CTKOS : Categorized Tag-based Knowledge Organization System

Donghee Yoo
Department of Information Science
Korea Military Academy
(donghee.info@gmail.com)

Gunwoo Kim
Department of Business Administration
Hanbat National University
(gkim@hanbat.ac.kr)

Keunho Choi
Korea University
Business School
(keunho@korea.ac.kr)

Yongmoo Suh
Korea University
Business School
(ymsuh@korea.ac.kr)

As more users are willingly participating in the creation of web contents, flat folksonomy using simple tags has emerged as a powerful instrument to classify and share a huge amount of knowledge on the web. However, flat folksonomy has semantic problems, such as ambiguity and misunderstanding of tags. To alleviate such problems, many studies have built structured folksonomy with a hierarchical structure or relationships among tags. However, structured folksonomy also has some fundamental problems, such as limited tagging to pre-defined vocabulary for new tags and the time-consuming manual effort required for selecting tags. To resolve these problems, we suggested a new method of attaching a categorized tag (CT), followed by its category, to web content. CTs are automatically integrated into collaboratively-built structured folksonomy (CSF) in real time, reflecting the tag-and-category relationships by majority users. Then, we developed a CT-based knowledge organization system (CTKOS), which builds the CSF to classify organizational knowledge and allows us to locate the appropriate knowledge.

Received : November 18, 2011 Accepted : December 15, 2011
Type of Submission : APJITM Special Corresponding author : Gunwoo Kim

1. Introduction

In an age of information overload, knowledge organization systems (KOSs) play an important role in organizing a huge number of materials (e.g., books, articles, and documents) by classifying the materials into more specific aspects of a topic, thereby helping us retrieve them easily and quickly (Hodge, 2000; Weller, 2007; Weller et al., 2010). KOSs typically impose a particular view of the world on the collected materials using various schemes for knowledge representation. The wide spectrum of the schemes used in KOSs ranges from a set of simple terms (e.g., keywords, glossaries, and dictionary) to a set of controlled terms with shallow hierarchy (e.g., taxonomies and classification schemes) and to a set of concepts and their relationships (e.g., thesauri and ontology).

However, since domain experts in a specific subject field or work area typically define most of the schemes used in these kinds of KOSs, they have the serious problem of not being flexible enough to adapt to the rapid changes in the online environment. In Web 2.0, users are willing to produce an enormous amount of user-generated information concerning web contents. Thus, the way in which experts or authorized groups classify information is inadequate to create proper information for indexing and retrieving all the web contents (Quintarelli, 2005; Matthews et al., 2010), especially new concepts captured from the social web (Mathes, 2004; Hammond et al., 2005; Kroski, 2005; Shirky, 2005).

In order to compensate for the shortcomings of KOSs, folksonomies are built from the perspective of knowledge organization and retrieval in the context of Web 2.0 (Weller et al., 2010). Folksonomies provide Internet users with a means of organizing various types of content by attaching meaningful tags to them. Folksonomies used by current KOSs are either flat or structured.

In KOSs with a flat folksonomy, users are free to attach a tag (or tags) to their web contents (e.g., web pages, images, videos) according to their own needs, because there are neither pre-defined, controlled keywords used as tags nor structures that categorize the tags. As more users participate in the creation of web contents, the totality of user-generated tags on the contents within a KOS forms a flat folksonomy (Brooks and Montanez, 2006; Angeletou et al., 2008), term that is used because no hierarchy is defined among the tags. However, although flat folksonomy has the significant advantage of flexible and easy choice of tags to organize and share user-generated contents (Mathes, 2004; Kroski, 2005; Weinberger, 2005), it has the fundamental problems of potential ambiguity and misunderstanding of tags among users. Such problems stem from semantic problems (Mathes, 2004) and/or cognitive problems (Golder and Huberman, 2006). These problems result in the retrieval of undesirable web contents when users try to search the web.

Because researchers understand that the main cause of these problems is the lack of semantics on tags, they have viewed the incorporation of semantics as the most important challenge (Hendler and Goldbeck, 2008; Pan et al., 2009; Limpens et al., 2010). Therefore, KOSs began to build structured folksonomies with a focus on augmenting the semantics on tags by building a hierarchical structure or relationships among tags. In those systems, users individually and manually define the meanings of their own tags by linking each tag to a term in a *pre-defined hierarchy of vocabularies* or to a concept that appeared in an *ontology* (Spyns et al., 2006; Marchetti et al., 2007, Passant, 2007; Buffa et al., 2008). Another approach is for machines to define tags automatically by capturing the semantics from the whole tag space through machine processing techniques (Abbasi et al., 2007; Mika, 2007; Specia and Motta, 2007). These structured folksonomy systems contributed to removing the ambiguity or misunderstanding of tags by resolving the semantic problem and the cognitive problem of flat folksonomy. However, current structured folksonomy systems also have fundamental problems, such as limited tagging to pre-defined vocabulary for new tags, time-consuming manual efforts for selecting tags, and difficulty in getting users' consensus for the meaning of tags. These problems may prohibit users from using tagging systems as a powerful instrument for facilitating knowledge organization and retrieval. So, we must build KOSs that allow users to construct a structured folksonomy that can be built collaboratively without limited vocabularies or concepts in real time.

To that end, we propose a new approach to the real-time construction of a hierarchical structure among the tags, augmented with semantics. In our approach, users are simply required to enter a tag along with the category to which the tag belongs. We call such a tag a categorized tag (CT). As more users enter CTs for the contents of a specific domain, CTs are integrated into a hierarchical structure of CTs, dynamically reflecting the tag-and-category relationships of majority users. Hereinafter, we will refer to collaboratively-built structured folksonomy as CSF. CSF allows us to alleviate the fundamental problems of structured folksonomy, while taking advantage of the features of flat folksonomy. Thus, the objective of the research was to propose a CT-based KOS (or CTKOS) that organizes web contents by building CSF dynamically.

The rest of the paper is organized as follows. Section 2 reviews several tagging systems that are closely related to our work. Section 3 describes CTKOS and explains how CSF helps users find more relevant documents than flat folksonomy does. Section 4 provides a summary of our conclusions and some limitations of our approach.

2. Previous Works

In this chapter, we review existing, structured folksonomy systems, address their problems collectively, and explain our approach to resolving the problems. Based on who assigns the proper meaning to a tag, two categories of existing tagging systems can be identified, i.e., one system that depends on human intervention and a second system that depends on machine-processing techniques. Systems that belong to the first category include DogmaBank (Spyns et al., 2006), semantic blogging system (Passant, 2007), and SweetWiki (Buffa et al., 2008). A feature these systems share is that humans define the meanings of tags by linking users' tags to the vocabularies defined in ontology (Gruber, 1993). In the DogmaBank, social bookmark system, when a user enters a keyword as a tag, the user receives a list of concepts associated with the keyword from WordNet Ontology (Spyns et al., 2006). Then, the user selects one of the concepts as a tag of a web page. If the user cannot find a relevant concept, he or she can suggest new tags and their definitions. In the *semantic blogging system*, the ambiguity of tags is removed by linking them to the concepts defined in a domain ontology (Passant, 2007). When entering a keyword in this system, a user must select a concept defined in the ontology associated with her or his blog post. In SweetWiki, which supports social tagging in order to categorize the Wiki documents, keywords that users enter are relocated by community experts into a more refined structure than before (Buffa et al., 2008).

Researches or systems belonging to the second category include T-ORG (Tag-ORGanizer) (Abbasi et al., 2007) and 'Integrating folksonomies with the semantic web' (Specia and Motta, 2007). In general, they define the meanings of tags from the tag space constructed by machine processing techniques, such as similarity computation, clustering, and network analysis. In the T-ORG system, users select ontologies that are associated with the categories of the resources to be tagged (Abbasi et al., 2007). Then, ontologies are modified by pruning unwanted concepts, refining redundant concepts, and adding missing concepts. Finally, the remaining concepts in the modified ontologies are used as categories, which, together with the tags and their contexts, are used as input to T-KNOW, which classifies the tags into the categories. In a research project entitled '*Integrating folksonomies with the semantic web*', Specia and Motta proposed an approach for making the semantics of tags explicit, which is based on the tag spaces in social tagging systems, such as Flickr and Del.icio.us (Specia and Motta, 2007). First, they filtered out unusual tags and then built clusters of similar tags. Finally, they defined the relationships between tags in each cluster.

Even though the tagging systems in two categories may solve the problems of flat folksonomy, they still have some problems in securing flexibility and ease in the choice of tags, which are the greatest advantages of flat folksonomy. In the first category of systems that require human intervention, when a user enters a new tag that does not match the concept defined in the ontologies, a flexibility problem occurs because the scope of individual tags should be restricted to the concepts pre-defined in the ontologies. Thus, a more flexible and simplified tagging method is required for incorporating tags with semantics. In the second category of systems that depend on machine-processing techniques, it is not guaranteed that the techniques correctly match the meanings of tags to the concepts or relationships in ontologies, because the ontologies may not include relevant concepts or relationships for the tags. In addition, manual intervention by domain experts sometimes is required to conduct decision making, such as selecting and modifying the ontologies. Thus, it is necessary to match simple, new tags to existing meaningful tags correctly based on users' consensus with minimum interventions by domain experts.

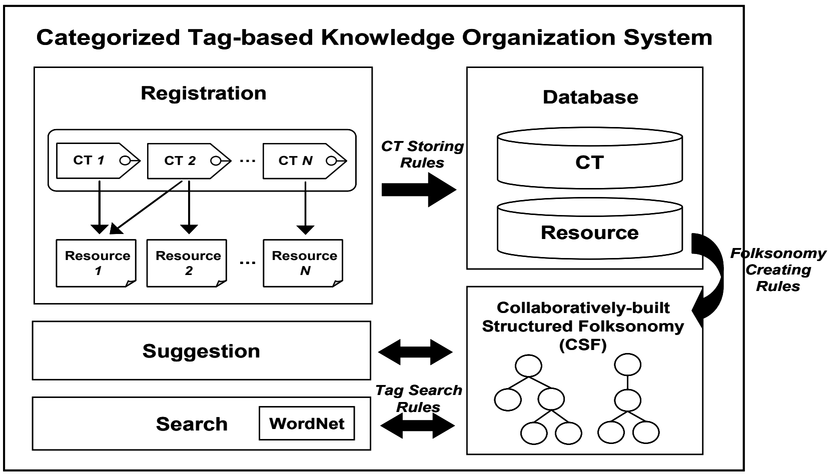
To alleviate these problems of existing, structured folksonomy systems, we propose a new approach to tagging, i.e., tagging with a category, or CT. In our approach, when a user attaches a tag to a resource, he or she must provide a tag and a category to which the tag belongs. Since our approach does not require the existence of a pre-defined ontology, it is more flexible in tagging than previous structured folksonomy systems. The next section describes our system in more detail.

3. CTKOS : Categorized Tag-based Knowledge Organization System

3.1 Overview of the System

CTKOS is a prototype system developed to create CSF by classifying and sharing the various types of organizational knowledge through social collaboration. <Figure 1> depicts the architecture of the system.

- Registration : When users register their organizational knowledge to share (e.g., documents), they should attach a new CT to it or select a specific CT for it among suggested CTs in order to categorize it.
- Database : Organizational knowledge is stored in the resource table and its CTs are stored in a CT table according to *CT storing rules*.
- CSF : When users enter organizational knowledge and its new CTs in the CT table collaboratively, CSF is built in real time according to *folksonomy creating rules*.
- Suggestion : When users register or search organizational knowledge, a list of CTs extracted from the CSF is shown to users to help them select appropriate CTs.



<Figure 1> CTKOS Architecture

- Search : This function allows users to search tagged organizational knowledge using either a keyword or CT. The keyword-based search consists of the simple text matching of keywords, while the CT-based search consists of semantic matching of tags, which provides various entry points in order to navigate documents along a tag hierarchy by *tag search rules*.

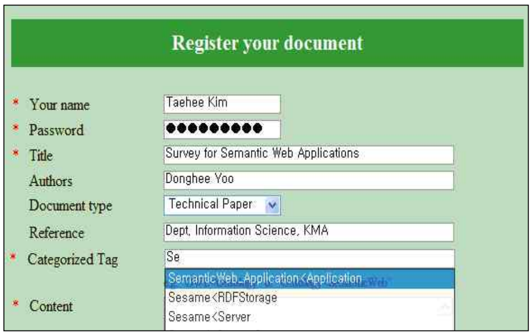
3.2 Registration of Organizational Knowledge

Users who want to store their organizational knowledge in CTKOS should enter a CT, which consists of two parts, i.e., a tag and a category. There should be an IS-A relationship between the two, i.e., the tag belongs to the category. They should enter a CT in the following form :

Tag < *Category*

When a tag is entered in this way, the meaning of the tag becomes more specific than it would otherwise be, because its category limits the context of the tag. CT would be more helpful, especially when tags have polysemy. For example, Bank < Financial_Institution implies that bank is used as a tag to mean a kind of financial institution, while Bank < Building implies that bank is used as a tag to mean a kind of building.

<Figure 2> shows a screenshot of a document registration page in CTKOS. First, a user enters various metadata about a document, e.g., title, authors, and document type, and then attaches one or more CTs to classify the document. As the user types CTs in the *Categorized Tag* field, an auto-completion mechanism suggests a list of potential CTs extracted from the CSF in the *Categorized Tag* field. The user can either select a particular CT from the list or enter a new CT. Through the registration page, the metadata of documents are stored in the resource table, and CTs are stored in the CT table.



<Figure 2> Registration Page and Auto-Completion in the Text Field of a CT

3.3 CT Storing Rules

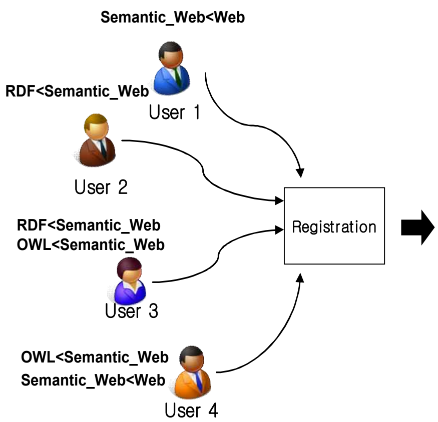
According to CT storing rules, CTs that are entered by users are stored in the CT table, the schema of which consists of {ID, Level_1_Tag, Level_2_Tag, Level_3_Tag, Count}, it is assumed that users can enter two CTs or less at a time. The CT storing rules are described as follows :

1. When one CT in the form of *tag* < *category* is entered by a user, the system checks to determine if there is an entry for the CT in the

CT table. If there is, it just increases the count field of the entry by 1. Otherwise, a new ID is assigned to the CT, *tag* of the CT is stored in Level_1_Tag, *category* of the CT in Level_2_Tag in the CT table and count field is set to 1.

2. When two CTs are entered in a row, which may look like "*tag* < *category1 category2* < *super_category*", the system stores each CT in the CT table as was done in the previous step.
3. When two CTs are entered in a row, which may look like "*tag* < *category category* < *super_category*", they are combined into a *three-level CT* in the form of "*tag* < *category* < *super_category*." If such a three-level CT were already stored in the CT table, only the count field of the three-level CT is increased by 1. Otherwise, each concept in the three-level CT is stored in Level_1_Tag, Level_2_Tag, and Level_3_Tag, respectively, and the count field is set to 1.

<Figure 3> describes an example of storing CTs that were entered by four different users into the CT table according to the CT storing rules described above. When user 1 enters '*Semantic_Web* < *Web*', a new ID (e.g., 1) is assigned to the CT, the tag (e.g., Semantic_Web) is stored in Level_1_Tag, the category (e.g., Web) is stored in Level_2_Tag in the CT table, and the count field is set to 1, indicating its first occurrence. Similar actions are taken for the input of user 2. If user 3 enters '*RDF* < *Semantic_Web OWL* < *Semantic_Web*', the count field of the first CT *RDF* < *Semantic_Web*, which was already stored in the table, is increased by 1. Then, one new entry in the table is made for the second CT *OWL* < *Semantic_Web*. When user 4 enters '*OWL* < *Semantic_Web Semantic_Web* < *Web*', the system checks whether the two CTs look like *tag* < *category category* < *super_category*. In this case, they can be combined into one three-level CT, '*OWL* < *Semantic_Web* < *Web*', and are stored, as



CT table

ID	Level_1_Tag	Level_2_Tag	Level_3_Tag	Count
1	Semantic_Web	Web		1
2	RDF	Semantic_Web		2
3	OWL	Semantic_Web		1
4	OWL	Semantic_Web	Web	1

<Figure 3> Example of storing CTs in the CT table Registration

shown in <Figure 3>. *Count* implies the degree of users' agreement for the category of a tag.

3.4 Folksonomy Creating Rules

From the CT table where two-level CTs and three-level CTs are stored, a hierarchical CSF is built automatically, according to the folksonomy creating rules, which can be described as follows:

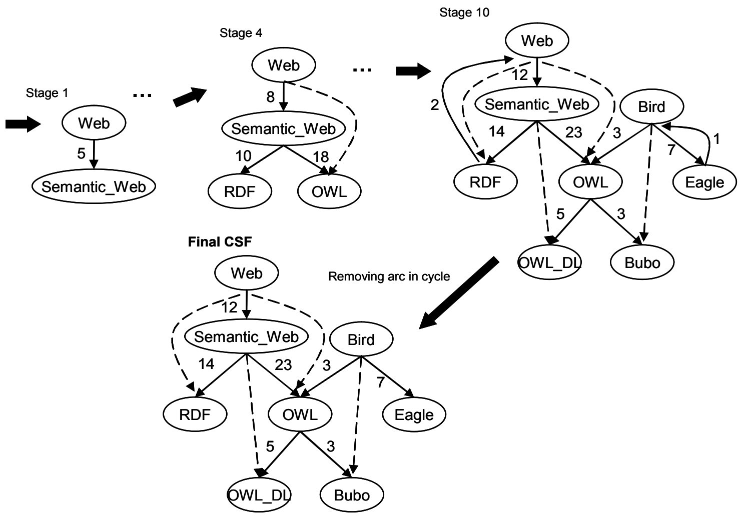
1. If there are two nodes and an arrow in the CSF graph, corresponding to a two-level CT in the CT table, just add the value of count field to the label of the arrow. Otherwise, create a node for each tag and connect them with an arrow from the Level_2_Tag to the Level_1_Tag and label the arrow with the value of the count field.
2. If there are nodes and an arrow in the CSF graph, corresponding to a two-level CT that is a part of a three-level CT, just add the value of the count field to the label of the arrow. Otherwise, create a node for each tag, connect them with arrows from Level_3_Tag to Level_2 _Tag and from Level_2_Tag to Level_1_Tag, and label the arrows with the value of the count field. In addition, connect the node representing the Level_3_Tag to that representing the Level_1_Tag with a dotted arrow.
3. If a cycle exists in the resulting graph, remove an arrow with the smallest label among the arrows that makes a cycle. (Existence of a cycle in the graph implies that some relationship among the concepts represented by the arrow is weaker than others in the cycle.)

<Figure 4> illustrates an example of creating CSF from the CT table. Note that, since the first CT OWL < Semantic_Web for the fourth entry of the CT table is already represented in the graph, its count is just added to the corresponding label. Its second CT is processed similarly, as indicated at stage 4 of <Figure 4>. Stage 10 in <Figure 4> represents the result of processing the 10th entry in the CT table. Also note that the resulting CSF graph has two cycles, the first among the nodes Web, Semantic_ Web and RDF, the second among the nodes Bird and Eagle. So, we removed the arrows with the smallest label, which are the arrow from RDF to Web in the first cycle, and the arrows from Eagle to Bird in the second cycle.

According to the CT storing rules and folksonomy creating rules, CSF is updated in real time when new CTs are entered in the table. As such, for the case in which several users enter their CTs that are related to each other at the same time, folksonomy creating rules may incorrectly remove an arrow among the arrows that make a cycle in CSF. Suppose that there is a cycle among three CTs, such as *RDF < Semantic_Web, Semantic_Web < Web*, and *Web < RDF*, and that their label values are 1, 2, and 3, respectively. Then, even though the first two CTs are correct and the last CT is incorrect, the arrow corresponding to the first CT will be removed. This potential error can be resolved as

CT table

ID	Level_1_ Tag	Level _2_ Tag	Level _3_ Tag	Count
1	Semantic_Web	Web		5
2	RDF	Semantic_Web		10
3	OWL	Semantic_Web		15
4	OWL	Semantic_Web	Web	3
5	Eagle	Bird		7
6	Bubo	OWL	Bird	3
7	OWL_DL	OWL	Semantic_Web	5
8	Bird	Eagle		1
9	RDF	Semantic_Web	Web	4
10	Web	RDF		2



——— From *Level 2 Tag* to *Level 1 Tag* or From *Level 3 Tag* to *Level 2 Tag*

- - - From *Level 3 Tag* to *Level 1 Tag*

<Figure 4> Example of Creating a CSF from the CT Table

more users who register CTs correctly participate in category tagging. In this way, CSF reflects more common, users' recognition for the relationship between each tag and its category, following the principle of *the wisdom of the crowds* (i.e., collective intelligence) (Lévy, 1999; Weiss, 2005; Tapscott and Williams, 2006).

3.5 Suggesting CT

When entering CTs, a user can choose one of several potential CTs extracted and suggested from the CSF or freely enter some new CTs. The suggested CTs are shown in a descending order of label values. After a user enters CTs, the CT table and the folksonomy are updated in real time based on the newly-entered CTs. This function saves the time required to enter CTs and also shows which pairs of tag and category were used most often.

3.6 Tag Search Rules

Tag search, which is conducted prior to document search, is conducted by CT-based search. It returns tag hierarchies following the tag search rules explained below, so that users can navigate the space of documents along each tag hierarchy. Tag search rules are as follows:

1. Retrieve the sub-tree of CSF whose root represents the *tag* or *category* part of the CT.
2. For each node at or below the third level of the sub-tree, remove it from the sub-tree if the node is not linked to its grandparent node by the dotted arrow.
3. Find all the paths from the root to each leaf of the sub-tree. Each path is returned as a tag hierarchy.

For example, suppose that a user enters a keyword 'web.' In the first step, a sub-tree that contains nodes *Web*, *Semantic_Web*, *RDF*, *OWL*, *OWL_DL*, *and Bubo* with root being *Web* is retrieved from the CSF in <Figure 4>. In the second step, nodes *RDF* and *OWL*, which are at the third level of the sub-tree, are checked to determine whether they are linked to their grandparent node *Web* by the dotted arrow. Since they are, they are not removed from the sub-tree. In the second step again, nodes *OWL_DL* and *Bubo*, which are at the fourth level of the sub-tree are checked. Since node *Bubo* is not linked to its grandparent node *Semantic_Web* by the dotted arrow, it is removed. As a result, we get a sub-tree that has two leaf nodes *RDF* and *OWL_DL*. As such, in the third step, we get only two tag hierarchies '*Web* → *Semantic_Web* → *RDF*' and '*Web* → *Semantic_Web* → *OWL* → *OWL_DL*.'

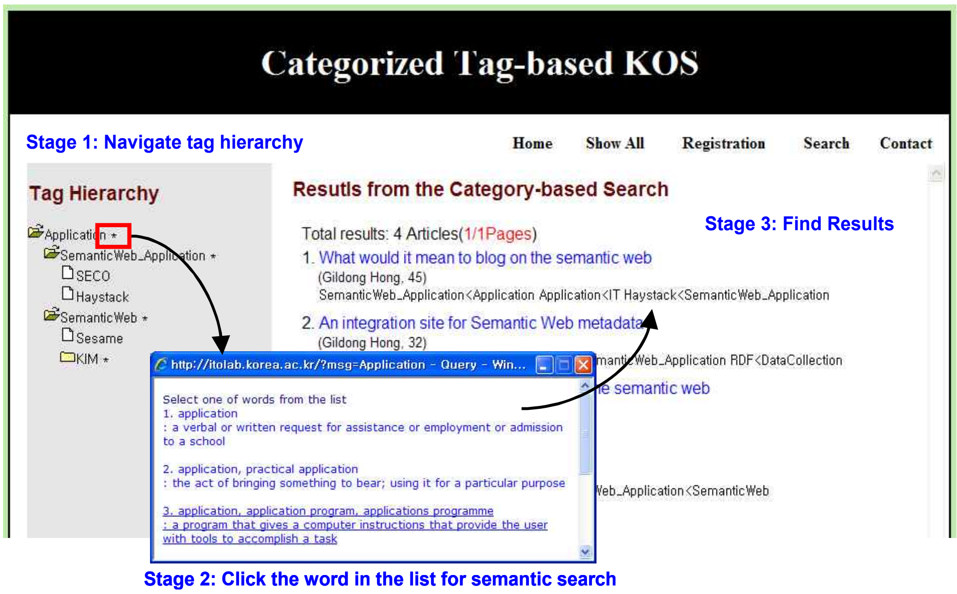
In the CT-based search function of CTKOS, each tag hierarchy from CSF is used to navigate organizational knowledge. The user enters a keyword or a CT in the text field of the search page and clicks the search button. Then, CTKOS, after referencing the CSF using the keyword or the CT, returns one or more tag hierarchies (stage 1). In tag hierarchies, each sibling tag is sorted in descending order of the label value of the corresponding arrow, and tags with the same label value are sorted by searching order. There are two types of emoticon in the tag hierarchy. A tag with a folder emoticon implies that it includes many sub-tags, and a tag with a document emoticon implies that it represents a document. Then, to supplement CT-based search semantically, synonyms of a tag selected in the hierarchy are extracted from WordNet and listed on the screen after a user selects a tag in the tag hierarchy (stage 2). Finally, all documents are searched and those that are related to the tag having the semantics of the selected synonym are shown to the user after the user clicks on one synonym in the list (stage 3).

CT-based search supports semantic search of documents by creating a tag hierarchy and by referencing WordNet. The tag hierarchy provides multiple entry points for search, while flat folksonomy systems provide a single entry point for search. Also, WordNet allows users to select a correct meaning of the tag among its synonyms, if any. As a consequence, CT-based search gives more opportunities to find web documents that match the user's interest.

In addition to the CT-based search, CTKOS also provides a means for a simple, keyword-based search, which is one of the functions that is used most frequently for searching documents. Given keywords by users, keyword search uses simple text matching algorithms to find, e.g., creator, title, author, and content.

3.7 Implementation

We used various web applications and open-source software to develop CTKOS. For example, we used a *tomcat* server to provide web-based interaction, and graphical interfaces and functions were implemented in JSP (Java Server Pages), Java Script, and Ajax. Three



<Figure 5> Overall Process of CT-based Search

rules, i.e., CT storing rules, folksonomy creating rules, and tag search rules, were all implemented in Java. All metadata and CTs for documents were stored as tables in a MySQL database.

4. Conclusions

Currently, flat folksonomy is known to effective for classifying a huge amount of Web contents regardless of the set of keywords. However, it has severe problems, such as the semantic problem and the cognitive problem, since flat folksonomy consists of only a set of simple tags. Many studies have been conducted to build structured folksonomy in order to resolve the problems of flat folksonomy. However, they still have some problems in terms of flexibility and ease in the choice of tags.

To alleviate the problems of structured folksonomies, we have proposed a new tagging approach, tagging with a category, called CT, to make simple tags into meaningful tags by real-time collaboration. Our approach provides the following contributions. First, users can freely input their tag with its category, without being limited to a pre-defined set of tags. Second, CSF is built from users' CTs collaboratively in real time, so that CSF always reflects user consensus for the relationship between each tag and its category. Third, we implemented a CTKOS in order to demonstrate how CSF can be used in document retrieval. CSF helps users navigate documents' space beginning at one of multiple entry points.

Also, our approach has an additional desirable feature. CSF in CTKOS reflects only the 'IS-A relationship' between each tag and its category, among many other potential relationships between tags. To make the CSF semantically more useful, it will be necessary to add other relationships, such as part-of, same-as, and inverse-of, to CSF. However, representing such relationships in CSF in a way that is useful for searching is another research problem. Even with the limitation, we contend that we have shown how CTs can be used to add structure to tags freely, collaboratively, and usefully. Therefore, we will extend the current work to many other folksonomy systems.

References

Abbasi, R., S. Staab, and P. Cimiano, "Organizing resources on tagging systems using T-ORG", *Proceedings of the ESWC 2007 Workshop Bridging the Gap between Semantic Web and Web 2.0*, Innsbruck, Austria(2007), 97~110.

Angeletou, S., M. Sabou, and E. Motta, "Semantically enriching folksonomies with FLOR", *Proceedings of the ESWC 2008 Workshop Collective Intelligence and the Semantic Web*, Tenerife, Spain, 2008.

Brooks, C. H. and N. Montanez, "Improved Annotation of the Blogopshere via Autotagging and Hierarchical Clustering", *Proceedings of the 15th international conference on World Wide Web*, Edinburgh, Scotland (2006), 625~632.

Buffa, M., F. Gandon, G. Ereteo, P. Sander, and C. Faron, "SweetWiki : a semantic wiki", *Journal of Web Semantics*, Vol.6, No.1(2008), 84~97.

Golder, S. and B. A. Huberman, "Usage patterns of collaborative tagging systems", *Journal of Information Science*, Vol.32, No.2(2006), 198~208.

Gruber, T., "A translation approach to portable ontology specifications", *Knowledge Acquisition*, Vol.5, No.2(1993), 199~220.

Hammond, T., T. Hannay, B. Lund, and J. Scott, "Social Bookmarking Tools (I) : A General Review", *D-Lib Magazine*, Vol.11(2005), http://www.dlib.org/dlib/april05/hammond/04hammond.html.

Hendler, J. and J. Goldbeck, "Metcalfe's law, web 2.0, and the semantic web", *Journal of Web Semantics*, Vol.6, No.1(2008), 14~20.

Hodge, G., "Systems of knowledge organization for digital libraries : beyond traditional authority files", *CLIR Publications*, (2000), http://www.clir.org/pubs/abstract/pub91abst.html.

Kroski, E., "The Hive Mind : Folksonomies and User-based Tagging", (2005), http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging.

Lévy, P., *Collective intelligence : Mankind's emerging world in cyberspace*, Perseus Books, Cambridge, MA, 1999.

Limpens, F., F. Gandon, and M. Buffa, "Helping online communities to semantically enrich folksonomies", *Proceedings of the Web Science Conference*, Raleigh, NC, USA 2010.

Marchetti, A., M. Tesconi, and F. Ronzano, "Semkey : a semantic collaborative tagging system", *Proceedings of the WWW 2007 Workshop Tagging and Metadata for Social In-*

formation Organization, Banff, Alberta, Canada, 2007.

Mathes, A., "Folksonomies : Cooperative Classification and Communication through Shared Metadata", 2004, http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html.

Matthews, B., C. Jones, B. Puzon, J. Moon, D. Tudhope, K. Golub, and M. L. Nielsen, "An evaluation of enhancing social tagging with a knowledge organization system", *Aslib Proceedings*, Vol.62, No.4/5(2010), 447~465.

McGuinness, D. L. and F. V. Harmelen, *OWL Web Ontology Language Overview (W3C Recommendation 10 February 2004)*, http://www.w3.org/TR/owl-features/.

Mika, P., "Ontologies Are Us : A Unified Model of Social Networks and Semantics", *Journal of Web Semantics*, Vol.5, No.1(2007), 5~15.

Pan, J. Z., S. Taylor, and E. Thomas, "Reducing ambiguity in tagging systems with folksonomy search expansion", *Proceedings of the 6th European Semantic Web Conference*, Heraklion, Greece, (2009), 669~683.

Passant, A., "Using ontologies to strengthen folksonomies and enrich information retrieval in weblogs", *Proceedings of the International Conference on Weblogs and Social Media*, Colorado, USA, 2007.

Quintarelli, E., "Folksonomies : power to the people", *Proceedings of the 1st International Society for Knowledge Organization*, Milan, Italy 2005, http://www.iskoi.org/doc/folksonomies.htm.

Shirky, C., "Ontology is Overrated : Categories, Links, and Tags", 2005, http://www.shirky.com/writings/ontology_overrated.html.

Specia, L. and E. Motta, "Integrating folksonomies with the semantic web", *Proceedings of the 4th European Semantic Web conference*, Innsbruck, Austria, 2007.

Spyns, P., A. Moor, J. Vandenbussche, and R. Meersman, "From folksologies to ontologies : how the twain meet", *Proceedings of the Move to Meaningful Internet Systems 2006 : CooPIS, DOA and ODBASE*, Montpellier, France, (2006), 738~755.

Tapscott, D. and A. D. Williams, *Wikinomics : How Mass Collaboration Changes Everything*, Penguin, New York, 2006.

Weinberger, D., "Taxonomies and tags : from trees to piles of leaves", *Esther Dyson's Monthly Report*, Vol.23(2005), 1~36.

Weiss, A., "The power of collective intelligence", *netWorker*, Vol.9(2009), 16~23.

Weller, K., "Folksonomies and ontologies : two new players in indexing and knowledge representation", *Proceedings of Online Information*, London, Great Britain (2007), 108~115.

Weller, K., I. Peters, and W. G. Stock, "Folksonomy : the collaborative knowledge organization system", In : Dumova T and Fiordo R (eds.) *Handbook of Research on Social Interaction Technologies and Collaborative Software : Concepts and Trends*, New York, USA, (2010), 132~146.

Abstract

카테고리형 태그 기반의 지식조직체계 구현

유동희* · 김건우** · 최근호*** · 서용무***

웹 2.0 환경에서는 사용자 참여를 통한 지식 생산이 증가하고 있으며, 이러한 지식들을 효과적으로 분류하고 공유하기 위한 방법으로 태그 기반의 단일 폭소노미가 활용되고 있다. 그러나 단일 폭소노미를 구성하는 태그들은 단일 태그로 구성되어 있어 그 의미가 불명확하다. 이를 위해 태그간의 관계성이 명시된 구조화된 폭소노미를 구축하는 연구들이 진행되었다. 그러나 이러한 연구들에서도 태그의 의미를 정의하기 위해 미리 정의된 용어집을 사용하기 때문에 용어집에 없는 새로운 태그는 정의할 수 없고, 수작업으로 태그의 의미를 정의할 때 시간이 소요되는 문제점이 발생하였다. 본 연구에서는 사용자가 태그를 입력할 때 입력된 태그가 속할 수 있는 카테고리를 미리 정의된 용어집 없이 자유롭게 정의할 수 있는 카테고리형 태깅 방식을 제안하고자 한다. 또한 사용자들로부터 수집된 카테고리형 태그들을 이용하여 다수의 사용자가 생각하는 태그의 관계를 구조화된 폭소노미에 반영하는 기술을 언급하고자 한다. 끝으로 본 연구에서 제시한 방법들이 조직의 지식 분류와 검색에 활용될 수 있음을 증명하기 위해, 카테고리형 태그 기반의 지식조직체계를 시스템으로 구현하였다.

Keywords : 폭소노미, 태그, 웹2.0, 집단지성, 지식조직체계

* 육군사관학교 이공학처 전자정보학과
** 한밭대학교 경상대학 경영학과
*** 고려대학교 경영대학 경영학과

저 자 소 개

Donghee Yoo
Is an assistant professor of Department of Electronics Engineering and Information Science at Korea Military Academy (KMA), South Korea. He received his Ph.D. degree in MS/IS from the Korea University Business School in 2009, and his Bachelor of Business Administration degree in MIS from Korea University in 2002. His current research interests include the semantic web, ontology, web 2.0, linked data, enterprise architecture, decision support system, recommendation system, and army tactical command information system.

Gunwoo Kim
Is an assistant professor in department of Business Administration at Hanbat National University, South Korea. He received his Ph.D. degree in MIS from the Korea University in 2010, his Master degree and Bachelor of Science degree in MIS and computer science from Korea University and Yonsei University, Korea, in 2005 and 2002, respectively. He worked as Information System consultants in several companies from 1996 to 2003. He has published papers on the application of semantic web technology to business domain such as business process management, knowledge management, tourism. His current research interests include use of ontology in several industry areas, industry convergence in management field, collaboration, recommendation system, and social network.

Keunho Choi
Is a Ph.D. candidate of Korea University Business School (KUBS), South Korea. He received his M.S. in Management Information Systems from Inha University in 2008, and BSs in Material Engineering and Techno Business from Inha University in 2006. His current research interests include recommendation systems, data mining, and semantic web.

Yongmoo Suh

Is a professor of Korea University Business School, South Korea (KUBS). He obtained his Ph.D. from the University of Texas at Austin in 1992, MS in Computer Science from Korea Advanced Institute of Science and Technology in 1980 and BS from Seoul National University in 1978. Before going abroad for Ph.D. study, he worked as a researcher of the Korea Institute of Science and Technology. And prior to joining KUBS, he taught at Sejong University and Konkuk University in Korea. He has published papers on various areas, such as object-oriented systems, collaboration, and data mining. His current research interests include use of ontology in business domain, recommendation, and business process management.