Personal knowledge management for knowledge workers using social semantic technologies

Haklae Kim*, John G. Breslin and Stefan Decker

Digital Enterprise Research Institute, National University Ireland, Galway, IDA Business Park, Upper Newcastle, Galway, Ireland
Fax: +353-91495541
E-mail: haklae.kim@deri.org
E-mail: john.breslin@deri.org
E-mail: stefan.decker@deri.org
*Corresponding author

Jaehwa Choi

Department of Business Administration, Dankook University, AnSeo-dong, Chonan, Chungnam, Korea
E-mail: jchoi@dankook.ac.kr

Honggee Kim

Biomedical Knowledge Engineering Lab, Seoul National University, Jong-Ro Gu, Yeon-Gun Dong, Seoul, Korea
E-mail: hgkim@snu.ac.kr

Abstract: Knowledge workers have different applications and resources in heterogeneous environments for doing their knowledge tasks and they often need to solve a problem through combining several resources. Typical personal knowledge management (PKM) systems do not provide effective ways for representing knowledge worker’s unstructured knowledge or idea. In order to provide better knowledge activity for them, we implement Wiki-based sociAl Network Thin client (WANT) that is a wiki-based semantic tagging system for collaborative and communicative knowledge creation and maintenance for a knowledge worker. And also, we suggest the social semantic cloud of tags (SCOT) ontology to represent tag data at a semantic level and combine this ontology in WANT. WANT supports a wide scope of social activities through online mash-up services and interlink resources with desktop and web environments. Our approach provides basic functionalities such as creating, organising and searching knowledge at individual level, as well as enhances social connections among knowledge workers based on their activities.

Keywords: semantic web; semantic wiki; personal knowledge management; PKM; folksonomy; social tagging.
1 Introduction

The goal of personal knowledge management (PKM) aims to make knowledge workers better at capturing, sharing and using knowledge and maximising their personal effectiveness in the social aspect of their jobs (KM Magazine, 2004). Knowledge workers have been using desktop and web applications to capture ideas or thoughts and to manage schedules, addresses or tasks, etc. There are many different applications such as e-mail clients, word processors and web browsers for knowledge worker’s daily routine. But since knowledge is scattered across applications and websites, it is not easy to aggregate or combine a right set of knowledge and operations required for their specific tasks for knowledge workers. In addition, knowledge can be evolved by social interaction among knowledge workers; knowledge worker’s activities have social characteristics such as connecting, communicating, and collaborating with others. In this sense, PKM should support both individual and social level for knowledge activities. Traditional PKM systems tend to be suitable for supporting specific functions such as managing schedule and address for individuals. This makes it difficult to extend and combine to other applications or services.

There are emerging trends associated with computing environments that support personal knowledge activities. ‘Web 2.0’ comprises of technologies and services to enable users to collaborate and share social contents. They include social software, content syndication, messaging protocols such as weblogs, wikis, podcasts, really simple syndication (RSS) feeds, etc. The majority of popular Web 2.0
sites, Flickr (http://www.flickr.com), del.icio.us (http://del.icio.us) and Technorati (http://www.technorati.com), are connecting people into communities creating networks of shared experiences using folksonomy and RSS. ‘Social semantic desktop’ can provide reliable technologies to enhance functionalities of PKM. The social semantic desktop is a new computing paradigm that provides an advanced way to create, automate and structure information and the technology convergence including the social network, community services and P2P services (Decker and Frank, 2004). It could provide a transformation of a typical desktop system into a collaborative environment that supports both personal computing and information sharing via social channels.

PKM is not only focused on managing data, but also on connecting people and sharing data among them. New social and semantic technologies can be able to provide knowledge workers to organise their thoughts and ideas in a relevant, timely manner. In this paper, we focus on social tagging as a way of representation and sharing for knowledge worker’s ideas or thoughts and wiki for organising their daily tasks. Social tagging can be an effective method to support, extend or derive values from human social behaviour. Wiki allows knowledge workers to make their internal knowledge more explicit and more formal. A proposed system aims to combine both social tagging and wiki features to improve effectiveness and efficiency for various activities of knowledge workers.

This paper is organised as follows: Section 2 describes a type of knowledge and collaborative tagging for social features of knowledge and summarise about semantic wikis. Section 3 describes limitations of current tagging systems and introduces Social Semantic Cloud of Tags (SCOT) – representation of tag data at a semantic level. Section 4 presents WANT – a wiki-based PKM system with design principles and its architecture. Section 5 describes the main features of WANT and explains how tags can be mapped to the SCOT ontology and we conclude in Section 6.

2 Related work

2.1 Types of knowledge

There are many definitions of knowledge and classifications or categorisations of knowledge. According to Spender (1994), knowledge can be classified into two dimensions: ‘explicit/tacit’ and ‘individual/social’ knowledge. Individual explicit knowledge (conscious knowledge) is located in an individual in the form of facts, documents and files that can be stored and represented from personal records. Individual tacit knowledge (automatic knowledge) means tacit knowing, including practical knowledge of people and performance of different types of skills. Social explicit knowledge (objectified knowledge) represents the shared corpus of knowledge by communities and social tacit knowledge is fundamentally embedded in the forms of social practice (Nahapiet and Ghoshal, 1998).

A knowledge process, in general, transforms individual knowledge to social knowledge. Knowledge is created through a social interaction in organisations or communities where knowledge workers are involved. But, traditional knowledge representation approaches constructed by domain experts (e.g., taxonomy, ontology) provide strict structures with a high-level formality for describing knowledge. These approaches are limited to represent individual knowledge worker’s unstructured thinking
and to support continuous feedbacks or interactions with others. We could come up with a collaborative tagging (also known as folksonomy, social classification, social indexing) as an alternative for knowledge representation and sharing from an individual knowledge worker’s point of view.

2.2 Tagging for representing knowledge

Tagging is a way of representing concepts by cognitive association techniques, without enforcing categorisation. A tagging system has been adopted in many social software applications such as weblogs, social bookmarking and social networking sites. This approach brings an important advantage to the knowledge workers in the form of a simple way to describe their knowledge in individual information spaces and to share it in online communities. A tag—a labelled keyword, is a type of metadata for a resource such as a resource link, a web page, a picture, a blog post, etc. The resources can be tagged with as many tags as desired because there are no restrictions on which or how many to use.

The result of participating tagging activities can be represented by folksonomies. Folksonomies, a term first coined by Tomas Vander Wal in 2004, are user-generated and distributed classification systems, emerging through bottom-up consensus (Merholz, 2004). The essence of folksonomies is user participation and internet-mediated social interaction. The tags in folksonomies are chosen by knowledge workers and may be reused and shared by other knowledge workers. Since a large number of users participate in creating, adding and sharing metadata in the form of keywords, folksonomic tagging is regarded as a social and democratic process (Golder and Huberman, 2006) and as a collective and social knowledge. Quintarelli (2005) points out that, ‘without social distributed environment that suggests aggregation, tags are just flat keywords’.

Knowledge workers do not necessarily have to be an expert, but can also be a creator or consumer of the content. They can collaboratively create and manage tags to annotate and categorise content. This activity establishes social connections among them and improves social reinforcement.

2.3 Semantics in wiki

A number of subsequent attempts have been made to solve the limitations of knowledge activities by various approaches. In particular, semantic web researchers have become increasingly interested in studying wiki (Hepp et al., 2007). Although wiki systems administrate collaborative contents, they only provide a limited number of functions for structuring the contents. Content in typical wikis is encoded in HTML, making it difficult to represent semantics for the content. A semantic wiki is a wiki system that has an underlying semantic model of the knowledge described in its pages (http://en.wikipedia.org/wiki/Semantic_wiki). These approaches aim to combine semantic data into HTML contents and to enhance machine-readable performance. There are several semantic wiki implementations such as Platypus Wiki (Campanini et al., 2004), Rhizome (Souzis, 2005), Semantic MediaWiki (Volkel et al., 2006), etc.
3 Social semantic tagging

Although a number of studies have been made on Web 2.0, little attention has been given to Web 2.0 from a 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 an 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 the social tagging and folksonomies (Gruber, 2007). For instance, a critical problem in typical tagging systems is that they do not provide a uniform way to share and reuse tag data amongst users or communities. Although most popular Web 2.0 sites such as del.icio.us and Flickr provide XML or JSON-based data using open APIs, there is no uniform structure or semantics to represent tag data. Therefore, it is not easy to meaningfully search, compare or merge ‘similar collective tagging data’ (Tagcommons, 2007) from different sources. This makes it difficult to share, reuse and integrate tag data among users or across different services. From a knowledge worker’s point of view, the limitations can be a barrier to adopt tag-based knowledge representation.

3.1 Overview of SCOT ontology

The SCOT ontology (http://scott-project.org) is an ontology for sharing and reusing tag data and for representing social relations across different sources (Kim et al., 2008). It provides the structure and semantics for describing resources, tags, and users and provides extended tag information such as synonyms, spelling variants, tag frequencies, tag cooccurrence frequencies and tag equivalence in order to reduce tag ambiguity. Our approach follows the principle ‘a little semantic 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 provides a consistent method for sharing existing sets of tags amongst users.

3.2 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 a folksonomy introduced in Hotho et al. (2006). A formal model of SCOT (S) is a tuple:

\[ S := (U, T, R, Y) \]  

where

- \( U \): set of users who participate in the tagging activity
- \( T \): set of tags that is assigned to resources
- \( R \): set of resources each of which has an 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 represents tagging.

There is an implication that $T$ has unique URIs for representing each tag for all resources. 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 actual resource information. From this perspective, a tag can be distinguished resources with a keyword or a single term without a specific URI. $R$ has an indefinitely unchanged link that is called a ‘permalink’. Therefore, an individual tag and its URI in $T$ are connected with the permalinks of $R$ on a web page.

Figure 1 shows a simplified model of the SCOT ontology with its top-level concepts and with relations to other existing vocabularies. The concepts user, tag and resource for the SCOT have links to FOAF (Brickley and Miller, 2005), SKOS (Brickley and Miles, 2005) and SIOC (Breslin et al., 2006), respectively.

**Figure 1** SCOT ontology model (see online version for colours)

We use SIOC concepts to describe site information and relationships among ‘container-item’, ‘site-site’ and use FOAF concepts to represent a human or machine agent as a tag. These 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 two aspects: ‘agent-agent’ and ‘agent-group’. Finally, we use SKOS to represent semantic relationships between each tag using properties such as ‘skos:broader’ and ‘skos:narrower’.

SCOT concepts and properties are formally defined to fully describe the tagging model. There are core concepts called ‘scot:Tagcloud’ and ‘scot:Tag’ in the SCOT ontology. ‘scot:Tagcloud’ is a class that can be used to represent information for the tag cloud itself. The ‘Tag’ class, a member of the ‘tag cloud’ class, is identified to represent the concept of a tag that has a name through URIs. All tags have an associated concept and can be represented by a hierarchy (‘skos:broader’ and ‘skos:narrower’) among tags using SKOS. It can provide a different structure to visualise a tag cloud beyond the flat organisation of tags and will be an alternative way to overcome problems with such a flat organisation.

4 Wiki-based social network thin client

4.1 Design principles

The proposed system aims to manage daily routine for knowledge workers more effectively and efficiently. We have used personal digital assistants such as ‘to do’ lists, calendars, address books, appointment books on both desktop and web environments. The proposed system must provide integrated functionalities to support seamless knowledge activities. In order to do this, the proposed system must support interoperability among different platforms without loss of information and functionalities. Also, there are two important design principles for developing the proposed system:

• easy to combine online mash-up services: knowledge workers tend to do their knowledge activities across desktop and web environments. For instance, they would use online services such as del.icio.us or YouTube in order to manage and share their knowledge. Therefore, the proposed system must get information from the web and integrate it with local data or vice versa.

• semantic tagging: as we mentioned in Section 3, tagging can be an effective method to represent knowledge for knowledge workers – both at individual and community level. But, a tag itself has limitations to describe the meaning of context in order to precisely search resources. In addition, most wiki systems do not provide tag features. The SCOT ontology can be helpful to describe semantic relationships among tags. In order to support this, a proposed system should have a method to map between tags and a SCOT ontology.

4.2 Architecture

We now describe our prototype application called WANT. WANT [as extension to the TiddlyWiki (http://tiddlywiki.com)] is a lightweight desktop wiki for PKM rather than full-fledged semantic desktop applications. Figure 2 shows the architecture of WANT. WANT is implemented as an interface with wiki-like features using JavaScript, AJAX and JSON (interface layer), thus, allowing to easily extend with certain functionalities using JavaScript plug-ins and to be easily deployed on web browser environments. All content in the system are stored in HTML itself, and published using RSS formats and
SCOT ontology in the decentralised knowledge source layer. The mash-up services layer plays an important role to connect desktop resources with web resources using semantic web technologies: ‘connection to semantic desktop’. All contents can be saved in HTML files, RSS1.0, RSS2.0 and Atom formats in WANT. There are two ways to connect the semantic desktop components. Firstly, since WANT is a simple and single HTML file, it could be opened by most semantic desktop applications. Secondly, all content of WANT would be generated by RSS1.0 format and be saved in RDF storage. When querying with RDF query languages such as SPARQL, users would be able to get the data set in the RDF storage and return back as value the new semantic data.

**Connection to Web 2.0**

Wikis are easy to use as collaboration platform and knowledge management systems. However, using wikis will significantly limit the potential for information sharing and collaboration in desktop environments. We overcome some of the weak points of wikis by using Web 2.0 technologies. WANT allows knowledge workers to organise their information or knowledge and provides various social content services such as folksonomies, social bookmarking and RSS/Atom feeds using HTTP, SOAP, XML RPC or REST web services. For instance, when knowledge workers make their own tags for certain content, they can use not only desktop tags which they made before, but also a folksonomy of weblogs or Technorati. All tags in the system can be translated and mapped to a SCOT instance.

WANT uses browser-based interfaces. Figure 3 shows a sample content about the picture ‘simplicity-desktop’ from Flickr as rendered in WANT. It can include pictures, desktop resources and links.
Figure 3 WANT interface: the picture ‘simplicity-desktop’ from Flickr (see online version for colours)

Note: WANT includes the RSS reader, tag cloud and social bookmark reader.

5 How it works

5.1 Enabling social collaboration and authority

A user can create new content with a new link and modify existing contents in an editing mode. All contents in the system could contain title, description, created date and tags. These elements could enrich the metadata of the contents. The description allows a user to organise his/her ideas or information about what they want to create. At the same time, they can add some tags to an item in order to classify and categorise the content. The system allows a user to assign multiple tags with separating comma like ‘Web 2.0, desktop’. After creating content with tags, the tags can be located at the bottom of the content and simultaneously the tags will be added in the tag cloud if the lists of tags are not defined in the system. The tag cloud shows a list of popular tags. The tags can be later used for navigating and finding contents. This can be a starting point to navigating contents via tags.

One of the main goals of WANT is to foster and employ social interactions for knowledge workers through various content services and Web 2.0 sites. To leverage the social collaboration, we need to shift from focusing on the individuals to focusing on
interactions. Social collaboration within WANT is in particular supported by the RSS reader, social bookmark reader and Flickr photo reader. Knowledge workers are able to participate in communities sharing common interests and to have the links for references. In WANT, it enables the users not only to capture web-based information, but also to organise the information together with desktop resources. The collaboratively user-added annotations are to improve social features on desktop resources. For instance, RSS reader and social bookmark reader can get the data from a user-given URL such as a certain RSS feed or bookmark URLs. The user can edit this data directly and add user-driven annotations as tags in WANT. These tags can be connected with social communities as specific links. So, contents on WANT easily reflect social content without the need for the user’s additional effort.

Figure 4 RSS reader (1), social bookmark reader (2) and Flickr photo reader (3) (see online version for colours)

5.2 Tag-based search

WANT provides two search methods: full-text search and tag-based search. The former returns results for matching one or multiple keywords in content, while the simple tag-based search navigates all tag lists in the tag cloud with the given tag. The tag cloud is a list of the most popular tags used by the user. The larger the font size of the tag, the greater its popularity. When the user navigates by a tag, he/she will be directly connected with other resources using the tag cloud, which is clickable. It will improve the
findability of content. If the user clicks on a certain tag, a list will pop up with links to the other content having the same tag. This is supposed to give an indication of the most popular topics. The availability of new semantic data will allow users to find and make use of relevant data quickly and accurately.

Figure 5  Tag-based search (see online version for colours)

Note: When a user clicks a certain tag in the tag cloud, it shows a list of contents using the tag.

5.3 Enhancing semantics

There are several ways to store contents in the system:

1. static HTML files
2. RSS files
3. RDF repository.

Basically, all contents can be stored in HTML files which includes URLs, tags, contents, links and date, etc. At the same time, all contents can be published with different syndication types such as RSS1.0, RSS2.0 and ATOM.

It is a slightly different approach to the other semantic wikis. Most of those have a way to generate semantic data in their tools directly. In our approach, we create contents and syndication files, then we generate semantic data automatically to reuse and share it. It is a more effective and efficient lightweight mechanism to connect with the semantic desktop. In addition, if necessary, it can be integrated with other semantic desktop applications because it is a simple HTML. It enables users to publish and share their contents as RSS on the web. Also, if they want to collaborate with someone, they exchange the RSS using an RSS reader. Thus, it is easy to cooperate in distributed computing environments.
5.4 Mapping tags to SCOT ontology

Although RSS provides minimum functionality to describe semantically the contents in the system, relationships among tags are still ambiguous. The SCOT ontology defines each tag and its relationships at a semantic level, as well as frequencies for each tag. Figure 6 presents a conceptual model for mapping tags to concepts in the SCOT ontology. All tags define a concept as a ‘scot:Tag’ and can be represented by a hierarchical structure among tags based on SKOS (Miles and Bechhofer, 2008). The ‘skos:broader’ property is used to describe a more general term and is the inverse of ‘skos:narrower’. There are two types of frequencies (i.e., absolute vs. relative) for representing a tag frequency: ‘scot:AFrequency’ and ‘scot:RFrequency’. The former is intended to describe an absolute value and the purpose of the latter is to represent the relative value as a proportion of total occurrence. A single tag can have both frequency formats. In particular, this information is represented by the ‘scot:ownAFrequency’ and ‘scot:ownRFrequency’ properties which are sub-properties of ‘scot:AFrequency’ and ‘scot:RFrequency’, respectively. Listing 2 displays the tags ‘blog’ and ‘web’. The tag ‘blog’ has 110 as an absolute frequency and 40.29% as a relative frequency. And it can be seen that the tag ‘web’ is broader than the tag ‘blog’ and also there is a cooccurrence relation between the two tags (i.e., ‘Cooccurrence_0’).
A certain tag may often appear together with other tags. The meaning of the tag can become more specific when the tag is combined with a set of tags and the frequency of cooccurrence of two tags has more significance than the two individual ones. To simply define the term ‘cooccurrence’: if an item contains both tags ‘blog’ and ‘web’, these tags are said to cooccur or have a first order cooccurrence that is observed when tags cooccur in the same items. It can play an important role in reducing ‘tag ambiguity’. When describing cooccurring tags and their frequency among them in Figure 6, we need to represent it by n-ary relations. The tag ‘blog’ has a cooccurrence with ‘web’ with AF and RF of 5 and 0.4, respectively. The facts are represented using an instance of the class ‘scot:Cooccurrence’. The individual ‘Cooccurrence_0’ here represents a single object encapsulating both the tag (web, a specific instance of ‘scot:Tag’) and the cooccurrence value among the individuals (5). A tag cooccurrence is represented by the ‘scot:cooccurTag’ property. The ‘scot:cooccurAFrequency’ and ‘scot:cooccurRFrequency’ property describe the frequency of the absolute and relative cooccurrence amongst a set of tags (see Listing 3), respectively.

**Listing 2** Representation tag ‘blog’ and ‘web’ in SCOT

```xml
<scot:hasTag>
  <scot:Tag rdf:about=“http://www.example.com/want/tag/blog”>
    <scot:name>blog</scot:name>
    <scot:ownAFrequency>110</scot:ownAFrequency>
    <scot:ownRFrequency>40.29</scot:ownRFrequency>
    <scot:lastUsedDate>2007–09–25T02:15:17</scot:lastUsedDate>
    <scot:cooccurWith rdf:nodeID=“Cooccurrence_0”/>

  </scot:Tag>
</scot:hasTag> <scot:hasTag>
    <scot:name>web</scot:name>
    <scot:ownAFrequency>41</scot:ownAFrequency>
    <scot:ownRFrequency>15.02</scot:ownRFrequency>
    <scot:lastUsedDate>2007–09–05T05:55:31</scot:lastUsedDate>
    <scot:cooccurWith rdf:nodeID=“Cooccurrence_0”/>

  </scot:Tag>
</scot:hasTag>
```
Personal knowledge management for knowledge workers

Figure 6  Conceptual model for mapping tags in a resource with concepts in the SCOT ontology (see online version for colours)

Listing 3  Cooccurring relationships among two tags

```xml
<scot:Cooccurrence rdf:nodeID= "Cooccurrence_0">
  <scot:cooccurTag rdf:resource= "http://example.com/tag/web"/>
  <scot:cooccurTag rdf:resource= "http://example.com/tag/blog"/>
  <scot:cooccurrenceAFrequency>5</scot:cooccurrenceAFrequency>
  <scot:cooccurrenceRFrequency>0.4</scot:cooccurrenceRFrequency>
</scot:Cooccurrence>
```

6  Summary and conclusions

In this paper, we discussed the method for enhancing personal knowledge activities for knowledge workers. In fact, traditional PKM applications intend to manage a specific task such as a schedule or an address book. Knowledge workers have been collaborating with their colleagues or community’s users across desktop and web environments. Therefore, PKM systems must support social features for knowledge activities. It means that this system is not only focused on managing data, but also on connecting people and enabling them to share data between them.
We suggested wiki-based content management and tag-based knowledge representation to support personal knowledge activities and implemented WANT system.

The main achievement of this work, distinguishing our approach from existing systems, is the establishment of an architecture to interact desktop and web, to create metadata based on the SCOT ontology and to enrich desktop systems with social intelligence. Our approach allows knowledge workers to interact and share their resources among their desktop and social software more easily.

Acknowledgements

This study was supported by the Dankook University Post-Doc Grant in 2008, a grant of the Science Foundation Ireland under Grant No. SFI/02/CE1/I131, the interoperable HER Research and Development (A050909) and Ministry of Health and Welfare, Republic of Korea.

References


