World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:3, No:5, 2009

FCA-based conceptual knowledge discovery in Folksonomy

Yu-Kyung Kang, Suk-Hyung Hwang, Kyoung-Mo Yang

***Abstract*— The tagging data of (users, tags and resources) constitutes a folksonomy that is the user-driven and bottom-up approach to organizing and classifying information on the Web. Tagging data stored in the folksonomy include a lot of very useful information and knowledge. However, appropriate approach for analyzing tagging data and discovering hidden knowledge from them still remains one of the main problems on the folksonomy mining researches. In this paper, we have proposed a folksonomy data mining approach based on FCA for discovering hidden knowledge easily from folksonomy. Also we have demonstrated how our proposed approach can be applied in the collaborative tagging system through our experiment. Our proposed approach can be applied to some interesting areas such as social network analysis, semantic web mining and so on.

***Keywords*—** Folksonomy Data Mining, Formal Concept Analysis, Collaborative tagging, Conceptual Knowledge Discovery, Classification.

I. INTRODUCTION

The best feature of the Web2.0 is that users participate in information production voluntarily and share them. Tagging is a basic method for activation of "Participation and Share" that are the core characteristics of Web2.0. It is that many web-users assign one or more descriptive keywords or tags freely to various resources, such as bookmarks, photos and videos, published on the web.

Recently, collaborative tagging has grown in popularity on the web. Collaborative tagging is a collaborative form of this process. Most collaborative tagging systems like Flickr, del.icio.us, YouTube, BibSonomy and others classify and organize resources with bottom-up approach based on such tagging data and provide various services. Flickr is a famous photo sharing system. Each user can attaches tags to photos and shares them. Also, the user can access and tag photos of other users. del.icio.us is a most popular social bookmarking system for storing, sharing, and discovering web bookmarks. In del.icio.us, users can tag each of their bookmarks with freely chosen keywords. YouTube is a video sharing website where users can upload, view and share video clips. BibSonomy is a social bookmarking and publication-sharing system. The tagging data of (users, tags and resources) constitutes a folksonomy that is the user-driven and bottom-up approach to organizing and classifying information. Tagging data stored in the folksonomy include a lot of very useful information and knowledge.

Many researchers[1-4] have been trying to mine useful knowledge out of the tagging data on the collaborative tagging systems. Ching-man Au Yeung et al.[1] proposed cluster analysis approach for understanding the semantics of ambiguous tags in folksonomies based on bipartite graph. Scott A. Golder et al.[2] suggested a method for extracting usage patterns from collaborative tagging systems. Suk-Hyung Hwang et al.[3] proposed an approach for applying hierarchical classes analysis to triadic data for extracting and discovering useful information from folksonomies. Christoph Schmitz et al.[4] presented a method for converting a folksonomy onto a two-dimensional structure and for mining association rules from folksonomy projected into two-dimensional structure.

However, appropriate approach for analyzing tagging data and discovering hidden knowledge from them still remains one of the main problems on the folksonomy mining researches. In this paper, we propose a new FCA-based approach for extracting useful hidden knowledge from tagging data of the collaborative tagging systems. Moreover, we show some experiments that demonstrate how our approach can be applied in web mining. Our research results would be helpful for clustering and classifying the tagging data.

This paper is organized as follows: we introduce some basic notions of the Formal Concept Analysis in Section II. In Section III, we introduce our approach and describe an experiment that applied our approach to tagging data. In the last section IV, we give some conclusions with a summary and our future directions of this research.

II. FORMAL CONCEPT ANALYSIS

In this section, we explain basic notions for understanding Formal Concept Analysis. Formal Concept Analysis(FCA)[5] is a method of data analysis which identifies conceptual structures among data sets. It extracts concepts from various data in given domain, grasps super-sub relations between concepts, and constructs conceptual hierarchy. FCA has been applied to various domains, such as medicine, bioinformatics, social science, ontology engineering, information science, software reengineering, and others[6,7].

Based on the following definitions, FCA classifies data based on the ordinary set into concept units which consists of objects and attributes that those objects have commonly.

Yu-Kyung Kang is with the Dept. of Computer Science & Engineering, SunMoon University, Korea (corresponding author to provide phone: +82-41-530-2265; fax:+82-41-530-2876; e-mail: yukyung.kang@gmail.com).

Suk-Hyung Hwang is Dept. of Computer Science & Engineering, SunMoon University, Korea (e-mail: shwang@sunmoon.ac.kr).

Kyoung-Mo Yang is with Dept. of Computer Science & Engineering, SunMoon University, Korea (e-mail: yjj0309@gmail.com).

Definition 1. A Formal Context K := (O, A, R) consists of two finite nonempty sets O and A and a relationship R between O and A. O is a set of objects and A is a set of attributes, and R ⊆ O X A is a binary relation between O and A. In order to express that an object o is in a relation R with an attribute a, we write (o, a) ∈ R and read it as "the object o has the attribute a".

TABLE I
Formal context

	a	b	c	d	e
O1	X		X	X	X
O2	X		X		
O3		X			X
O4		X			X
O5	X				
O6	X	X			X

TABLE I shows an example of formal context based on O = {O1, O2, O3, O4, O5, O6} and A = {a, b, c, d, e} and the binary relation R is represented as "X" in the cross table.

Definition 2. Let (O, A, R) be a context, X⊆O and Y⊆A. The function intent maps a set of objects into the set of attributes common to the objects in X (intent : $2^X \rightarrow 2^Y$), whereas extent is the dual for an attributes set (extent : $2^Y \rightarrow 2^X$) :

$$intent(X) := \{a \in A \mid \forall o \in X : (o, a) \in R\},$$
$$extent(Y) := \{o \in O \mid \forall a \in Y : (o, a) \in R\}.$$

Definintion 3. Let (O, A, R) be a context. A formal concept is a pair (X, Y) with X⊆O is called extension, Y⊆A is called intension, and

$$(X = extent(Y)) \wedge (Y = intent(X)).$$

We can extract every concept in a context by above Galois connection. The set of all concepts B(X) are presented at TABLE II and it is defined as follows:

$$B(K)=\{(X, Y) \in 2^O \text{ X } 2^A \mid intent(X) = Y \wedge extent(Y) = X\}.$$

TABLE II
All formal concepts for TABLE I

ID	Extents	Intents
C_1	{ }	{a, b, c, d, e}
C_2	{O1}	{a, c, d, e}
C_3	{O6}	{a, b, e}
C_4	{O1, O2}	{a, c}
C_5	{O1, O6}	{a, e}
C_6	{O3, O4, O6}	{b, e}
C_7	{O1, O2, O5, O6}	{a}
C_8	{O1, O3, O4, O6}	{e}
C_9	{O1, O2, O3, O4, O5, O6}	{ }

Definition 4. For a Formal Context K = (O, A, R) and two Concepts C_1 = (O_1, A_1), C_2 = (O_2, A_2) ∈ B(K) the subconcept-superconcept relation is given by:

$$(O_1, A_1) \leq (O_2, A_2) \Leftrightarrow O_1 \subseteq O_2 \ (\Leftrightarrow A_1 \supseteq A_2).$$

Concept lattice L := (B(K), ≤) can be obtained by all formal concepts of a context K with the subconcept-superconcept relation. Its graphical representation is line (or Hasse) diagram. Fig. 1 is example of line diagram for TABLE I. Each circle denotes concept. The labels of circle represent a pair of extent and intent of the concept. The number represented in the node is ID of the formal concept. Whole intents are obtained by upward path of concepts and extents are obtained by downward path.

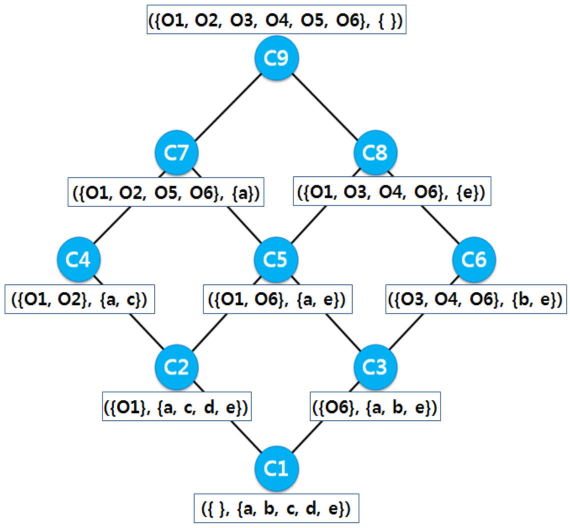

Fig. 1. Formal Concept Lattice for TABLE I

For example, in Fig. 1, C_5 is composed of a pair of objects O1 and O6 spread from subconcept and attributes a and e inherited from superconcept.

III. Discovering Conceptual Knowledge form Folksonomy

In this section, we describe our approach for discovering useful hidden knowledge from folksonomy data. The folksonomy data are analyzed based on FCA mentioned above. The process to analyze is described with 4 steps as follows(Fig. 2):

Step 1. Building formal context for a given seed

We can choose a set of user or tag or resource with a seed. We make a formal context composed of other two elements related with the seed from folksonomy data.

Step 2. Applying FCA

The formal concepts are extracted by applying FCA to the context made in step 1. Each concept is a cluster, which consist of objects and attributes that the objects have commonly. The formal concept lattice is constructed to grasp the subconcept-superconcept relationships between the concepts.

Step 3. Building formal context for a selected concept

We select a concept in the concept lattice and choose a seed on objects or attributes of the selected concept. Similar with step 1, we create the formal context based on tagging data related with the seed in given folksonomy.

Step 4. Apply FCA and Extracting Knowledge

The formal concept lattice for the concept selected in step 3 is constructed with FCA application to the formal context created in step 3. We can extract hidden useful knowledge from the concept lattice based on the information that user group is composed of tags used by the users commonly and resources attached common tags.

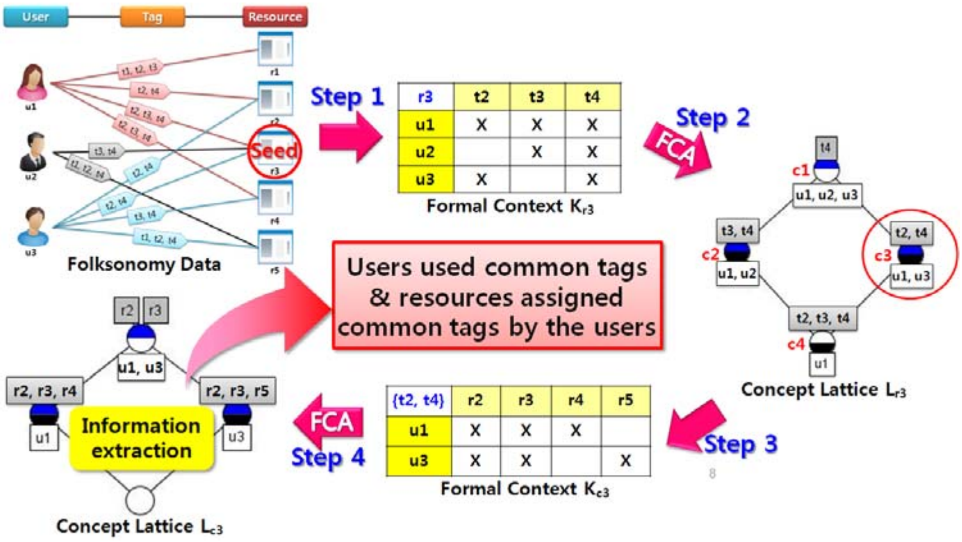

Fig. 2 FCA-based folksonomy data mining process

TABLE III
Formal Context gathered from BibSonomy for "web2.0" and "tagging" tags

users \ resources	r0	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	...	r44	r45
kochm														X					
pecim7af										X									
ewm								X											
yish									X										
jboy701							X												
dbenz															X				
teresina					X			X			X	X					...		X
lysander07	X		X	X		X	X									X		X	
domenico79							X						X						
michael					X														
iamjason		X								X									
schmitz									X				X						
beate							X						X						
...	...																		
stumme															X				

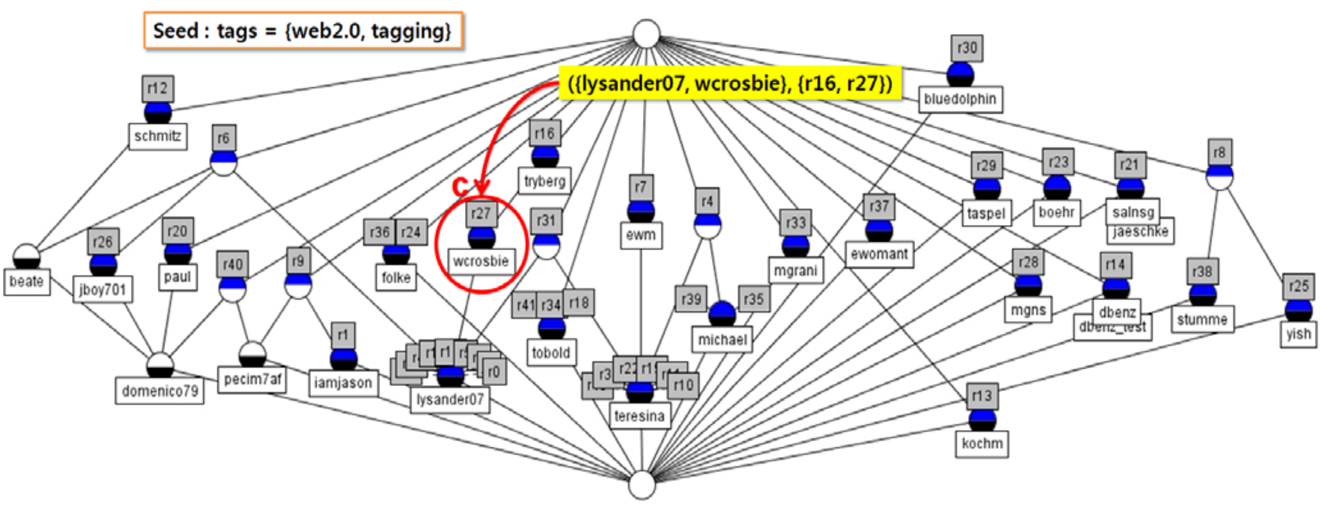

Fig. 3 Formal Concept Lattice for TABLE III

TABLE IV
Formal Context for a concept C of Fig. 3

tags \ resources	r0	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10
2005											X
history				X							
web2.0		X			X						
semantic						X					
foaf			X								
wikipedia					X						
eswc2006							X				
flickr		X									
semwiki2006						X		X	X		
memex				X							
semanticweb					X						
tagging	X	X			X					X	

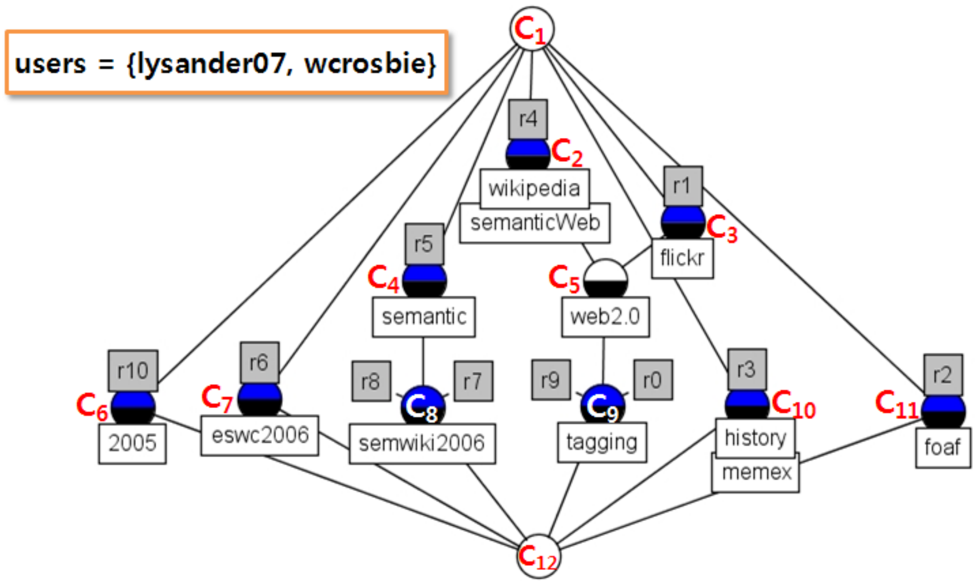

Fig. 4 Formal Concept Lattice for TABLE IV

In order to show the potentiality and usefulness of our proposed approach, we experiment with some tagging data captured from collaborative tagging systems and explain the result of the experimant. For this experiment, we gathered the tagging data of (users, tags and resources) from BibSonomy, which is one of the famous social bookmarking and publication -sharing system. We selected “web2.0” and “tagging” tags as a seed. And, based on the seed, we collected folksonomy data of users and resources that are related with the seed(i.e. tags). That is, given the tags, the data table of users and resources(TABLE III) is gathered from BibSonomy. For “web2.0” and “tagging” tags, the context is composed of 28 users, 46 resources, and relations between them. We extracted 34 formal concepts and relationships among concepts by apply FCA to TABLE III and constructed a formal concept lattice like Fig. 3.

In the formal concept lattice of Fig. 3, we chose a concept C, which is composed of 2 users(lysander07 and wcrosbie) and 2 resources(r16 and r27). We again created a formal context about the concept C in TABLE IV with 12 tags and 10 resources and applied FCA to the context. We extracted 23 concepts and subconcept-superconcept relations and built the formal concept lattice of Fig. 4. The concept lattice indicates information about 12 tags used by users and 10 resources attached such the tags, commonly. We can discover useful information as follows: For example, from the concept C_5 of Fig. 4. we can extract that two users lysander07 and wcrosbie assigned tags “web2.0” and “tagging” to resources r1 and r2, commonly. Also, two users lysander07 and wcrosbie attached tags “Wikipedia” and “semanticWeb” as well as “web2.0” and “tagging” to a resources r4, commonly. This shows that, by analyze tagging data using proposed our approach, we extract useful hidden knowledge from folksonomy easily.

IV. CONCLUSIONS

In this paper, we have proposed a folksonomy data mining for discovering hidden knowledge easily from folksonomy.

Also we have demonstrated how our proposed approach can be applied in the collaborative tagging system through our experiment. From our experiment result, we can confirm that the proposed approach is useful and has some potentialities on clustering for tagging data. Our proposed approach can be applied to some interesting areas such as social network analysis, semantic web mining and so on. In particular, it can be applied to analyze web users' tagging patterns and to web search engine based on new classification

As a future work, in order to apply our approach widely, we will develop a supporting tool for folksonomy data mining. Also, we will do experiments for discovering hidden knowledgs from various collaborative tagging systems.

REFERENCES

[1] Ching-man Au Yeung, Nicholas Gibbins, and Nigel Shadbolt, "Understanding the Semantics of Ambiguous Tags in Folksonomies", The International Workshop on Emergent Semantics and Ontology Evolution(ESOE2007) at ISWC/ASWC 2007, pp. 108~121, 2007.
[2] Scott A. Golder and Bernardo A. Huberman, " Usage patterns of collaborative tagging systems ", Journal of Information Science, 32 (2), pp. 198–208, 2006.
[3] Suk-Hyung Hwang, Yu-Kyung Kang, "Applying Hierarchical Classes Analysis to Triadic context for Folksonomy Mining", 2007 International Conference on Convergence Information Technology (ICCIT'07), pp.103-109, 2007.
[4] Christoph Schmitz, Andreas Hotho, Robert J¨aschke, Gerd Stumme, "Mining Association Rules in Folksonomies", In: Data Science and Classification: Proc. of the 10th IFCS Conf.Berlin, Heidelberg: Springer (2006) , p. 261--270..
[5] B. Ganter, R. Wille, *Formal Concept Analysis : Mathematical Foundations.*, Springer, 1999.D
[6] C. Carpineto, G. Romano, *Concept Data Analysis: Theory and Applications.*, Wiley, September, 2004.
[7] B. A. Davey, H. A. Priestley, *Introduction to Lattices and Order.*, Cambridge University Press, 2002.
[8] D. Doubois, W. Ostasiewicz and H. Prade, Fuzzy sets: History and Basic Notions., Kluwer Academic Boston, 1999.
[9] B. Walczak and D. L. Massart, "Tutorial Rough sets theory", *Chemometrics and Intelligent Laboratory Systems.*, vol.47, pp. 1-16, 1999.

Yu-Kyung Kang received B.S degrees in computer and information science from SunMoon University, Korea in 2003. She received the M.S. and PhD. degrees in computer science and engineering from SunMoon University, Korea, 2005 and 2009, respectively. Her research interests include Formal Concept Analysis, Semantic Web Mining, Semantic Web and Software Engineering, etc.

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 the Department of Computer Science & Engineering, SunMoon University, Korea. His research interests include Object-Oriented Analysis and Design, Ontology, Semantic Web, and Formal Concept Analysis, etc.

Kyoung-Mo Yang received B.S degrees in computer and information science from SunMoon University, Korea in 2006. Currently, he is a M.S. candidate in the Department of Computer Science & Engineering, SunMoon University, Korea. His research interests include Data Mining, Semantic Web and Ontology Engineering.