A Triadic Approach of Hierarchical Classes Analysis on Folksonomy Mining

Suk-Hyung Hwang

Div. of Computer and Information Science, SunMoon University, A-San, Chung-Nam, 336-840 Korea

Summary

As the number of tagged data on the Web is increased, there are many needs on the folksonomy-based systems to provide some proper functionalities, such as identifying user interests, recommending relevant resources or constructing ontologies for sharing and reusing tag data, etc. However, some proper data mining approaches to folksonomies are necessary to better understand their characteristics and extract valuable information from folksonomies. In this paper, we propose a triadic approach for mining folksonomies based on hierarchical classes analysis, and demonstrate how triadic elements of folksonomies can be analyzed and mined by applying the proposed approach. We also discuss how the results can be used for better understanding of the characteristics of folksonomies.

Key words:
Hierarchical Classes Analysis, Folksonomy, Ontology, Tagging..

1. Introduction

Many websites, such as Wikipedia, Flickr, Technorati, Del.icio.us, Yahoo!, Bibsonomy and others employ a simple tagging mechanism, where users assign metadata in the form of keywords(i.e., tags) to any web resources(e.g., photos, bookmarks, videos, blog entries, etc.) without relying on any controlled vocabulary. Tags are free-formed and user-generated textual keywords that describe a resource, or certain aspects of that resource, from the perspective of the individual. Tags have recently become popular as a coremeans of annotating, classifying and organizing resources on the Web.

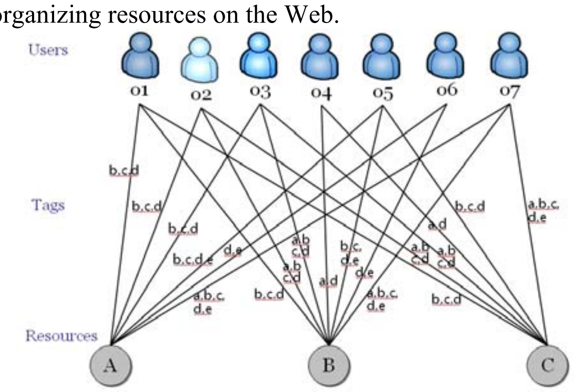


Fig. 1 An example of a Folksonomy

Manuscript received August 5, 2007
Manuscript revised August 20, 2007

Collaborative social tagging systems capture a large collection of diverse web resources that are collaboratively and socially annotated and used by the community of users. In collaborative social tagging systems, tags are used to enable the organization of information within a personal information space, but are also shared, thus allowing the browsing and searching of tags attached to information resources by other users. A folksonomy is a lightweight conceptual structure which is built by a broad community or people(”folks”) in order to categorize various types of Web resource by collaborative social tagging.

As shown in figure 1, folksonomies are constituted by basically three main elements: users, tags, and resources. Resources can be categorized with tag that defines a relationship between the resource and a concept in the user’s mind. Users share various resources under a particular same tag, or share different tags assigned to a group of resource. Tags are linked to each other and so are the resources. Thus three main elements of tagging assignments in folksonomies can be seen as a rich source of data to be mined for semantical and social insights.

In this paper we focus on conceptual analysis of triadic context of folksonomies. We propose a triadic approach of hierarchical classes analysis for mining folksonomy, and present an investigation into extracting and discovering useful information from folksonomies. The organization of the paper is as follows: we introduce the basic notions of folksonomy and some related works on folksonomy models and analysis approaches in Section 2. In Section 3, we present a new approach of conceptual analysis for mining folksonomies based on the hierarchical classes analysis. Finally, we conclude the paper with a summary and ongoing research in section 4.

2. Background

There are many studies already emerging in the folksonomies. We show some folksonomy models and related works on folksonomy analysis.

2.1 Folksonomy models

'Folksonomy' is a blend of keywords 'folk' and 'taxonomy', and stands for a lightweight conceptual structure which is built by a broad community or people in order to categorize various types of Web resource by collaborative social tagging. Folksonomies appear in many web applications, such as Wikipedia, Flickr, Technorati, Yahoo!, YouTube, del.icio.us employ the folksonomy as their social tagging mechanism, where users assign tags to resources and share it with each other within their community[15].

Some well-known formal models of folksonomies are as follows: Guruber[6] proposes a "tag ontology" which formalizes the activity of tagging as a 'tagging relation':*Tagging(object, tag, tagger, source, [+/−])*, with object being the Web resources being tagged, tagger being the user who assigns tags, source being the system from which this annotation originates, and [+/−] representing either a positive or negative vote places on this annotation by the tagger.

On the other hand, Mika[14] represents a folksonomy $F \subseteq A \times T \times O$ as a 'tripartite hypergraph', $G(F) = (V,E)$, where the node set V is the disjunctive union of *actors*(A), *concepts*(T), and *objects*(O) being tagged: $V = A \cup T \cup O$ and $E = \{\{a, t, o\} | (a, t, o) \in F\}$ is the set of hyperedges connecting an actor a who tagged an object o with the tag t.

Recently, Hotho et al.[7, 9] introduce a formal model of folksonomy as a tuple $F := (U, T, R, Y, \lessdot)$, where U, T, and R are finite sets of users, tags, and resources, respectively., and Y is a triadic relation between them, i.e., $Y \subseteq U \times T \times R$, stands for 'tag assignments'. $<$ is a user-specific relation which defines the sub/super relations between tags.

From the above models, we can then understand that there are three main elements that compose any folksonomy-based system:

- The users who actually do the tagging,
- The tags, metadata in the form of user-generated textual keywords,
- The resources being tagged.

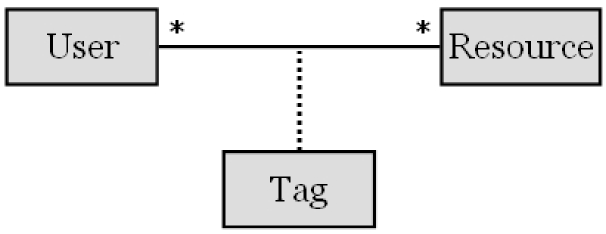


Fig. 2 A conceptual model for a folksonomy

The UML class diagram in Figure 2 represents a conceptual model of folksonomies. Folksonomies allow users to assign tags to resources in order to manage bookmarks, photos, etc. Therefore, tags can be thought of as a bridge between users and resources, with these connections defining relationships among users(several users may use the same tags) and among resources(resources can be tagged with the same words).

2.2 Folksonomy Analysis

There are many studies already emerging in the analysis of tagging and folksonomy. We briefly summarize below some related works including Extracting some characteristics of folksonomies; Analyzing the structure of collaborative tagging systems; Harvesting social knowledge from tags; Deriving ontologies from folksonomies, etc.

Mathes[13] investigates the strengths and limitations of tags in collaborative environment. He claims that despite the chaotic and imprecise nature of folksonomies, the freeform tagging approach works well for users because it allows users to organize information in their own ways and that folksonomies lower the barriers to cooperation and lead to asymmetric communication through tags. Moreover, there have been some attempts of analyzing the structure of collaborative tagging systems and their dynamic aspects[5] and harvesting social knowledge from tags[16]: community identification, expert and document recommendation, and ontology generation.

On the other hand, based on tripartite hypergraph model of a folksonomy, Mika[14] examines two lightweight ontologies(one based on sub-communities of interest and another on object overlaps) on a data set of the deli.cio.us system and reveals broader/narrower relations. The authors concluded that analyzing a lightweight ontology of a subcommunity is a good mean for discovering the emergent semantics of a community. In [4], Ce'line Van Damme et al., proposed a comprehensive approach for deriving ontologies from folksonomies by intergating some related techniques(such as statistical analysis, ontology mapping and matching) and resources(online lexical and semantic web resources). More recently, some significant works on folksonomies have been undertaken by Hotho et al which used a triadic context model of formal concept analysis[1, 11] and then investigated some properties: ranking of contents[7, 8], discovering trends in the tagging behaviour of users[10] and network properties of folksonomies[2].

3. Conceptual Analysis of Folksonomy

In this section, we give a short review of the Hierarchical Classes Analysis[3] and then propose a new approach for triadic context of folksonomy.

3.1 Hierarchical Class Analysis

Hierarchical Class Analysis(HCA)[3] is a set theoreticl clustering technique. It represents the hierarchical structure of the data in a numerical and a graphical way based on a set-theoretical framework. That is, based on set-theoretical relations among objects and attributes, the objects and attributes can be grouped into classes, and then hierarchies of classes can be constructured. In order to grasp the main ideas of the hierarchical class analysis, some definitions are recapitulated in the following:

Definition 1 A dyadic formal context is a triple (**O**, **A**, **R**) that is comprised of a set of objects **O**, a set of attributes **A** and a relation **R**$\subseteq$ **O**×**A** describing which objects possess which attributes.

Table 1 shows an example of a dyadic formal context that is composed of objects **O** = {o1, o2, o3, o4, o5, o6, o7}, attributes **A** = {a, b, c, d, e} and some relations between them.

Table 1: Dyadic formal context

	a	b	c	d	e
o1		×	×	×	
o2	×	×	×	×	
o3	×	×	×	×	
o4	×			×	
o5		×	×	×	×
o6				×	×
o7	×	×	×	×	×

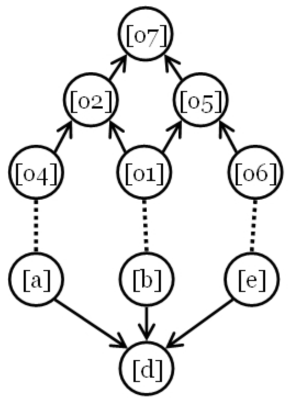


Fig. 3 Object and Attribute hierarchies of the dyadic context of Table 1

For a set $O \subseteq$ **O** of objects and a set $A \subseteq$ **A** of attributes , two derivation operators, intent and extent, are given by:

intent(O) := {a $\in$ **A** | $\forall$o $\in$ O : (o, a) $\in$ **R**},

extent(A) := {o $\in$ **O** | $\forall$a $\in$ A : (o, a) $\in$ **R** }.

Intuitively speaking, **intent**(O) is the set of attributes common to all objects in $O \subseteq$ **O**. Dually, **extent**(A) is the set of objects that have all attributes from $A \subseteq$ **A**. For example, **intent**({o1, o2, o3}) = {b, c, d} and **extent**({b, c, d}) = {o1, o2, o3, o5, o7}.

Definition 2 Given two objects o1, o2$\in$ **O**, o1 and o2 are equivalent if **intent**({o1}) = **intent**({o2}),

o1 $\equiv_O$ o2 $\Leftrightarrow$ **intent**({o1}) = **intent**({o2}).

Correspondingly, the equivalence relation $\equiv_A$ on **A** is defined by:

a1 $\equiv_A$ a2 $\Leftrightarrow$ **extent**({a1}) = **extent**({a2}).

The equivalent objects(attributes) can be grouped together into a same set, which is called an object class(attribute class).

Definition 3 An equivalence class of objects including the object o is defined by

[o] = {x $\in$ **O** | x $\equiv_O$ o}.

Similarly, an equivalence class of attributes including the attribute a is defined by

[a] = {y $\in$ A | y $\equiv_A$ a}.

The hierarchies of the object and attribute classes are constructed based on the order relation as follows:

Definition 4 Given two object classes [o1] and [o2] on **O**, an order relation v between them is defined by:

[o1] $\leq$ [o2] $\Leftrightarrow$ **intent**({o1}) $\subseteq$ **intent**({o2}).

Correspondingly, for two attribute classes [a1] and [a2] on **A**, the order relation $\leq$ between them is defined by:

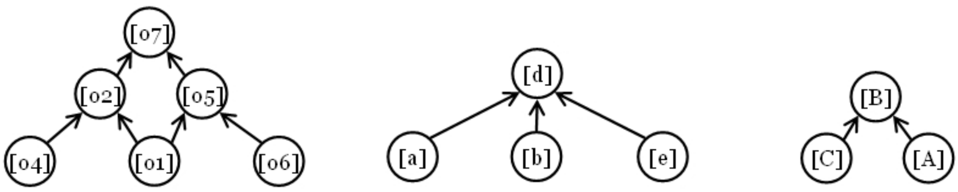
[a1] $\leq$ [a2] $\Leftrightarrow$ **extent**({a1}) $\subseteq$ **extent**({a2}).

[o1]([a1]) is called a sub-class of [o2]([a2]) . Conversely, [o2]([a2]) is called a super-class of [o1]([a1]) . This definition means that one can order the object(attribute) classes by using their associated set of attributes(objects). It is easy to know that if an object class associates with an attribute class, it also associates with all super-classes of the attribute class. Likewise, an attribute class associates with an object class and all its the super-classes.

Based on the order relations between the classes, we can construct two hierarchical structures of the classes(figure 3): Object hierarchy and Attribute hierarchy. The object hierarchy shows up in the upper half of the representation. Equivalent objects are enclosed by boxes, representing the object classes, and the hierarchical relations between the object classes are indicated by lines between the respective boxes. Similarly, the attribute hierarchy shows up upside down in the lower half of the representation. The object and attribute hierarchies are further linked by paths between the leaf classes of the object and attribute hierarchies that are associated with each other.

Table 2 : Triadic context

	A					B					C				
	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e
o1		×	×	×			×	×	×			×	×	×	
o2		×	×	×		×	×	×	×		×	×	×	×	
o3		×	×	×		×	×	×	×		×	×	×	×	
o4						×			×		×			×	
o5		×	×	×	×		×	×	×	×		×	×	×	
o6				×	×				×	×					
o7	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×



(a) (b) (c)

Fig. 4 The hierarchical structures of the users(a), tags(b) and resources classes(c)

3.2 Concept Analysis for Triadic context of Folksonomy

Folksonomies allow users to assign tags to resources in order to manage and classify the resources. These resources can be URLs, photos, movies, blog entries or just about anything else on the web. There are three main elements(users, tags and resources) that compose any folksonomy-based system. To express the notion of folksonomies, we use the triadic context[12].

Definition 5(**Triadic Context**) A triadic context K is a quadruple (K_1, K_2, K_3, Y) that is comprised of three sets K_1, K_2, K_3 and a ternary relation Y between them, i.e., $Y \subseteq K_1 \times K_2 \times K_3$. The elements of K_1, K_2, and K_3 are called *objects*, *attributes*, and *conditions*, respectively. $(x, y, z) \in Y$ is read as "*the object* x *has the attribute* y *under (or according to) the condition* z".

Table 2 shows an example of a triadic context for a folksonomy in figure 1.

We now introduce some notation: For {i, j, k} = {1, 2, 3} with j < k and given an element $a_i \in K_i$, the associated set $M_{(i)}(a_i)$ is defined by

$$M^{(i)}(a_i) := \{(a_j, a_k) \in K_j \times K_k \mid (a_i, a_j, a_k) \in Y\}.$$

For example, for $o_2 \in K_1$, $a \in K_2$, $C \in K_3$ in Table 2,

$M^{(1)}(o_2)$ = {(b,A), (c,A), (d,A), (a,B), (b,B), (c,B), (d,B), (a,C), (b,C), (c,C), (d,C)}

$M^{(2)}(a)$ = {(o_7,A), (o_2,B), (o_3,B), (o_4,B), (o_7,B), (o_2,C), (o_3,C), (o_4,C), (o_7,C)}

$M^{(3)}(C)$ = {(o_1, b), (o_1, c), (o_1, d), (o_2, a), (o_2, b), (o_2, c), (o_2, d), (o_3, a), (o_3, b), (o_3, c), (o_3, d), (o_4, a), (o_4, d), (o_5, b), (o_5, c), (o_5, d), (o_7, a), (o_7, b), (o_7, c), (o_7, d), (o_7, e)}.

Definition 6 For each $i \in \{1, 2, 3\}$, and given two elements x, x' $\in K_i$, x and x' are equivalent if $M^{(i)}(x) = M^{(i)}(x')$, that is

$$x \equiv^i x' \Leftrightarrow M^{(i)}(x) = M^{(i)}(x').$$

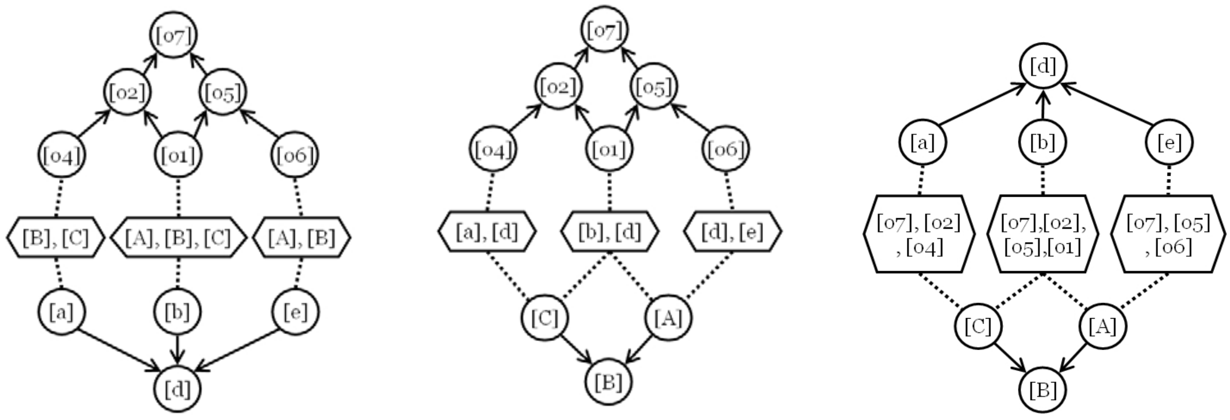
In the example in Table 2, o_2 and o_3 are equivalent($o_2 \equiv^1 o_3$):

$M^{(1)}(o_2) = M^{(1)}(o_3)$
= {(b,A), (c,A), (d,A), (a,B), (b,B), (c,B), (d,B), (a,C), (b,C), (c,C), (d,C)}.

The equivalent objects(attributes, conditions) can be grouped together into a set, called an object(attribute, condition) class as follows:

Definition 7 For each $i \in \{1, 2, 3\}$, an equivalent class including the element $x \in K_i$ is defined by

$$[x]_i := \{x' \in K_i \mid x \equiv^i x'\}$$
$$:= \{x' \in K_i \mid M^{(i)}(x) = M^{(i)}(x')\}.$$



(a) (b) (c)

Fig. 5 Triadic class hierarchies of Table 2

For example,

$[o1]_1 = \{o1\}$, $[a]_2 = \{a\}$,
$[o2]_1 = [o3]_1 = \{o2, o3\}$, $[b]_2 = [c]_2 = \{b, c\}$,
$[o4]_1 = \{o4\}$, $[d]_2 = \{d\}$,
$[o5]_1 = \{o5\}$, $[e]_2 = \{e\}$,
$[o6]_1 = \{o6\}$, $[A]_3 = \{A\}$,
$[o7]_1 = \{o7\}$, $[B]_3 = \{B\}$,
$[C]_3 = \{C\}$.

The sets of objects K_1, attributes K_2 and conditions K_3 can be expressed by the unions of object, attribute and condition classes, respectively. We have,

$K_1 = \{[o1]_1, [o2]_1, [o4]_1, [o5]_1, [o6]_1, [o7]_1\}$,
$K_2 = \{[a]_2, [b]_2, [d]_2, [e]_2\}$,
$K_3 = \{[A]_3, [B]_3, [C]_3\}$.

Furthermore, the order relation between the object or attribute or condition classes can be defined based on the set inclusion relation.

Definition 8 For each $i \in \{1, 2, 3\}$, given two classes $[x]_i$ and $[x']_i$ on K_i, an order relation $\leq$ between them is defined by:

$$[x]_i \leq [x']_i \Leftrightarrow M^{(i)}(x) \subseteq M^{(i)}(x').$$

For each $i \in \{1, 2, 3\}$, a class $[x]_i$ is a subclass of class $[x']_i$ iff the associated set of x is a subset of the associated set of x'. $[x]_i$ is called a subclass of $[x']_i$ and, $[x']_i$ is called a superclass of $[x]_i$. Based on the order relation between the object or attribute or condition classes, three hierarchical structures can be organized from the triadic context(figure 3).

The hierarchies of the object, attribute and condition classes can be integrated into a "Triadic class hierarchy" based on the order relations. There exist connections between object classes and attribute classes by some paths that go via condition classes. Furthermore, if an object class associates with an attribute class, it also associates with all superclasses of the attribute class. That is, an object class should associates with a union of attribute classes via condition classes. In the same way, an attribute class associates with a union of object classes and all its the superclasses via condition classes.

Figure 5 shows some alternative triadic class hierarchies of table 3. The triadic class hierarchy (a) is associated with two class hierarchies(object and attribute class hierarchy) via condition classes. In the object(attribute) class hierarchy, each node denotes object(attribute) class, and each link represents an order relation between object(attribute) classes. The hexagonal nodes with dot lines indicate connections between object classes and attribute classes via some condition classes. The association relation combined in the dual class hierarchy can be read as follows: An object $x \in [x]$ is associated with attribute $y \in [y]$ by condition $z \in [z]$ iff object class $[x]$ and attribute class $[y]$ are connected with each other by a path that goes via condition class $[z]$. For example, in figure 4 (a), there exists a path from object class [o2] to attribute class [a] via [B], [C] and hence, objects {o2, o3} are associated with attribute a by condition B and C.

4. Conclusion

With the growing use of tagging on the Web resources, there are many needs on the folksonomy-based systems to provide some proper functionalities, such as identifying user interests, recommending relevant resources or constructing ontologies for sharing and reusing tag data, etc. By analyzing the distributions of how users apply tags, how tags are applied to links, and how users pick content, we should be able to gain better insights and useful information from the folksonomies.

We presented a triadic approach for mining folksonomies based on hierarchical class analysis. More precisely, we build triadic class hierarchies from a triadic context representing the folksonomy. In order to deal with the complexity of the intrinsic triadic relations among users, tags and resources, the triadic equivalent class and an order relation between them are defined. Based on the order relation between the object or attribute or condition classes, three hierarchical structures can be organized from the triadic context. The hierarchies of the object, attribute and condition classes can be integrated into a Triadic class hierarchy.

We are under development with a prototype implementation of the proposed approach that is helpful to analyze important characteristics on the triadic relationships of the folksonomies. Also, our approach has a wide range of applications and potentialities to explore, classsify and organize the folksonomies more efficiently.

References

[1] R.W. Bernhard Ganter, Formal Concept Analysis: Mathematical Foundations. Springer-Verlag, 1999.

[2] C. Catutto, C. Schmitz, A. Baldassarri, V. D. P. Servedio, V. Loreto, , A. Hotho, M. Grahl, and G. Stumme. Network properties of folksonomies. AI Communications Journal, Special Issue on "Network Analysis in Natural Sciences and Engineering", 2007.

[3] Y.H. Chen and Y. Y. Yao. Formal concept analysis based on hierarchical class analysis. In Proceedings of the 4th IEEE International Conference on Cognitive Informatics(ICCI05), pp.285-292, 2005.

[4] C. V. Damme, M. Hepp, and K. Siorpaes. Folksontology: An integrated approach for turning folksonomies into ontologies. In Bridging the Gap between Semantic Web and Web2.0(SemNet 2007), pages 57-70, 2007.

[5] S. Golder and B. A. Huberman. The structure of collaborative tagging systems. Journal of Information Sciences, 32(2):198–208, April 2006.

[6] T. Gruber. Ontology of folksonomy, 2005.

[7] A. Hotho, R. J¨achke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In Proceedings of the 3rd European Semantic Web Conference, pages 411–426, Budva, Montenegro, June 2006.

[8] A. Hotho, R. J¨aschke, C. Schmitz, and G. Stumme. Folkrank: A ranking algorithm for folksonomies. In K.-D. Althoff and M. Schaaf, editors, LWA, volume 1/2006 of Hildesheimer Informatik-Berichte, pages 111–114. University of Hildesheim, Institute of Computer Science, 2006.

[9] A. Hotho, R. J¨aschke, C. Schmitz, and G. Stumme. Trend detection in folksonomies. In Y. S. Avrithis, Y. Kompatsiaris, S. Staab, and N. E. O'Connor, editors, Proc. First International Conference on Semantics And Digital Media Technology (SAMT), volume 4306 of LNCS, pages 56–70, Heidelberg, dec 2006. Springer.

[10] A. Hotho, R. J¨aschke, C. Schmitz, and G. Stumme. Trend detection in folksonomies. In Y. S. Avrithis, Y. Kompatsiaris, S. Staab, and N. E. O'Connor, editors, Proc. First International Conference on Semantics And Digital Media Technology (SAMT), volume 4306 of LNCS, pages 56–70, Heidelberg, dec 2006. Springer.

[11] F. Lehmann and R. Wille. A triadic approach to formal concept analysis. In G. Ellis, R. Levinson, W. Rich, and J. F. Sowa, editors, Conceptual structures: applications, implementation and theory, number 954 in Lecture Notes in Artificial Intelligence, pages 32–43, Berlin–Heidelberg–New York, 1995. Springer–Verlag.

[12] F. Lehmann and R. Wille. A triadic approach to formal concept analysis. In G. Ellis, R. Levinson, W. Rich, and J. F. Sowa, editors, Conceptual structures: applications, implementation and theory, volume 954 of Lecture Notes in Arti-ficial Intelligence, pages 32–43. Springer Verlag, 1995.

[13] A. Mathes. Folksonomies – cooperative classification and communication through shared metadata. http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html, December 2004.

[14] P. Mika. Ontologies are us: A unified model of social networks and semantics. In Proceedings of the 4th International Semantic Web Conference, volume 3729 of LNCS, pages 522–536. Springer-Verlag, 2005.

[15] T. V. Wal. Folksonomy, 2007. http://vanderwal.net/folksonomy.html.

[16] H. Wu, M. Zubair, and K. Maly. Harvesting social knowledge from folksonomies. In HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia, pages 111–114, New York, NY, USA, 2006. ACM Press.

Suk-Hyung Hwang received the B.S degrees in Computer Science from Kangwon National University, Korea in 1991. He received the M.S. and PhD. degrees in Information and Computer Science from Osaka University, Japan, 1994 and 1997, respectively. He is an Associate Professor of Computer and Information Science at SunMoon University, Korea. His research interests include Object-Oriented Analysis and Design, Ontology, Semantic Web, and Formal Concept Analysis, etc.