int.ere.st: Building a Tag Sharing Service with the SCOT Ontology

HakLae Kim, John G. Breslin
Digital Enterprise Research Institute
National University of Ireland, Galway
IDA Business Park, Lower Dangan
Galway, Ireland

SungKwon Yang, SeongJae Song, HongGee Kim
Biomedical Knowledge Engineering Lab
Seoul National University
28-22 Yeonkun-Dong, Chongno-Ku
Seoul, Korea

Abstract

Although a number of people have participated in tagging activity on many Web 2.0 sites and the sites allow users to get their tagging data through Mash-up APIs, there are many issues (e.g. tag semantics) associated with social tagging or folksonomies. Many studies have focused on semantics of a tag that offers a way of exact meaning of an individual tag or among the tags. Surprisingly, there has been minimal research regarding tag sharing or exchange.

We suggest a semantic model, is called the 'SCOT (Semantic Social Cloud of Tags)', to represent the structure and the semantics of tagging data. We also discuss social tagging processes in order to provide semantic interoperability of tagging data among diverse sources.

The `int.ere.st` web site aims to allow users to search and bookmark their social tagging data and to help the sharing and exchange of tagging data among people or heterogenous sources. We give an overview of `int.ere.st` with the social tagging processes and describe specific algorithms to support the processes.

Introduction

As user generated content on the Web has become a mainstream phenomenon, "more and more people will be exposed to the concept of tagging" (TagCommons 2007). 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 that 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 social bookmarking and tagging systems do not provide a uniform way to share and reuse tag data among users or communities. Although most popular Web 2.0 sites such as Del.icio.us, Flickr, or YouTube provide XML or JSON based data 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*" (TagCommons 2007) on different sources. It makes difficult to share and reuse tag data among users or across different services.

It is necessary to adopt Semantic Web technologies to solve the limitations. Although a number of studies have been made on Web 2.0, little attention has been given to bridge Web2.0 from Semantic Web perspective. There is a gap between Semantic Web research topics and Web 2.0 applications, since much Semantic Web research has thus far been focused on developing standards and recommendations. On the other hand, Web 2.0 plays important role by leading users to participate in online communities. Technologies for Web 2.0, however, are not mature enough to deal with effective and efficient services, in particular, those associated with social tagging and folksonomy. We believe bridging between two technologies will be an optimized solution to solve these limitations.

In this paper, we focus on how to solve the limitations related to social tagging using Semantic Web and Web 2.0 technologies. In particular, the contributions of this paper are:

- We suggest a semantic model to represent the structure and semantics of tagging data in social tagging spaces. The SCOT as a semantic model provide a complete set of information for social tagging.
- `int.ere.st` helps users to enhance the sharing and exchange of tag data among people or various online communities. `int.ere.st`[1] provides interlinked semantic data among FOAF, SIOC, and SCOT. Therefore, results of user contributions in the site are a good way to build social networks based on social tags with semantic links.

Introduction to SCOT

The SCOT (Social Semantic Cloud of Tags) ontology[2] is an ontology for sharing and reusing tag data and for representing social relations across different sources. It provides the structure and semantics for describing resources, tags, and users, and provides extended tag information such as synonym, spelling variant, tag frequency, tag co-occurrence frequency, and tag equivalence in order to reduce tag ambiguity. Our approach follows the principle "*a little semantic*

[1] `http://int.ere.st`
[2] `http://scot-project.org/scot/ns`

goes a long way" (Hendler 2007). The ontology model is designed both with minimal structure and minimal semantics in a simple RDF format. In order to share and reuse the data with other applications, the ontology model provide a consistent method for sharing existing sets of tags among users.

The SCOT Ontology Model

The SCOT ontology generically models tagging activities for typical online communities and relations between components(i.e. users, tags, resources etc) of the activity. We recapitulate the formal model for folksonomy introduced in (Hotho *et al.* 2006a).

A formal model of the SCOT ($\mathbb{S}$) is a tuple

$$\mathbb{S} := (U, T, R, Y)$$

where

- U: set of users who participate in tagging activity
- T: set of tags that is assigned in resources
- R: set of resources that has a indefinitely unchanged link that is called permalink
- Y: a ternary relation between U, T, and R (i.e. $Y \subseteq U \times T \times R$), that represent tagging.

There is an implication that T has unique URIs for representing a tag in a resource. For instance, del.icio.us and Flickr have their unique tag URIs such as *http://del.icio.us/tag* and *http://www.flickr.com/photos/tags*, respectively. Using the URI of T, we can connect and navigate the resources even if we do not have an actual resource information. In this perspective, a tag can be distinguished with a keyword or a single term without a specific URI. R has an indefinitely unchanged link that is called *permalink*. Therefore, an individual tag and its URI of T are connected with the permalinks of R in real world.

Figure 1 shows the simplified model of SCOT ontology with its top-level concepts and with relations with other existing vocabularies. The concepts user, tag, resource for the SCOT have the links to FOAF(Brickley & Miller 2005), SKOS(Brickley & Miles 2005), SIOC(Breslin *et al.* 2005), respectively.

We use the SIOC concepts to describe site information and relationships among *container-item*, *site-site*, and use the FOAF concepts to represent a human or machine agent as a tag can be generated either manually by a human user or automatically by a machine. Also the model attempts to represent the relationships among users. This relationship has the two aspects: *agent-agent* and *agent-group*. Finally, we use the SKOS to represent semantically relationships between each tag using properties such as `skos:broader` and `skos:narrower`.

SCOT concepts and properties are formally defined to achieve the model. There are core concepts `scot:Tagcloud` and `scot:Tag` in the SCOT ontology. The `scot:Tagcloud` is a class that can be used to represent this information for the tagcloud itself. The `Tag` class, a member of Tagcloud class, is used to represent the concept of a tag that has a name through URIs. All

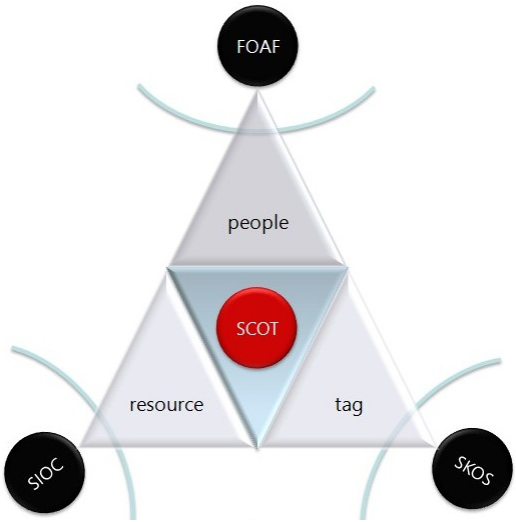

Figure 1: SCOT Ontology Model

tags have a concept and can be represented by a hierarchy(`skos:broader` and `skos:narrower`) among tags in SKOS. It can provide a different structure to visualize a tagcloud beyond the flat organization of the tags and will be an alternative way to overcome the problem of flat organization of tags.

Overview of int.ere.st

The main purpose of `int.ere.st` is to enhance the sharing and exchanging of tag data among people or various online communities(Kim *et al.* 2007). A number of social bookmarking and tagging sites have become popular recently, and tagging in traditional web sites is being adopted at a good pace. However, tagging data from these sites without a social exchange is regarded as an individual set of metadata rather than a social one. Although tagging captures our individual conceptual associations, upon the tagging system itself does not promote a social transmission that is converged by both creator and consumers.

To achieve social transmission environments for tagging, we need a comprehensive semantic model to represent tagging activity and a service to encourage its exchange. We already introduce the SCOT ontology for the former requirement. The latter can be realized by `int.ere.st`. Figure 2 shows the interface for the site. The left of Figure 2 shows search results with core information for identifying tag data and the right shows more detailed tag information of a selected result such as member name, list of tags and co-occurring tags with their frequencies, and items (resources).

We take some use cases on tag sharing from TagCommons[3] to make implementation guidelines. The use cases (TagCommons 2007) are very informative in that they provide technical and functional requirements for tag sharing. Qualitative analysis help with functionalities for `int.ere.st`. The requirements include *Personal Book-*

[3] `http://tagcommons.org`

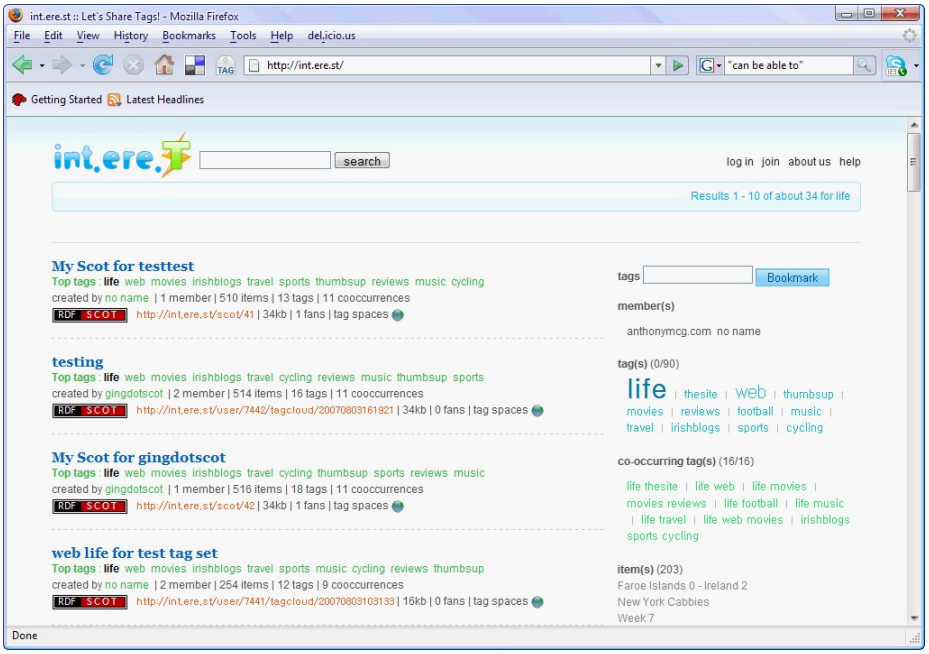

Figure 2: int.ere.st

marking across tagging sites, Browsing and Searching Others Tag Data Across Sources, Social Re-Search Using Tag Data, Multimedia Cross Reference, Organizing Documents Using Tags, Tag MetaSearch and Meta-Monitoring, Social Research on Collective Intelligence, and Distributing Tagged Information to the Semantic Web for identifying tag sharing as shown in Table 1. We find that all the use cases except use case 7 are very close to the services of `int.ere.st`.

Cases	support	services
Case 1	✓	integration among different sources
Case 2	✓	tag-based search, tagcloud browsing
Case 3	✓	user search, my fans
Case 4	✓	integration among different sources
Case 5	-	-
Case 6	✓	meta tag search, bookmark with tags
Case 7	✓	tag sharing (let's share menu)
Case 8	✓	interlinks among SIOC, FOAF, and SCOT

Table 1: Services for int.ere.st and the use cases for tag sharing

How int.ere.st works?

`int.ere.st` is a social tagging, bookmarking, and sharing service for SCOT ontology metadata. With int.ere.st, users can import, search, bookmark, and share their own as well as others' SCOT ontologies (see Figure 3). This functionality, ultimately, help users to exchange and share their tagging data based on the SCOT ontology. In addition, `int.ere.st` enables users to create Semantic Web data, such as FOAF, SIOC automatically. The RDF vocabularies can be interlinked with the URIs of SCOT ontologies that are generated in the site and shared in online communities.

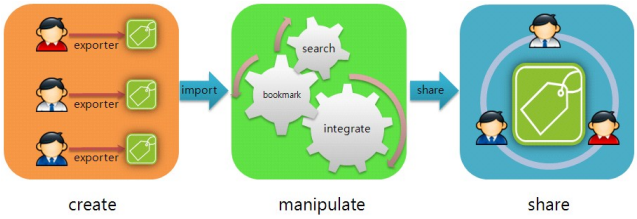

Figure 3: Functional Workflow for int.ere.st

Social Tagging Process with SCOT

One of the goals for `int.ere.st` is to provide a way for a user to manage tagging data from different contents and different tag spaces. With the SCOT Exporter[4], a user can create an instance of the SCOT ontology from a single online community such as weblog. The ontologies can be shared and reused as RDF itself. The Exporter, however, provides a simple method for exposing a SCOT ontology, it is necessary to support a method for managing and retrieving one. To achieve comprehensive management of personal tagging data, we provide an importing method in which a user can import his/her SCOT as a file or URL. Then, the aggregator for the imported SCOT run by periodically and automatically.

There are several ways to search tag information on the site. First, a tag search allows users to look for similar pat-

[4] `http://scot-project.org/?page_id=12`

terns of tagging or persons with their interests based on tags. A user can find out tags or resources using SPARQL-based semantic search methods as described the search operators:

- and : '&' sign (ex. web & blog)
- or : space (ex. web blog)
- co-occurring tags: '+' sign (ex. web + blog)
- broader relationship: '>' sign (ex. web > blog)
- narrower search: '<' sign (ex. blog < web)

The operators enable users to restrict their search conditions. Each operator is translated into a SPARQL-based query format on run time. When choosing one of the search results, the users can get meta-information for each SCOT ontology such as members, top tags, creator, total posts, and total tags (see Figure 4).

Social search, influenced by human judgement, takes many forms, ranging from simple shared bookmarks or content tagging to more sophisticated approaches that combine human intelligence with algorithmic searches.

When the '*created by*' from search results is clicked for a specific SCOT ontology, all SCOT ontologies created by the creator are listed. We provide the 'fans' as a concept for a list of people: when someone has added a certain SCOT as a bookmark, a fan connection can be created. If a user makes a new group SCOT with existing group SCOT ontologies, the user can get various types of information such as networks and bookmarks. This will help users to find interesting new people in the system, much as a user refers to ontologies to find interesting new ones.

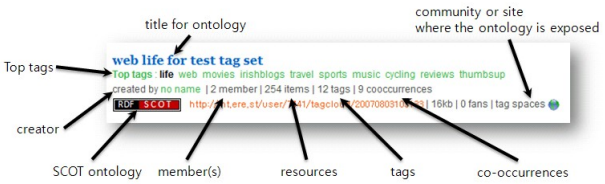

Figure 4: a search result

Social Tag Network with SCOT

`int.ere.st` provides a bookmarking and meta-tagging method for each SCOT ontology so that the user can participate in the tagging activity and share the experiences of other people. If users is interested in certain users or tags in their ontologies, they can create a bookmark, with tags, for the ontologies. The SCOT ontology in `int.ere.st` can be classified with several types such as *imported*, *interested*, and *grouped* ontology according to a creator. The interested type is created by other users while the grouped type is created by the logged in user themselves. The imported ontology can be one of two. The interested ontology can be made by bookmarking process and a list of interested ontologies is located in the "***my interests***" menu. The grouped ontology listing in "***my scots***" can be built by integrating interested ontologies. `int.ere.st` allows users to make their network both a positive and a passive way. The former means that a user can add to his network by bookmarking or by building grouped ontologies, while the latter is a result of other's activities that someone bookmarks your ontology. There is "***my fans***" menu to describe a tag network. The number of fans tells you how many users you are referring to, or you are referred by. Social connections among users can be made with their tag usage patterns. In other words, this approach focuses on people-centric social network based on tags.

`int.ere.st` exposes various and structured types of user contributions in the system and also connects to other sources of data using Semantic Web technologies. For instance, personal information can be exposed as FOAF and a SCOT ontology in the system can be mapped into the SIOC ontology. An intested ontology is described as foaf:interest and the grouped ontology is mapped as foaf:maker. In addition, all types of the SCOT ontology for a certain user are mapped with sioc:Item. This process can be done automatically. The mapping among FOAF, SIOC, and SCOT together provides a way to enhance social connections that are distributed and shared among people with semantic links.

Algorithms

In this section we describe methods to build a hierarchical structure and to integrate different ontologies based on tag frequencies. Frequencies can be expressed as absolute frequencies or relative frequencies. The absolute frequencies are raw observations, that have not been normalized with respect to the base rates of the event in question. When you speak about the frequency of tags it usually means the absolute one. The 'relative frequency' means a frequency that is expressed in relation with to a sample size or rate. This format is used to compare the occurrence of objects in two or more groups. Accordingly, this format is used in the tag clouds where the size of each tag represents the proportion of the tags. Simply stated, an absolute frequency, actual number of frequency, tells us how a frequent tag is in a given domain while a relative frequency of each tag means its proportion in the total tag occurrence.

Tag Hierarchy Computation

The objective of this algorithm is to construct a conceptual hierarchy of tags based on their relative frequencies and co-occurrence frequencies. It provides a statistically organized structure over large social tagging spaces. The hierarchy of tag data is driven by the Co-occurrence-based Tag Hierarchy Computation, as shown in Algorithm 1. We assume that a hierarchical structure from this algorithm can be changed if tagging data is updated. Thus, it might not be the same as general taxonomy or classification systems.

Note that *cooccur.getTagList()* in the `Algorithm 1` returns the cooccurring tag list. *tag.narrower()* and the *tag.broader()* describes the relationships among tags.

Suppose that the tags ***semanticweb*** and ***blog*** have 0.4 and 0.14 values of the relative frequency[5] respectively and they appear 14 times together. We called it (i.e. 14) as *ACF(Absolute Co-occurrence Frequency)* for the tag set

[5] `http://www.blogweb.co.kr/scot/scot.rdf`

Algorithm 1 Co-occurrence-based Tag Hierarchy Computation

Require: ACFrequency is a absolute frequency of co-occurrence for a certain tag
var *tagList* ← *empty list*
for all $tag \in tagList$ **do**
  $tag.ACFrequency \leftarrow 0$
end for
for all $cooccur \in cooccurList$ **do**
  for all $tag \in$ cooccur.getTagList() **do**
    tag.ACFrequency ← *tag.ACFreqeuency* + *cooccur.AFrequency*
  end for
end for
for all $cooccur \in cooccurList$ **do**
  $tags \leftarrow cooccur.getTagList()$
  if $tags.count == 2$ **then**
    if *tags*[0].*ACFrequency* × *tags*[0].*RFrequency* > *tags*[1].*ACFrequency* × *tags*[1].*RFrequency* **then**
      $tags[0].narrower(tags[1])$
      $tags[1].narrower(tags[0])$
    end if
  end if
end for

(***semanticweb*** and ***blog***). We can get the proportions 5.6 (=14×0.4) and 1.96 (=14×0.14) for ***semanticweb*** and ***blog***, respectively. Therefore, we can say the ***semanticweb*** is broader than ***blog***. It is described using skos:broader and skos:narrower properties in the SCOT ontology as shown in Listing 1.1.

Listing 1: broader/narrower Relationship among tags

```
<scot:hasTag>
  <scot:Tag rdf:about="http://www.example.com/tag/semanticweb">
    <scot:name>Semantic Web</scot:name>
    <scot:ownAFrequency>158</scot:ownAFrequency>
    <scot:ownRFrequency>0.4</scot:ownRFrequency>
    <skos:narrower rdf:resource="http://www.example.com/tag/blog" />
  </scot:Tag>
  <scot:Tag rdf:about="http://www.example.com/tag/blog">
    <scot:name>blog</scot:name>
    <scot:ownAFrequency>54</scot:ownAFrequency>
    <scot:ownRFrequency>0.14</scot:ownRFrequency>
    <skos:broader rdf:resource="http://www.example.com/tag/
          semanticweb" />
  </scot:Tag>
</scot:hasTag>
```

Multiple Ontology Integration

We provide a common mechanism to integrate multiple SCOT ontologies from different sources. An integrated ontology gives a customized perspective for a certain user. A user can create or manipulate a SCOT ontology as his/her interests.

Existing folksonomy systems are covered in various domains with whole set of data obtained from the systems. At the moment in these systems, it is not easy to produce specific and restricted subjects across different spaces. For instance, some SCOT ontologies in the system have multiple references such as WordPress[6], CiteULike[7], mar.gar.in[8] and so on. So it is possible to provide a rich reference of connections for tag data. In addition, the integrated ontology gives an useful information to analyze social relations among users that are included in the ontology.

Algorithm 2 describe the process to integrate multiple ontologies. The objective of this algorithm is to construct unified tag data based on their tag frequencies and co-occurrence frequencies.

Algorithm 2 Frequency-based Ontology Integration

Require: : tagcloudList is a list of tagclouds
var tagList ← empty list
var totalTagFrequency ← 0
var cooccurList ← empty list
var totalCooccurFrequency ← 0
for all tagcloud ∈ tagcloudList do **do**
  for all tag ∈ tagcloud.getTagList() do **do**
    if tagList.exist(tag) **then**
      t ← tagList .same(tag)
      t.AFrequency ← t.AFrequency + tag.AFrequency
    else
      tagList.add(tag)
    end if
    totalTagFrequency ← totalTagFrequency + tag.AFrequency
  end for
  for all cooccur ∈ tagcloud.getCooccurList() do **do**
    if coocurList.exist(cooccur) **then**
      c ← cooccurList.same(cooccur)
      c. AFrequency ← c. AFrequency + cooccur.AFrequency
    else
      cooccurList.add(cooccur)
    end if
    totalCooccurFrequency ← totalCooccurFrequency + cooccur.AFrequency
  end for
end for
for all tag ∈ tagList do **do**
  tag.RFrequency ← tag.AFrequency / totalTagFrequency
end for
for all cooccur ∈ cooccurList do **do**
  cooccur.RFrequency ← cooccur.AFrequency / totalCooccurFrequency
end for

A user can see a list of bookmarked ontologies, that are saved by the user, in the "***my interests***" menu and can make an integrated ontology with checking and clicking the 'build' button. It is also possible to give a title for the ontology.

[6] http://wordpress.org
[7] http://www.citeulike.org
[8] http://mar.gar.in

Related Works

A number of attempts have been made to copy with solving the limitations of folksonomies by various approaches. In particular, Semantic Web researchers have become increasingly interested in studying these topics. Gruber(Gruber 2007) holds the view that the Semantic Web technologies including ontologies can be interoperated with collective metadata.

Schmitz(Schmitz *et al.* 2007; Schmitz 2006) observes small world effects by analyzing a network structure of folksonomies of Bibsonomy (Hotho *et al.* 2006b) and del.icio.us. He introduces the notions of clustering and characteristic path length to describe the small world effects. According to his study, folksonomies exhibit a small world structure and have a sort of social networks. Mika (Mika 2005) carries out a study to construct community-based semantics based on tripartite model of actors, concepts, and instances. He emphasizes the social context for a representation of ontologies and generates the well-known co-occurrence network of ontology learning and a novel semantic network based on community relationships using del.icio.us data. Schmitz focuses on that folksonomies have some features of social networks reflecting social context.

The semantics of a tag is primarily about the agreement on 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(Gruber 2005; Newman 2005; Knerr 2006; TagCommons 2007). 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 ways 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.

Conclusions

`int.ere.st` provides better metadata creating and sharing support across online communities. In addition, `int.ere.st` makes it possible to exchange tagging data and to navigate resources using the SCOT ontology. All kinds of user contributions in the system will be exposed as RDF vocabularies based on SIOC, FOAF, and SCOT.

We believe it is a good starting point to build social semantic spaces based on using tagging data. Even if our approach and tool do not promise to solve all problems related to social tagging, we believe our approach can limit the impact of some of them. We will provide further information through the project web site (http://scot-project.org).

Acknowledgments

This material was supported in part by Science Foundation Ireland under Grant No. SFI/02/CE1/I131.

References

Breslin, J. G.; Harth, A.; Bojars, U.; and Decker, S. 2005. Towards semantically-interlinked online communities. In Gomez-Perez, A., and Euzenat, J., eds., *European Semantic Web Conference (ESWC)*, volume 3532 of *Lecture Notes on Computer Science*, 500–514. Springer.

Brickley, D., and Miles, A. 2005. SKOS core vocabulary specification. W3C working draft, W3C. http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102.

Brickley, D., and Miller, L. 2005. Foaf vocabulary specification. Working draft.

Gruber, T. 2005. Ontology of folksonomy.

Gruber, T. 2007. Collective knowledge systems: Where the social web meets the semantic web. *To appear in Journal of Web Semantics* 2–4.

Hendler, J. 2007. The dark side of the semantic web. *IEEE Intelligent Systems* 22(1):2–4.

Hotho, A.; J?schke, R.; Schmitz, C.; and Stumme, G. 2006a. Information retrieval in folksonomies: Search and ranking. In Sure, Y., and Domingue, J., eds., *The Semantic Web: Research and Applications*, volume 4011 of *LNAI*, 411–426. Heidelberg: Springer.

Hotho, A.; schke, R. J.; Schmitz, C.; and Stumme, G. 2006b. Bibsonomy: A social bookmark and publication sharing system. In de Moor, A.; Polovina, S.; and Delugach, H., eds., *Proceedings of the First Conceptual Structures Tool Interoperability Workshop at the 14 th International Conference on Conceptual Structures*, 87–102. Aalborg Universitetsforlag.

Kim, H.; Yang, S.; Song, S.; Breslin, J. G.; and Kim, H. 2007. Tag mediated society with scot ontology. In *Semantic Web Challenge on International Semantic Web Conference*.

Knerr, T. 2006. Tagging ontology- towards a common ontology for folksonomies.

Mika, P. 2005. Ontologies are us: A unified model of social networks and semantics. In *International Semantic Web Conference*, LNCS, 522–536. Springer.

Newman, R. 2005. Tag ontology design.

Schmitz, C.; Grahl, M.; Hotho, A.; Stumme, G.; Catutto, C.; Baldassarri, A.; Loreto, V.; and Servedio, V. D. P. 2007. Network properties of folksonomies. In *Proc. WWW2007 Workshop "Tagging and Metadata for Social Information Organization"*.

Schmitz, C. 2006. Small world folksonomies: Clustering in tri-partite hypergraphs.

TagCommons. 2007. Functional requirements for sharing tag data.