

# Social Semantic Cloud of Tag: Semantic Model for Social Tagging

Hak-Lae Kim<sup>1</sup>, John G. Breslin<sup>1</sup>, Sung-Kwon Yang<sup>2</sup>, and Hong-Gee Kim<sup>2</sup>

<sup>1</sup> Digital Enterprise Research Institute, National University of Ireland, Galway, IDA Business Park, Lower Dangan, Galway, Ireland  
haklae.kim@deri.org, john.breslin@deri.org

<sup>2</sup> Biomedical Knowledge Engineering Lab, Seoul National University, 28-22 Yeonkun-Dong, Chongno-Ku, Seoul 110-749, Korea  
sungkwon.yang, hgkim@snu.ac.kr

**Abstract.** Tagging has proven to be a successful and efficient way for creating metadata through a human collective intelligence. It can be considered not only an application of individuals for expressing one's interests, but also as a starting point for leveraging social connections through collaborative user participations. A number of users have contributed to tag resources in web sites such as Del.icio.us, Flickr etc.

However, there is no uniform structure to describe tags and user's activities. This makes difficult to share and represent tag data among people. The SCOT (Social Semantic Cloud of Tags) ontology is aimed to represent the structure and semantics of a set of tags and promotes their global sharing. The paper introduce the SCOT ontology and methods of its representation.

## 1 Introduction

A number of social bookmarking and tagging sites have become popular, and tagging in traditional web sites is getting adopted at a good pace. Many people already know how to bookmark and tag online resources such as Web sites, bookmarks, photos, and blog posts. Tagging is a way for representing concepts by cognitive association techniques, but does not force us to categorize. Each tag tells us about what we are interested in and improves social reinforcement through enabling social connections and search. There is an advantage that social bookmarking and tagging is a simple way that allows a user to save and share anything in online communities.

But the critical problem is that the social bookmarking and tagging systems do not provide a uniform way to share and reuse tag data among users or communities. Although some systems support an export functionality using open APIs, there are no uniform structure and semantics to represent tag data. Therefore, it is not easy to meaningfully search, compare or merge “*similar collective tagging data*” [2] on different sources.

This paper will explore the idea of defining the structure and the semantics of social tagging through a new approach we have called SCOT (Social Semantic

Cloud of Tags)[3]. This approach can help to represent the context required in the use of social tagging and to provide methods for sharing and reusing tagged data.

To remedy the shortcoming of social tagging, we advocate using Semantic Web technologies semantically to represent tagging data. In particular, the contributions of the paper are as follows:

- We suggest the semantic model to represent the structure and semantics of tagging data on social tagging spaces.
- The SCOT can enhance sharing and reuse of social tagging data.

The next section will discuss the motivation for presented ideas and address a number of issues for social tagging. In Section 3 the design principles of the SCOT ontology will be introduced. Section 4 will be devoted to the core concepts of the SCOT ontology to describe the structure and the semantics of tag data. Section 5 explains how a SCOT space expands. Section 6 describes some works related to the topic. Finally, future directions for further development of the SCOT ontology will be discussed.

## 2 Motivations

We propose the SCOT ontology to address a number of issues of note in the domain of tagged social content.

Firstly, people have a certain amount of inherent laziness when it comes to tagging content. Over time, they may build up a rich collection of tags on a certain site or application, but the size and scope of this collection will depend on how regularly the site or application is being used. When beginning to tag content on a new system, the user may have a certain amount of reluctance to create new tags if they do not expect to be using the system that much, and also they may not wish to recreate their existing tag information on other sites.

Related to this issue is a problem with the lack of reuse of tags between an individual's various applications. At the moment, there is no consistent method for reusing one's personal set of tags between either web-based systems, desktop applications, or for transferring tags between the desktop and the Web. For example, in Windows Vista users can now tag files and use those keywords to organise and search all documents on their desktop. However, they cannot reuse this collection of tags from their desktop when they signup and start annotating content on a new social bookmarking website.

The third issue is in relation to a consistent use of tagging. Even a single user may forget the tags that they have used previously on a single site (not to mention all the sites or applications that they may use) when they are prompted with an empty tag field for annotating a new content item. Some systems may provide tag recommendations based on previously used tags (or from a user's social network), but this could be augmented by suggesting tags from any system that a user has previously tagged content on. In order to provide a consistent and wider view of all the content that a person has tagged with a particular keyword through multiple applications, such cross-platform exchanges of tag cloud metadata is necessary.

Another motivation is to allow tag reuse between people in existing social networks, and to connect people who may have a common interest of set of interests. On signing up for a site, one may quickly realise that a particular tagging scheme used by a network contact is of interest when tagging one's own content. Also, social networks of people may be tagging content across various sites, and therefore a tag cloud for not only one's own content but also for one's friends' can be useful for maintaining a cohesive social network across different systems. Even if there is no existing link between people with similar interests, these can be formed through a serendipitous use of similar tags or by browsing from one's community tag cloud to others who use tags in that cloud.

Finally, another problem that has been noticed with the exporting of tag information from social applications is the lack of explicit information on tag structure, how often tags occur together, and who has used a particular set of tags. We aim to address this lack of semantics of tag structure with a number of interesting properties in the SCOT ontology.

### 3 Design Principles of the SCOT

In this section, the principles of the development of the SCOT ontology are described. The SCOT ontology<sup>1</sup> is an ontology for sharing and reusing tag data and representing social relations among individuals. It provides the structure and semantics for describing resources, tags, users and the extended tag information such as tag frequency, tag co-occurrence frequency, and tag equivalence. As the SCOT ontology is established to support semantically social tagging, the framework is designed to represent tag data at both personal and group-level. There is an opportunity to build group-level SCOT through integrating multiple personal SCOT ontologies. This idea has emerged on the following major principles:

- *lightweight*. Our approach follows the principle “a little semantic goes a long way[6].” The ontology model must be designed both with minimal structure and minimal semantics. Accordingly, the SCOT is expressed in a simple RDF format.
- *share and reuse*. In order to share and reuse the data to other applications, the ontology model must provide a consistent method for sharing existing sets of tags among users.
- *compatibility*. Except for the core concepts of our ontology that represent tagging activities, there is no intention to define redundant classes and properties that already exist in other RDF vocabularies.

### 4 Overview of the SCOT

The SCOT ontology generically models tagging activities for typical online communities and relations between each component of the activity. Figure 1 shows simplified model of SCOT ontology with its main top-level concepts and relations.

---

<sup>1</sup> <http://scot-project.org/scot/ns/>

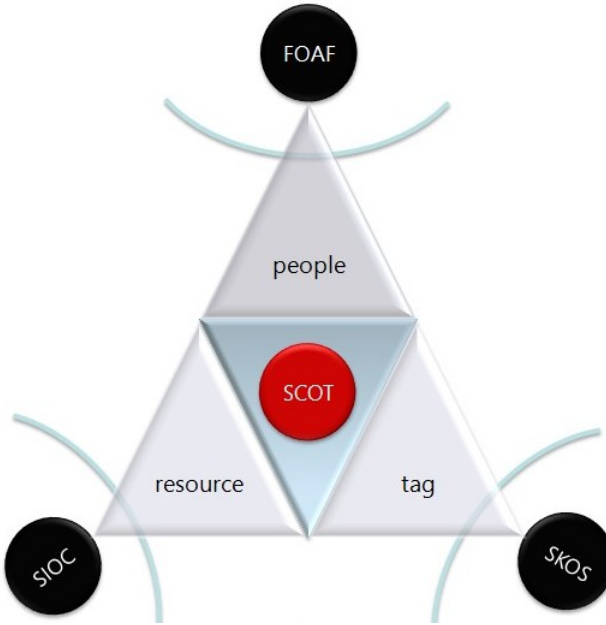


Fig. 1. SCOT Ontology Model

The SCOT ontology is linked to the three dimensional relationships that are represented in SIOC[7,10], FOAF[9], and SKOS[8]. We use the SIOC concepts to describe site information and relationships among container-item as well as site-site, and use the FOAF concepts to represent either a human or machine agent because a tag can be generated manually by a human user or automatically by a machine. Also we try to represent the relationships among users. This relationship has the two aspects: agent-agent and agent-group. When we are tagging and are using them, we assume these relationships. Finally, we use the SKOS to allow semantically relate a tag with another tag using properties such as `skos:broader` and `skos:narrower`. The SCOT consists of the Tagcloud, the Tag and the Cooccurrence class as the core concepts. In the following section, we describe in detail.

#### 4.1 Tagcloud Class

The `Tagcloud` class is the central element of the ontology. Basically the term *tagcloud* is used for navigation and visualization of content tags used on a website. Often, more frequently used tags are displayed in a larger font or otherwise emphasized, while the displayed order is generally by alphabet or by popularity. In the sense, “*text cloud*” can be considered a more accurate term than tagcloud from the visualization point of view.

Tagging itself, however, includes users, tags, and the relation among them. The tagcloud, the result of tagging, represents *user-tag* relation and *tag-tag*

relations for participants of tagging activities. The term tagcloud in our model has comprehensive perspectives to represent the entire tagging activity rather than just a simple visualization.

The `scot:Tagcloud` is a class that can be used to represent groups of users and a set of tags for users in certain sites or forums. For instance, there are the properties such as `scot:hasUsergroup`, `scot:hasMember`, `scot:composedOf`, `scot:totalTagFrequency`, `scot:totalCooccurFrequency`, `scot:totalItems`, `scot:totalTags`, `dcterms:created`, and `dcterms:modified` and so on.

The range of `scot:hasUsergroup` property is an instance of any of the classes such as the `Usergroup`, `User`, `Site`, and `Forum` from SIOC and `Agent`, `Group` from FOAF. Thus, a user group of the `scot:Tagcloud` class can be considered a common interest group according to set of tags. The `scot:composedOf` property is to describe that if a `Tagcloud` has more than two `Tagcloud` (i.e. group SCOT), each `Tagcloud` is part of the `Tagcloud`.

## 4.2 Tag Class

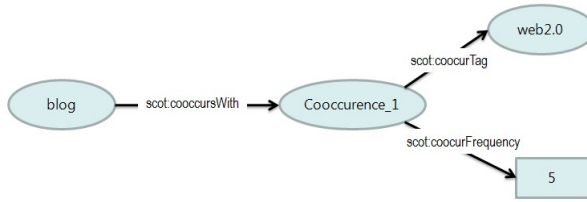
The `Tag` class is used to represent the concept of a tag and its statistical and linguistic properties. This class, a member of `Tagcloud` class, has a name through URIs. There are some well-known flaws in using free-tagging classification such as “*tags variation*”, “*tags ambiguity*”, and “*flat organization of the tags*” [11]. Those flaws are critical barriers to more precise categorization and to methods for better navigation.

***statistical properties.*** A single tag has a frequency which is the number of occurrences. In SCOT, a tag’s frequency is represented by the `scot:Frequency` property. This property has a value from the XML schema datatype `xsd:integer`.

A certain tag would often appear tags, and the meaning of the tag become more specific when the tag is combined with a set of tags. Generally it is so called ‘*tag co-occurrence*’ with those other tags. It can play an important role to reduce ‘*tag ambiguity*’. In SCOT, a tag co-occurrence is represented by the `scot:cooccurTag` property. When we use this property, we must consider the frequency of the individual tag itself and also the frequency among co-occurring tags. Thus we define two properties for this purposes. The First is `scot:ownFrequency` property that is used for representing a frequency of a single tag. The next is `scot:cooccurFrequency` to which describes co-occurrence frequency amongst a set of tags.

Both properties are subproperty of `scot:Frequency` property which is an instance of `rdf:Property`. These three properties (`scot:Frequency`, `scot:ownFrequency`, `scot:cooccurFrequency`) are called “*statistical properties*” of tag.

When we want to describe cooccurrence and its frequency among tags, we need to represent it by n-ary relations. For instance, let us assume that the tag *blog* has a cooccurrence with *web2.0* with a frequency of 5 (see Figure 2). There is a binary relation between the Tag *blog* and the tag *web2.0* and there is a cooccurance value describing this relation. We try to represent this information using an



**Fig. 2.** Cooccurrence and its frequency between blog and web2.0

instance of the class `scot:Cooccurrence`. The individual `Cooccurrence_1` here represents a single object encapsulating both the tag (`web2.0`, a specific instance of `scot:Tag`) and the cooccurrence value among the individuals (5).

**linguistical properties.** These properties are used to enhance the semantics of individual tags and to create connections between different variations of tags. In SCOT all tags have a concept and can be represented by a hierarchy among tags in SKOS. The `skos:broader` property is used to describe a more general term inverse of `skos:narrower`. In fact, it is not easy to build a tag hierarchy from a set of individual tag. However, we can build tag frequencies using the frequency of the tag and co-occurrence among tags, and can then represent their concept hierarchy based on this information. It can be provided a different structure to visualize a tagcloud beyond the flat organization of the tags.

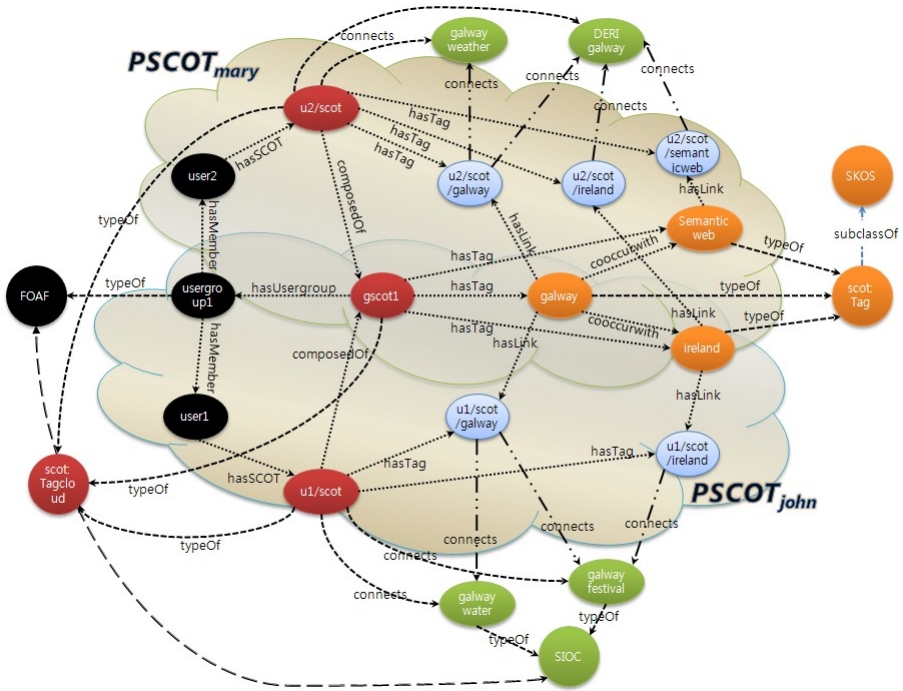
A tag would often appear in different conventions across a systems or web sites. Firstly, let us think of the term '*Weblog*'. We can find some different conventions such as '*weblogs*' as the plural, '*weblog*' in the lowercase, or '*blog*' as the short term. All terms can be considered as having the same meaning, if not the intended purpose. In addition, we can consider different examples like '*Semantic Web*'. In this case, the issue is placed on how to combine a compound word rather a number or case-sensitive version of the term. We can intuitively think '*semantic\_web*', '*SemanticWeb*', or '*Semantic-Web*' as having same meaning with '*Semantic Web*'.

The above examples are typical real-world cases encounter when we use a tag. Even if some popular Web 2.0 sites provide methods to handle different conventions, they do not allow one to represent this information explicitly. For solving these flaws, we define several properties such as `scot:spellingVariant`, `scot:acronym`, `scot:plural`, `scot:singular`, `scot:synonym`, `scot:delimited`, `scot:hyphenated`, `scot:underscored`, `scot:slashed`, and `scot:spaced` and so on. It can reduce tag ambiguity from acronyms and even recommend more common patterns of tag name.

## 5 How SCOT Space Expands

A huge amount of tagging data has been created from different users, applications, and domains. But it is difficult to share and reuse the tagging data across the different environments. SCOT provides a comprehensive way for sharing and

reusing the tagging data. Basically SCOT can be created by the efforts of a single user or a machine agent. In this case, the user owns the tagcloud. We are called it as the *Personal SCOT*. However, if a tagcloud is made by a group of users or reflects tagging efforts from different sites, its ownership is not an individual, but rather includes all those who participate in this activity. Such group of users would share the tagcloud under a boundary. We call it as the *Group SCOT*.



**Fig. 3.** Relationship between RDF vocabularies and SCOT: *gscot1* is composed of both *PSCOT\_mary* and *PSCOT\_john*. FOAF has information about Mary and John and also has URIs for their SCOTs. Each SCOT can be connected with *sioc:Post* as the URIs.

A user can create a personal SCOT from his/her data source and then publish and share it. The personal SCOT can be used for searching in the user’s site or blog, but also can be contained in other’s SCOT. This means that SCOT can be composed of combinations of either multiple personal SCOTs or multiple group SCOTs. Moreover, SCOT can be created across different types of applications such as weblogs, social bookmarking sites etc.

We make it possible to exchange the tagging data and to navigate resources using SCOT across various and varied applications. It make it possible to build the SCOT spaces which is the network based on user-driven tagging data. In addition, It can be considered as a kind of folksonomy for those in the group. Figure 3 illustrates the scenario to interconnect across users or communities and shows how the SCOT is connected with other RDF vocabularies.

## 6 Related Work

In [1], the author describes purposes of “semantics for tagging” with two aspects: to help in people’s understanding and to allow computers to process the tag for supporting people’s understanding.

The semantics of a tag is primarily about the meaning among people or a community group in the social space. There are several efforts that try to represent the concept of tagging, the operation of tagging, and the tag themselves.

Newman[14] describes the relationship between an agent, an arbitrary resource, and one or more tags. In his ontology, there are three core concepts such as Taggers, Tagging, and Tags to represent tagging activity. Taggers are represented by `foaf:Agents` and Taggings “reify the n-ary relationship between a tagger, a tag, a resource, and a date.” Tags are represented by the Tag class with URIs and are linked to `skos:Concept` and `skos:subject`.

Gruber[15] describes the core idea of tagging that consists of object, tag, tagger, and source. One notable thing is that he defines the source as the “scope of namespaces or universe of quantification” for objects. This allows one to differentiate between tagging data from different systems and is the basis for “collaborative tagging across multiple applications.”

Knerr[17] describes the concept of tagging in the *Tagging Ontology*. Since his approach is based on the ideas from [14] and [15], the core element of the ontology is Tagging. The ontology consists of time, user, domain, visibility, tag, resource, and type.

The approaches in the related work are focused on tagging activities or events that people used to tag in resources using terms. Therefore the core concept of the ontologies is Tagging, and there are Tagger and Resource class to represent user and resource respectively. However, there are no way to describe frequency of tags in the ontologies. The SCOT ontology is easy to represent this information using three properties of frequency. In addition, we provide a number of properties to represent social tagging activity and relationships among elements occurring on online community.

## 7 Conclusion

We have introduced the SCOT ontology, which focuses primarily on representing the uniform structure and the semantics for tag data. And we also have proposed how to share tag data among different sources or different people.

A tag can be used with many different meanings depending on user’s contexts. Thus, when tags are represented by kinds of ontology, we should take into account relationships among a tagger, a resource, and a tag. The SCOT ontology can describe three concepts and the relationships to define context for a tagging activity. Furthermore, the SCOT has some properties for handling tag frequencies and tagging activities such as total posts, number of users etc. In other words, the ontology has numerical and linguistic properties. The classes and the properties of SCOT is essential to represent tag data and share them.



Our approach is a starting point to share tag data beyond sites or people. To realize SCOT space, we have provided the SCOT Exporter that exposes a SCOT instance from legacy databases and the `int.ere.st` web site to search, bookmark, integrate, and share SCOT instances. We plan to extend and improve the Exporter and the `int.ere.st` for applying to various online communities. We provide more detailed information about the SCOT ontology at <http://scot-project.org>.

## Acknowledgments

This study was partly supported by a grant of the interoperable EHR Research and Development (A050909), Ministry of Health & Welfare, Republic of Korea and the Science Foundation Ireland under Grant No. SFI/02/CE1/I131.

## References

1. Beckett, D.: Semantics Through the Tag. In: Proceeding of XTech 2006: Building Web 2.0 (2006), Available at: <http://xtech06.usefulinc.com/schedule/paper/135>
2. TagCommons project website, <http://tagcommons.org>
3. SCOT project website, <http://scot-project.org>
4. Kim, H.L., Yang, S.K., Breslin, G.J., Kim, H.G.: Simple Algorithms for the SCOT Exporters using Tag Frequency and Cooccurrence. In: Proceedings of the 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2007), Silicon Valley, USA (2007)
5. Kim, H.L., Yang, S.K., Song, S.J., Breslin, G.J., Kim, H.G.: Tag Mediated Society with SCOT Ontology. In: Proceedings of the Semantic Web Challenge 2007 in conjunction with the Sixth International Semantic Web Conference, Busan, Korea, November 11-15 (2007)
6. Hendler, J.: The Dark Side of the Semantic Web. *IEEE Intelligent Systems and Their Applications* 22(1), 2–4 (2007)
7. Breslin, J.G., Harth, A., Bojars, U., Decker, S.: Towards Semantically-Interlinked Online Communities. *The UPGRADE European Journal for the Informatics Professional* 6(6) (2005)
8. Brickley, D., Miles, A.: SKOS Core Guide-W3C Working Draft 2 (2005), Available at: <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>
9. Brickley, D., Miller, L.: FOAF Vocabulary Specification (2005), Available at: <http://xmlns.com/foaf/0.1>
10. Brickley, D., Stefan, S., Miles, A., Miller, L., Caoimh, D.O., Neville, C.M.: SIOC Ontology Specification (2007), Available at: <http://rdfs.org/sioc/spec/>
11. Passant, A.: Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs: Theoretical background and corporate use-case. In: Proceeding of International Conference on Weblogs and Social Media, Boulder, Colorado, U.S.A. (2007)
12. Finin, T., Ding, L., Zhou, L., Joshi, A.: Social networking on the semantic web. *The Learning Organization: An International Journal* 12(18), 418–435, May 1 (2005)
13. Lux, M., Granitzer, M., Kern, R.: Aspects of Broad Folksonomies. In: 4th International Workshop on Text-Based Information Retrieval (2007)

14. Newman, R.: Tag Ontology design, (viewed 03/04/2007) (2005), Available at: <http://www.holygoat.co.uk/projects/tags/>
15. Gruber, T.: Ontology of Folksonomy: A Mash-up of Apples and Oranges, First on-Line conference on Metadata and Semantics Research (MTRSR 2005), <http://www.metadata-semantics.org/>, (to appear in International Journal on Semantic Web and Information Systems vol 3(1) (2007)
16. Gruber, T.: Collective Knowledge Systems: Where the Social Web meets the Semantic Web. To appear in Journal of Web Semantics (2007), Available at: <http://tomgruber.org/writing/Collective-KnowledgeSystems.htm>
17. Knerr, T.: Tagging Ontology- Towards a Common Ontology for Folksonomies (2006), Available at: <http://code.google.com/p/tagont/>