam, pp. 137-46.

Garcia, R., Gil, R. and Delgado, J. (2007), "A web ontologies framework for digital rights management", *Artificial Intelligence Law*, Vol. 15 No. 2, pp. 137-54.

Golder, S. and Hubermann, B.A. (2006), "The structure of collaborative tagging systems", *Journal of Information Sciences*, Vol. 32 No. 2, pp. 198-208.

Graves, M., Constabaris, A. and Brickley, D. (2007), "FOAF: connecting people on the Semantic Web", *Cataloging Classification Quarterly*, Vol. 43 No. 12, pp. 191-202.

Gruber, T. (2005), "TagOntology – a way to agree on the semantics of tagging data", presentation to Tag Camp, available at: http://tomgruber.org/writing/tagontology-tagcamp-talk.pdf (accessed 23 February 2010).

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.

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.

Hammersley, B. (2002), "RDF site summary 1.0 modules: mod_cc", available at: http://web.resource.org/rss/1.0/modules/cc/ (accessed 23 February 2010).

Iannella, R. (2002), "Open Digital Rights Language (ODRL) Version 1.1", available at: www.w3.org/TR/odrl/ (accessed 23 February 2010).

ISO (2004), "Multimedia framework (MPEG-21) – part 5: rights expression language", technical report, ISO/IEC TR 21000-5:2004, available at: http://mpeg.chiariglione.org/standards/mpeg-21/mpeg-21.htm#_Toc23297977 (accessed 3 May 2010).

Kim, H.L., Yang, S.K., Breslin, J.G. and Kim, H.G. (2007), "Simple algorithms for representing tag frequencies in the SCOT exporter", *IAT '07: Proceedings of the 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, IEEE Computer Society, Washington, DC, pp. 536-9.

Kim, H.L., Yang, S.K., Breslin, J.G. and Kim, H.G. (2008a), "int.ere.st: building a tag sharing service with the SCOT ontology", paper presented at American Association for Artificial Intelligence Spring Symposium on Social Information Processing, Stanford University, CA, 26-28 March.

Kim, H.L., Passant, A., Breslin, J.G., Scerri, S. and Decker, S. (2008b), "Review and alignment of tag ontologies for semantically-linked data in collaborative tagging spaces", *Proceedings of the 2nd IEEE International Conference on Semantic Computing*, IEEE Computer Society, Washington, DC, pp. 315-22.

Kim, H.L., Scerri, S., Breslin, J.G., Decker, S. and Kim, H.G. (2008c), "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*, Dublin Core Metadata Initiative, Berlin, pp. 128-37.

Kim, H.L., Yang, S.K., Breslin, J.G., Scerri, S., Decker, S., Yang, S.K. and Kim, H.G. (2009), "SCOT ontology specification", available at: http://scot-project.org/scot/index.html (accessed 23 February 2010).

Kroski, E. (2005), "The hive mind: folksonomies and user-based tagging", available at: http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/ (accessed 23 February 2010).

Lee, S.B., Choi, H.J. and Sim, S.G. (2008), "A multiple authorship model in user-generated content", *Proceedings of the 2008 IEEE Consumer Communications and Networking Conference*, IEEE Computer Society, Washington, DC, pp. 1070-4.

Mathes, A. (2004), "Folksonomies – cooperative classification and communication through shared metadata", available at: www.adammathes.com/academic/computer-mediated-communication/folksonomies.html (accessed 23 February 2010).

Miles, A. and Bechhofer, S. (2008), "SKOS simple knowledge organization system reference", available at: www.w3.org/TR/skos-reference (accessed 23 February 2010).

Newman, R., Ayers, D. and Russell, S. (2005), "Tag ontology", available at: www.holygoat.co.uk/owl/redwood/0.1/tags/ (accessed 23 February 2010).

O'Reilly, T. (2005), "What is Web 2.0? Design patterns and business models for the next generation of software", available at: www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html (accessed 23 February 2010).

Quintarelli, E. (2005), "Folksonomies: power to the people", available at: www.iskoi.org/doc/folksonomies.htm (accessed 23 February 2010).

Rosati, L., Resmini, A. and Quintarelli, E. (2006), "FaceTag: integrating bottom-up and top-down classification", available at: www.facetag.org/download/facetag.pdf (accessed 23 February 2010).

Sawyer, M. (2009), "Filters, fair use, and feedback: user-generated content principles and the DMCA", *Berkley Technology Law Journal*, forthcoming, available at http://ssrn.com/abstract=1369665 (accessed 23 February 2010).

Seaborne, A. and Prud'hommeaux, E. (2008), "SPARQL Query Language for RDF", available at: www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/ (accessed 23 February 2010).

Spivack, N. (2008), "Folktologies – beyond the folksonomy vs ontology distinction", available at: http://novaspivack.typepad.com/nova_spivacks_weblog/2005/01/whats_after_fol.html (accessed 24 September 2008).

TagCommons Project (2007), "Functional requirements for sharing tag data", available at: http://tagcommons.org/2007/02/28/functional-requirements-for-sharing-tag-data/ (accessed 23 February 2010).

US Copyright Office (1998), *The Digital Millennium Copyright Act of 1998*, Pub. L. No. 105-304, 112 Stat. 2860, US Copyright Office, Washington, DC.

Van der Wal, T. (2005), "Folksonomy definition and Wikipedia", available at: www.vanderwal.net/random/entrysel.php?blog=1750 (accessed 23 February 2010).

Wang, X., DeMartini, T., Wragg, B., Paramasivam, M. and Barlas, C. (2005), "The MPEG-21 rights expression language and rights data dictionary", *IEEE Transactions on Multimedia*, Vol. 7 No. 3, pp. 408-17.

Weibel, S. (1998), "The Dublin Core: a simple content description model for electronic resources", *Bulletin of the American Society for Information Science and Technology*, Vol. 24 No. 1, pp. 9-11.

About the authors
Haklae Kim is currently a Senior Engineer in the Media Solution Centre at Samsung Electronics. He is also a director of the Web Science Workgroup in Korea. He was a researcher at the Digital Enterprise Research Institute at National University of Ireland, Galway, and his research efforts focused on creating social spaces based on tagging activities. He is the founder of the SCOT (Social Semantic Cloud of Tags) ontology project, which aims to provide semantic relations for tagging data, and has implemented the int.ere.st web site for sharing tagging data. Haklae Kim received a BS in Business Administration, and an MS and PhD in Management Information Systems from Dankook University, as well as a PhD in Computer Science from NUI Galway. Haklae Kim is the corresponding author and can be contacted on: haklae.kim@gmail.com

John Breslin is currently a Lecturer in Electrical and Electronic Engineering at NUI Galway. He is also an Associate Researcher and Leader of the Social Software Unit at the Digital Enterprise Research Institute at NUI Galway, researching semantically-enabled social networks and community portals. He is the founder of the SIOC project. He was previously a postdoctoral researcher at DERI from 2004 to 2008, a lecturer at the University from 2000 to 2004, and was a research officer and postgraduate student at the Power Electronics Research Centre in NUI Galway before that. He also worked as a visiting scholar at Virginia Tech in 1996. He has received a number of awards for web site design, including a Golden Spider for the Irish community websiteboards.ie, which he co-founded in 2000. The Irish Internet Association presented him with Net Visionary awards in 2005 and 2006.

Jae-Hwa Choi is a Professor of Information Systems in the Department of Management at Dankook University in the Republic of Korea. He received his PhD in Information Systems from the University of Maryland in College Park. Before joining Dankook University he was an Assistant Professor at the University of Michigan-Dearborn. His recent research interests include collaborative innovation and collective intelligence in knowledge management and business intelligence and decision support systems.