

WordNet기반의 형식개념분석기법을 이용한 사용자태그 분류체계의 구축

황석형*

Construction of Hierarchical Classification of User Tags using WordNet-based Formal Concept Analysis

Suk-Hyung Hwang*

요약

본 논문에서는, 어휘정보를 체계화한 워드넷(WordNet)을 토대로 형식개념분석기법(Formal Concept Analysis)을 사용하여 폭소노미 사용자들의 태그데이터를 분석하여 사용자태그의 분류체계를 구축하기 위한 기법을 제안하였다. 또한, 제안된 기법을 지원하기 위한 도구(TagLighter)의 개발에 대하여 설명하고, 그 유용성을 확인하기 위하여 Bibsonomy.org의 사용자태그데이터를 대상으로 실시한 실험결과를 기술하였다. TagLighter에 의해 구축된 사용자태그 분류계층구조는 사용자태그에 대한 보다 상세한 이해를 제공하기 때문에, 폭소노미기반 시스템에서의 정보검색과 데이터분석에 유용하다. 본 논문의 연구결과는, 폭소노미기반의 웹서비스와 소셜네트워크시스템, 시맨틱웹 어플리케이션 등을 대상으로 하는 웹데이터 마이닝분야에 응용할 수 있다.

▶ Keywords : 폭소노미, 사용자태그, 워드넷, 형식개념분석, 분류체계

Abstract

In this paper, we propose a novel approach to construction of classification hierarchies for user tags of folksonomies, using WordNet-based Formal Concept Analysis tool, called TagLighter, which is developed on this research. Finally, to give evidence of the usefulness of this approach in practice, we describe some experiments on user tag data of Bibsonomy.org site. The classification hierarchies of user tags constructed by our approach allow us to gain a better and further understanding and insight in tagged data during information retrieval and data analysis on the folksonomy-based systems. We expect that the proposed approach can be used in the fields of web

•제1저자 : 황석형 •교신저자 : 황석형

•투고일 : 2013. 8. 2, 심사일 : 2013. 8. 22, 게재확정일 : 2013. 9. 6.

* 선문대학교 컴퓨터공학과(Dept. of Computer Science and Engineering, SunMoon University)

* 이 연구는 2013학년도 선문대학교 교내학술연구비 지원에 의하여 이루어졌음.

data mining for folksonomy-based web services, social networking systems and semantic web applications.

▶ Keywords : Folksonomy, User Tag, WordNet, Formal Concept Analysis, Classification

I. 서론

첨단 정보통신인프라와 TGIF(Twitter, Google, iPhone, Facebook)으로 대표되는 스마트기기, 그리고 소셜네트워크서비스(SNS) 등과 같은 소셜미디어의 급격한 발전을 토대로, 최근의 웹환경은 사용자들의 참여와 정보공유, 협력이 중심되는 웹2.0세대로 진화발전하면서, 웹자원의 분류체계는 과거 소수의 전문가집단에 의한 톱다운 방식의 택소노미(Taxonomy)로부터 다수의 대중 참여에 의한 바텀업방식의 폭소노미(Folksonomy)로 변화하고 있다[1-3]. 폭소노미는 사용자들에 의해 자유롭게 선택된 어휘(태그)를 이용하여 웹자원을 효율적으로 분류하고 체계화하기 위한 협력형 정보분류체계로서, 이미지공유사이트Flickr.com, 북마킹공유사이트Del.icio.us, 연구관련논문 및 북마킹 공유사이트www. Bibsonomy.org 등이 대표적인 폭소노미기반시스템이다.

폭소노미는 택소노미에 비해서 체계적이지 않고 무작위로 생성되지만, 사용자들의 자발적인 참여에 의해 웹자원에 의미가 부여된 태그데이터(tagged data)를 공유함으로써 사용자 관점에서 편리하고 자유롭게 웹자원을 분류하고 체계화할 수 있다. 따라서 개인과 단체는 물론 정부와 기업에게 새로운 가치를 창출할 수 있는 자원으로 인식되어, 폭소노미기반 웹자원의 분류체계 구축에 관한 다양한 연구[4-11]가 수행되었다. 그러나 최근 폭소노미기반 애플리케이션의 급격한 증가에도 불구하고, 태깅에 사용되는 통제되지 않은 어휘들이 갖고 있는 모호성, 다의어(polysemy) 및 동형어의어(synonymy)의 문제, 그리고 정제된 어휘계층구조의 부재 등으로 인해 웹자원의 검색 및 관련정보의 브라우징에 관한 품질이 저하될 수 있는 등의 문제점이 있다[12,13].

본 논문에서는 폭소노미에 축적되어 있는 웹자원의 태그데이터에 대하여 어휘정보체계를 참조하여 개념적으로 분석하여 사용자태그데이터의 분류체계를 구축하기 위한 기법을 제안한다. 구체적으로 본 연구에서는, 어휘의 의미와 어휘들 사이의 관계를 토대로 어휘정보를 체계화한 워드넷(WordNet)[14]을 토대로 형식개념분석기법(Formal

Concept Analysis)[15]을 사용하여 폭소노미 사용자들의 태그데이터를 분석하여 계층구조를 구축하기 위한 기법을 제안한다. 또한, 지원도구(TagLighter)를 개발하였으며, 제안한 기법의 유용성을 검토하기 위하여 연구관련 논문 및 북마킹 공유사이트 Bibsonomy.org의 사용자태그데이터를 대상으로 실시한 실험결과를 기술하였다.

본 논문의 구성은 다음과 같다. 제2장에서는 관련연구들을 소개하고, 제3장에서는 어휘정보체계를 기반으로 하는 폭소노미 사용자태그분류체계의 구축에 관하여 기술한다. 제4장에서는 지원도구의 개발 및 실험에 대하여 설명하고, 제5장에서는 결론 및 향후 연구 과제를 기술한다.

II. 관련 연구

참고문헌[4]에서는 웹 이미지에 대한 태깅으로 구축된 폭소노미의 효과적인 검색을 위하여 워드넷과 위키피디아의를 기반으로 태그들의 의미적 연관성을 계산하는 방법을 제시하였다. 한편, Stock[5]는 고전적인 전문과학기술데이터베이스(INSPEC, MEDLINE)와 폭소노미를 매쉬업하여 관련성 높은 태그를 추출함으로써 폭소노미의 검색효율을 향상시키기 위한 기법을 제안하였다. Finnsen 등[6]은 도서관분야에 폭소노미를 도입한 사례에 관한 연구로서, 전통적인 분류체계를 사용하고 있는 영국Brunel대학도서관에 대하여, 도서관이용자 참여형 폭소노미기반 도서관정보화시스템을 구축함으로써, 보다 효율적인 도서관분류체계를 구축하고 이용자만족도를 향상시킨 사례를 소개하고 있다. 한편, 참고문헌[7]에서는 워드넷과 태그출현빈도계산 등을 토대로 동형어의어문제를 해결하고 태그품질을 향상시킨 폭소노미기반의 북마킹시스템을 개발하여 30%이상의 태그품질향상과 사용자만족도증가 결과를 얻고 있다. 또한, Kiu와 Tsui[8]는 데이터마이닝 기법을 활용하여 택소노미와 폭소노미를 결합시킨 하이브리드형 분류체계 구축기법을 제안하였고, Yoo 등[9]은 폭소노미가 갖는 의미론적 문제점을 해결하기 위하여 카테고리화된 태그를 사용하여 웹 콘텐츠를 체계화하는 지식체계화 시스템(CTKOS)을 제안하였다. 참고문헌[10]에서는 일반적인 폭

소노미에 그룹개념을 도입한 그룹화된 폭소노미 모델과 관련 기본연산들을 제안하였다. 또한, 참고문헌[11]에서는 폭소노미기반의 소셜태깅기능을 음악추천시스템에 융합한 시스템을 제안하였다.

이상의 관련연구들을 종합해보면, 전통적인 분류체계(택소노미)보다는 사용자참여형 분류체계(폭소노미)가 사용자만족도가 높지만, 폭소노미의 문제점 해결과 분류체계 품질향상을 위하여 표준화된 어휘체계 또는 카테고리화된 태그사용, 택소노미와 폭소노미의 융합 등이 시도되고 있음을 알 수 있다. 따라서 기존의 관련연구들을 토대로, 본 연구에서는 표준화된 어휘정보체계(워드넷)와 정보추출 및 계층화를 위한 분석방법(형식개념분석)을 이용하여, 폭소노미기반 시스템에 축적되어 있는 사용자태그데이터를 효율적으로 체계화하기 위한 새로운 기법을 제안하고자 한다.

III. 사용자태그분류체계의 구축

1. 폭소노미

폭소노미는 웹에 존재하는 다양한 자원에 대하여 사용자들이 자유롭게 선택한 키워드를 부여하는 태깅(tagging)작업에 의해서 자체적으로 만들어가는 협력형 정보분류체계이다. 폭소노미는 빠르게 변하는 다양한 콘텐츠에 대해서 유연한 적응력을 가지고 있어서 보다 효율적이고 효과적인 정보검색 및 공유가 가능하며, 현재 다양한 폭소노미기반 웹서비스가 등장하여 활용되고 있다. 본 논문에서는 연구관련 논문 및 북마킹 공유사이트로서 널리 활용되고 있는 Bibsonomy.org에 대하여, 그림1과 같이 웹자원과 사용자, 그리고 태그에 초점을 맞추도록 한다.

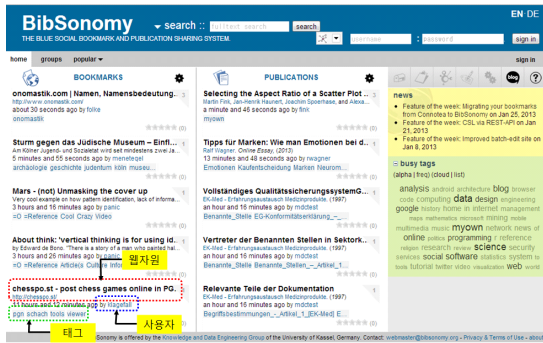


그림 1. 폭소노미기반시스템의 예(Bibsonomy.org)
Fig. 1. Example of Folksonomy-based system(Bibsonomy.org)

다양한 폭소노미기반 웹서비스에 공통적으로 내장되어 있는 폭소노미에 대한 일반적인 정의는 Hotho[16]에 의해 다음과 같이 정형화되었다.

[정의1]폭소노미 $F := (U, T, R, A)$ 는 다음과 같은 요소들로 구성되는 집합체이다.

- U : 태깅을 수행하는 사용자들의 집합,
- T : 태깅에 사용되는 키워드들의 집합,
- R : 태깅대상이 되는 웹자원들의 집합,
- $A \subseteq U \times T \times R$: 사용자들이 웹자원에 대하여 수행한 태깅작업에 의해 형성된 3항관계(⟨사용자, 키워드, 웹자원⟩)들의 집합. ■

그림2는 사용자 $U = \{u_1, u_2, u_3\}$, 태그 $T = \{t_1, t_2, t_3, t_4\}$, 웹자원 $R = \{r_1, r_2, r_3\}$ 으로 구성된 폭소노미의 예를 보여주고 있다. 예를 들어, 사용자 u_1 의 태그집합(T_{u_1})은 $T_{u_1} = \{t_1, t_2\}$ 이고, 웹자원집합(R_{u_1})은 $R_{u_1} = \{r_1\}$ 이다. 또한, 태그 t_1 의 사용자집합(U_{t_1})은 $U_{t_1} = \{u_1, u_2\}$ 이고, 웹자원집합(R_{t_1})은 $R_{t_1} = \{r_1\}$ 이다. 웹자원 r_1 의 사용자집합(U_{r_1})과 태그집합(T_{r_1})은 각각, $U_{r_1} = \{u_1, u_2, u_3\}$, $T_{r_1} = \{t_1, t_2\}$ 이다. 특히, 웹자원 r_1 과 r_2 에는 동시사용태그로서 각각, $\{t_1, t_2\}$ 와 $\{t_3, t_4\}$ 가 태깅되어 있음을 알 수 있다.

폭소노미는 다수의 사용자들이 참여하는 협업적 태깅시스템이기 때문에 어떤 웹리소스에 대하여 태깅할 때 다른 사람의 태그를 참조하거나, 의미를 보다 명확하게 하기 위하여 여러 개의 태그들을 함께 사용해서 태깅할 수 있다. 이와 같이 임의의 웹자원에 대하여 동시에 사용된 태그들을 동시사용태그(co-occurrence tag 또는 co-tag라고 부름)라고 부른다.

[정의2]폭소노미 $F := (U, T, R, A)$ 에서, 임의의 태그들 $\{t_1, t_2, \dots, t_n\} \subseteq T$ 을 동시 사용한 사용자들의 집합 $U_{\{t_1, t_2, \dots, t_n\}}$ 과 동시사용태그 $\{t_1, t_2, \dots, t_n\}$ 가 태깅되어 있는 웹자원들의 집합 $R_{\{t_1, t_2, \dots, t_n\}}$ 을 다음과 같이 정의한다(단, $1 \leq n \leq |T|$).

$$U_{\{t_1, t_2, \dots, t_n\}} := \{u \in U \mid \exists r \in R : (u, \{t_1, t_2, \dots, t_n\}, r) \in A\},$$

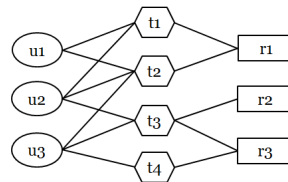


그림 2. 폭소노미의 구성요소들의 관계
Fig. 2. Relationships between Folksonomy elements

$$R_{\{t_1, t_2, \dots, t_n\}} := \{r \in R \mid \exists u \in U : (u, \{t_1, t_2, \dots, t_n\}, r) \in A\}$$

그림2에서 $U_{\{t_1, t_2\}} = \{u_1, u_2\}$, $R_{\{t_1, t_2\}} = \{r_1\}$ 이다. 즉, 태그 t_1 과 t_2 를 동시에 사용한 사용자는 u_1 과 u_2 이고, 이러한 태그들은 웹자원 r_1 에 태그되어 있다. 따라서 사용자 u_1 과 u_2 는 웹자원 r_1 에 대한 지식(태그 t_1 과 t_2)을 공유하고 있다.

2. 워드넷

워드넷(14)은 단순한 어휘의 의미 뿐만 아니라 어휘들 사이의 의미관계정보를 망라하고 있는 사전으로서 자연언어처리와 정보검색 분야에서 널리 이용되고 있다. 워드넷은, 어휘들(W)에 대하여 유사한 의미를 갖는 어휘들로 구성된 동의어집합(synset이라고 부름)들의 집합 $S \subseteq 2^W$ 과 동의어집합들 사이의 다양한 의미관계들(상위어, 동위어, 전체어, 부분어, 양태어, 수반어 등)의 집합 R 에 의해 구성되는 의미연결체계(비순환형 유방향그래프 $WordNet = (S, R)$)로 구성되어 있다. 최신버전의 워드넷(WordNet 3.0)에는 총 155,287개의 어휘들과 117,659개의 개념(synset)들로 구성되어 있다.

그림3은 워드넷에서 “Java”라는 어휘를 검색한 결과(대상 어휘의 의미, 동의어와 상위어를 표시)를 나타내고 있다. 그림3의 결과로부터, “Java”는 총3개의 의미들(Sense 1 : 섬, Sense 2 : 커피, Sense 3 : 프로그래밍언어)을 가지고 있고, “Java”의 상위어들이 포함된 동의어집합(synset)들을 “=”기호를 사용하여 나타내고 있다.

본 논문에서는 워드넷에 포함된 어휘의 83%이상을 차지

```

3 senses of java
Sense 1
Java -- (an island in Indonesia south of Borneo; one of the world's most densely populated regions)
INSTANCE OF= island -- (a land mass (smaller than a continent) that is surrounded by water)
=> land, dry land, earth, ground, solid ground, terra firma -- (the solid part of the earth's surface; "the plane turned away from the sea and moved back over land"; "the earth shook for several minutes"; "he dropped the legs on the ground")
=> object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")
=> physical entity -- (an entity that has physical existence)
=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Sense 2
coffee, java -- (a beverage consisting of an infusion of ground coffee beans; "he ordered a cup of coffee")
=> beverage, drink, drinkable, potable -- (any liquid suitable for drinking; "may I take your beverage order?")
=> food, nutrient -- (any substance that can be metabolized by an organism to give energy and build tissue)
=> substance, matter -- (that which has mass and occupies space; "an atom is the smallest indivisible unit of matter")
=> physical entity -- (an entity that has physical existence)
=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
=> liquid -- (a substance that is liquid at room temperature and pressure)
=> fluid -- (a substance that is fluid at room temperature and pressure)
=> substance, matter -- (that which has mass and occupies space; "an atom is the smallest indivisible unit of matter")
=> physical entity -- (an entity that has physical existence)
=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Sense 3
Java -- (a simple platform-independent object-oriented programming language used for writing applets that are downloaded from the World Wide Web by a client and run on the client's machine)
=> object-oriented programming language, object-oriented programming language -- ((computer science) a programming language that enables the programmer to associate a set of procedures with each type of data structure; "C++ is an object-oriented programming language that is an extension of C")
=> programming language, programming language -- ((computer science) a language designed for programming computers)
=> artificial language -- (a language that is deliberately created for a specific purpose)
=> language, linguistic communication -- (a systematic means of communicating by the use of sounds or conventional symbols; "he taught foreign languages"; "the language introduced is standard throughout the text"; "the speed with which a program can be executed depends on the language in which it is written")
=> communication -- (something that is communicated by or to or between people or groups)
=> abstraction -- (a general concept formed by extracting common features from specific examples)
=> abstract entity -- (an entity that exists only abstractly)
=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
    
```

그림 3. 워드넷을 이용한 어휘검색결과 의 예
Fig. 3. Example of Word search results in WordNet

하고 있는 명사어휘들과 동의어집합 및 어휘들 간의 상하위어 관계(hypernym-hyponym)에 주목한다(이후, 본 논문에서 “임의의 어휘”라 함은 “명사어휘”를 의미한다). 특히, 워드넷에 포함된 명사관련 동의어집합들은 Abstraction, Act, Entity, Event, Group, Phenomenon, Possession, Psychological feature, State와 같은 9가지 개념들로 분류되어 각각을 루트로 하여 적어도 1개의 상하위어 관계들을 포함하는 어휘계층구조를 형성한다. 이와 같은 어휘계층구조는 다음과 같은 재반정의들을 토대로 비순환형 유방향그래프(Directed Acyclic Graph)로 정의할 수 있다.

[정의3]워드넷 $WordNet = (S, R)$ 에 수록된 임의의 어휘 $x, y \in W$ 에 대하여,

- (1) x 가 y 와 같은 의미를 갖는 경우, x 는 y 의 동의어라고 부르고 $x \equiv y$ 로 나타낸다.
- (2) x 가 y 의 일종('x is a y' 또는 'x is a kind of y')인 경우, x 는 y 의 하위어(또는 y 는 x 의 상위어)라고 부르고 $x < y$ 로 나타낸다.

이때, 어휘 $x \in W$ 에 대하여, 동의어들의 집합을 $synset(x) := \{y \in W \mid x \equiv y\}$ 로, 상위어들의 집합을 $hypset(x) := \{y \in W \mid x < y\}$ 로 나타낸다. 특히, 어휘 $x \in W$ 의 동의어집합과 상위어집합이 1개 이상 존재하는 경우에는 $synset_i(x)$ 와 $hypset_i(x)$ 로 나타낸다(단, $1 \leq i \leq n$).

동의어관계(\equiv)는 일종의 동치관계(equivalence relation)로서 반사성($v \equiv v$)과 대칭성($(v \equiv w) \Rightarrow (w \equiv v)$), 추이성($(v \equiv t) \wedge (t \equiv w) \Rightarrow (v \equiv w)$)을 만족하고, 상하위어 관계($<$)는 일종의 반순서관계(partially order relation)로서, 반사성($v < v$)과 반대칭성($(v < w) \Rightarrow (w \not< v)$), 그리고 추이성($(v < t) \wedge (t < w) \Rightarrow (v < w)$)을 만족한다. 또한, 동의어관계와 상하위어관계의 추이적 폐쇄(transitive closure)를 각각 \equiv^* 와 $<^*$ 로 나타낸다. 특히, 동의어관계에 의해 구성되는 동의어집합은 일종의 동치류를 형성하므로, 어휘들 사이의 상하위어관계($<$)는 동의어집합들 사이의 상하위어관계로 확장될 수 있다.

즉, $\forall v, w \in W [(v < w) \Rightarrow (synset(v) < synset(w))]$ 이다. 임의의 어휘 $w \in W$ 에 대하여, 동의어집합들($synset_1(w), \dots, synset_n(w)$)의 전체집합 $SynSet(w)$ 과 상위어집합들의 전체집합 $HypSet(w)$ 을 각각 다음과 같이 나타낸다. 즉,

$$SynSet(w) := \bigcup_{i=1}^n \{synset_i(w) \mid w \in W\},$$

$$HypSet(w) := \bigcup_{\text{for each } s \in SynSet(w)} \{t \in S \mid s <^* t\}.$$

위의 제반 개념들을 토대로, 본 논문에서는 다음과 같이 어휘계층구조를 정의한다.

[정의4] 임의의 어휘들 $ws = \{w_1, w_2, \dots, w_n\}$ 에 대한 어휘 계층구조 $WHG_{ws} := (V, E)$ 는 다음과 같은 요소들로 구성되는 비순환형 유방향그래프(Directed Acyclic Graph)이다.

- $V := \bigcup_{i=1}^n (SynSet(w_i) \cup HypSet(w_i))$:
어휘들 $ws = \{w_1, w_2, \dots, w_n\}$ 의 동의어집합 및 각 동의어집합들의 모든 상위어들에 대한 동의어집합들의 집합
- $E \subseteq V \times V$:
각 동의어집합들사이의 상하위어관계들의 집합 ■

임의의 어휘들 $ws = \{w_1, w_2, \dots, w_n\}$ 에 대한 어휘계층구조 $WHG_{ws} := (V, E)$ 는 알고리즘 **Build_WordHierarchyGraph**에 의해 구축할 수 있다. 그림4는 정의4와 알고리즘 **Build_WordHierarchyGraph**를 사용하여, 주어진 2개 어휘

Function FindHypSet(WordNet , v)

```

1. // 입력 : (1)동의어집합들의 상하위어연결체계를 나타낸 //
2.   워드넷그래프 WordNet := (S, R),
3. //   (2)경로탐색을 위한 시작점 v ∈ S
4. // 출력 : 워드넷의 상하위어관계(<)를 토대로 v부터
5. //   최상위정점까지 깊이우선탐색에 의해 구해지는
6. //   경로 Hpath(v) := <v → v1 → v2 → ... → vk>
7. setLabel(v, VISITED);
8. for each w ∈ S : v < w do
9.   if getLabel(w) ≠ VISITED then
10.    FindHypSet( WordNet , w);
11.  else
12.    Hpath(v).append(v); //Hpath(v)에 정점v를 추가
13.  end if
14. end for
End_FindHypSet.

```

$ws = \{java, software\}$ 에 대한 어휘계층구조 $WHG_{(java, software)}$ 를 구성한 예를 보여주고 있다. 말단정점들은 주어진 어휘 "java"와 "software"에 해당하고, 중간정점들은 "java"와 "software"의 상위어에 해당하는 동의어집합들 $HypSet(java)$ 와 $HypSet(software)$ 의 요소들이다. 각 정점들을 연결하는 유방향화살표는 상하위어관계를 나타내고 있다. 최상위정점은, 본 논문에서 편의상, 모든 어휘들의 공통 최상위개념을 나타내는

Algorithm 1 Build_WordHierarchyGraph($ws,$

$WordNet$)

-
- ```

1. // 입력 : 임의의 어휘들의 집합 $ws = \{w_1, w_2, \dots, w_n\}$,
2. // 워드넷 WordNet := (S, R)
3. // 출력 : 어휘계층구조 $WHG_{ws} := (V, E)$
4. $V_{ws} := \bigcup_{i=1}^n synset_k(w_i)$;
 //어휘 w_i 의 동의어집합(synset)들의 합집합을 구한다.
 //이때, 어휘 w_i 가 $n(\geq 2)$ 개의 동의어를 갖는 경우,
 //사용자가 적합한 의미의 동의어를 선택한다.
5. $V := \{\top\} \cup V_{ws}$;
6. $E := \emptyset$;
7. for each $v \in V_{ws}$ do
 //워드넷의 상하위어관계를 토대로 v부터 최상위정점까지
 //깊이우선탐색에 의해
 //경로 Hpath(v) := <v → v1 → v2 → ... → vk>를 구한다.
8. Hpath(v) := FindHypSet(WordNet , v);
 //경로 Hpath(v)상의 정점들과 변들을
 //어휘계층구조 $WHG_{ws} := (V, E)$ 에 추가한다.
9. $V := V \cup \{v_1, v_2, \dots, v_k\}$;
10. $E := E \cup \{(v, v_1), (v, v_2), \dots, (v, v_k)\}$;
11. return $WHG_{ws} := (V, E)$;

```

End\_Build\_WordHierarchyGraph

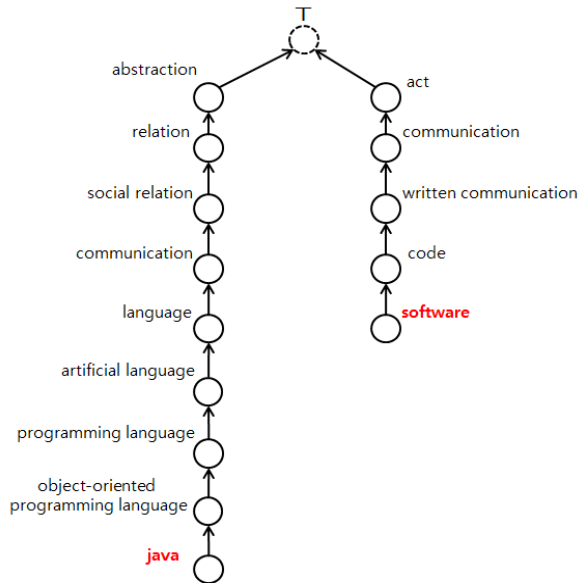


그림 4. "java"와 "software"에 대한 어휘계층구조  $WHG_{(java, software)}$   
 Fig. 4. word hierarchy graph  $WHG_{(java, software)}$

“ $\top$ ”(top)으로 설정하였다.

### 3. 형식개념분석

형식개념분석기법[15]은 개념격자라는 수학적 모델을 기반으로 하는 데이터분석기법의 일종으로서, 개념적인 데이터 분석과 지식기반처리분야의 제반문제들에 대한 수학적 해법을 제공하고 있다. 형식개념분석기법에서는, 분석대상데이터를 객체와 속성, 그리고 이들 사이의 포함관계를 이진데이터테이블(formal context라고 부름)로 나타낸다. 즉, 이진데이터테이블  $K := (G, M, I)$ 는 객체들(Objects)의 집합  $G$ 와 속성들(Attributes)의 집합  $M$ , 그리고  $G$ 와  $M$ 사이의 이항관계  $I \subseteq G \times M$ 으로 구성된다. 즉, 어떤 객체  $g \in G$ 가 속성  $m \in M$ 을 가지고 있을 경우,  $gIm$  또는  $(g, m) \in I$ 로 나타내며, 객체  $g$ 는 속성  $m$ 을 갖는다는 것을 의미한다.

이진데이터테이블의 행과 열의 헤드부분은 각각 객체들과 속성들로 구성된다. 해당 셀에 관련된 객체와 속성이 이항관계  $I$ 를 만족할 경우에는 표1에서와 같이  $\times$ 표시하고, 이외의 경우에는 빈 공간으로 남겨둔다.

이때, 이진데이터테이블  $K := (G, M, I)$ 에 대하여,  $O \subseteq G$ ,  $A \subseteq M$ 일때,  $intent(O) = A \wedge extent(A) = O$ 를 만족하는  $(O, A)$ 를 형식개념(formal concept)이라고 한다. 단,

$$intent(O) := \{a \in M \mid \forall o \in O: (o, a) \in I\},$$

$$extent(A) := \{o \in G \mid \forall a \in A: (o, a) \in I\}.$$

또한, 이진데이터테이블  $K := (G, M, I)$ 로부터 추출한 모든 개념들의 집합을  $B(K) := \{(O, A) \in 2^G \times 2^M \mid intent(O) = A \wedge extent(A) = O\}$ 와 같이 정의한다.

즉, 형식개념  $(O, A)$ 는 객체집합  $O$ 의 모든 원소들(객체들)에 공통적인 속성집합이 집합  $A$ 와 같고, 속성집합  $A$ 의 모든 원소들(속성들)을 공통적으로 갖는 객체들의 집합이 집합  $O$ 와 같음을 의미한다. 또한, 임의의 형식개념  $(O_1, A_1)$ 과  $(O_2, A_2)$ 에 대하여,  $O_1 \subseteq O_2 (\Leftrightarrow A_1 \supseteq A_2)$ 라면,  $(O_1, A_1)$ 은  $(O_2, A_2)$ 의 하위개념, 또는,  $(O_2, A_2)$ 는  $(O_1, A_1)$ 의 상위개념이며,  $(O_1, A_1) \leq (O_2, A_2)$ 와 같이 표현한다. 상·하위개념관계의 정의는, 외연적(extensional) 및 내포적(intensional)

인 관점에서 상위의 개념에 비해 하위의 개념이 보다 적은 외연(extension)과 보다 큰 내포(intension)를 갖는다는 철학적인 개념계층구조와 매우 유사한 특성을 반영하고 있다.

이진데이터테이블  $K := (G, M, I)$ 로부터 만들어진 모든 개념들  $B(K)$ 과 그들 사이의 상·하위개념관계  $\leq$ 로 이루어진 계층구조  $L(K) := (B(K), \leq)$ 를 개념격자(Concept Lattice 또는 Galois Lattice)라고 부른다. 아래에는 이진데이터테이블로부터 개념격자를 구축하기 위한 알고리즘  $FCA(K)$ 를 정의하였다.

```

Algorithm 2 $FCA(K)$
// 입력 : 이진데이터테이블 $K := (G, M, I)$
// 출력 : 개념격자 $L(K) := (B(K), E_{\leq})$
 $B(K) \leftarrow \emptyset$; $E_{\leq} \leftarrow \emptyset$;
for each $g \in G$ do
 $B(K) \leftarrow B(K) \cup (extent(intent(g)), intent(g))$;
end for
for each $(O, A) \in B(K)$ do
 for each $g \in (G - extent(A))$ do
 $X \leftarrow extent(A) \cup \{g\}$;
 if $(extent(intent(X), intent(X)) \notin B(K))$ then
 $B(K) \leftarrow B(K) \cup (extent(intent(X)), intent(X))$;
 end if
 end for
end for
for each $c1 \in B(K)$ do
 for each $c2 \in (B(K) - \{c1\})$ do
 if $(c1 \leq c2) \wedge (\nexists c3 \in (B(K) - \{c1, c2\})[(c1 \leq c3) \wedge (c3 \leq c2)])$ then
 $E_{\leq} \leftarrow E_{\leq} \cup \{(c1, c2)\}$;
 end if
 end for
end for
return $L(K) := (B(K), E_{\leq})$.

```

개념격자는 Hasse Diagram을 사용하여 그림5와같은 계층적인 개념구조형태로 가시화할 수 있다. Hasse Diagram에서는, 각 개념들과 이들 사이의 상·하위관계가 링크에 의해 표시되며, 특히, 개념들 간의 링크에 의해 만들어지는 경로에 의해 상위개념으로부터 하위개념으로 속성들이 상속되며, 하위개념으로부터 상위개념으로 해당 객체들이 전파된다. 예를 들어, o2객체는 고유한 속성으로서 a2를 갖고 있으며, 더불어서, 상위개념들로부터 a1과 a5, 그리고 a3과 같은 속성을 상속받는다. 한편, 속성a3을 갖는 객체로서는 o1, o2, o4가 됨을 알 수

표1. Formal context의 예  
Table 1. Example of Formal context

|    | a1 | a2 | a3 | a4 | a5 |
|----|----|----|----|----|----|
| o1 |    |    | ×  | ×  |    |
| o2 | ×  | ×  | ×  |    | ×  |
| o3 | ×  |    |    | ×  | ×  |
| o4 |    |    | ×  |    |    |

있다. 이와 같은 방법을 사용함으로써, 주어진 문제영역의 객

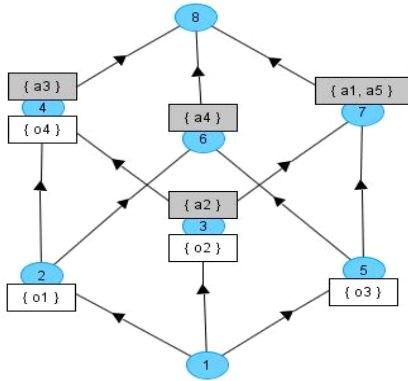


그림 5. 표1에 대한 개념격자

Fig. 5. Concept lattice for Table 1

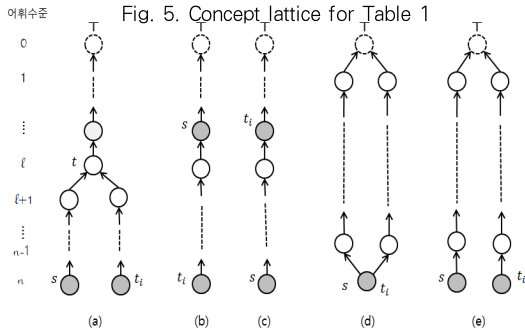


그림 6. 동시사용태그( $s, t_i$ )에 대한  
공통어휘계층구조  $WHG_i := (V, E)$ 의 유형

Fig. 6. Categories of Word Hierarchy  $WHG_i := (V, E)$   
for co-usage of tags ( $s, t_i$ )

체들과 이들이 갖는 속성들을 이진데이터테이블형태로 파악하여, 개념을 추출하고 개념격자형태로 나타냄으로써, 도메인 내의 개념들을 분류하고 체계화할 수 있는 계층적 개념구조를 수월하게 구축할 수 있다.

#### 4. 사용자태그데이터에 대한 워드넷기반 형식개념분석

폭소노미기반 시스템에 축적되어 있는 사용자태그데이터에 대하여, 워드넷을 기반으로 형식개념분석을 수행하기 위하여, 먼저, 폭소노미의 사용자태그데이터를 수집하여 데이터테이블을 구성해야 된다. 사용자태그데이터는 관심대상이 되는 어떤 태그  $s$ 에 대하여 사용자가 동시에 사용하고 있는 태그들, 즉, 태그  $s$ 의 동시사용태그들에 관한 정보가 기록되는 데이터테이블로서, 아래와 같이 정의된다.

[정의5] 폭소노미  $F := (U, T, R, A)$ 의 임의의 태그  $s \in T$ 에

대하여, 동시사용태그 데이터테이블  $K_s := (U_s, T_s, I_s)$ 는 다음과 같은 요소들로 구성된 이진데이터테이블이다:

- $U_s := \{u \in U \mid \exists r \in R: (u, s, r) \in A\} \subseteq U$  : 태그  $s$ 의 사용자 집합.
- $T_s := \{s\} \cup \{t \in T \mid \exists u \in U_s, r \in R[(u, \{s, t\}, r) \in A]\}$  : 태그  $s$ 의 사용자에 의해 동시 사용된 태그들로 구성되는 집합
- $I_s \subseteq U_s \times T_s$  : 태그  $s$  사용자들과 동시사용태그들 사이의 사용관계집합. ■

예를 들어 표2는 태그  $s$ 를 사용하는 4명의 사용자들 ( $u_1, u_2, u_3, u_4$ )에 대한 동시사용태그들에 관한 정보를 나타내고 있는 동시사용태그데이터테이블이다.

표 2. 동시사용태그 데이터테이블  $K_s$ 의 예  
Table 2. Example of data table  $K_s$  for co-tags

| $K_s$ | $s$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ |
|-------|-----|-------|-------|-------|-------|-------|
| $u_1$ | ×   | ×     |       | ×     | ×     | ×     |
| $u_2$ | ×   | ×     | ×     |       | ×     | ×     |
| $u_3$ | ×   |       | ×     | ×     |       | ×     |
| $u_4$ | ×   |       |       | ×     | ×     |       |

표2에서 사용자  $u_1$ 은 태그  $s$ 와  $t_1, s$ 와  $t_3, s$ 와  $t_4, s$ 와  $t_5$ 를, 사용자  $u_2$ 는 태그  $s$ 와  $t_1, s$ 와  $t_2, s$ 와  $t_4, s$ 와  $t_5$ 를, 그리고 사용자  $u_3$ 는 태그  $s$ 와  $t_2, s$ 와  $t_3, s$ 와  $t_5$ 를, 사용자  $u_4$ 는 태그  $s$ 와  $t_3, s$ 와  $t_4$ 를 동시에 사용하고 있음을 나타내고 있다.

임의의 폭소노미로부터 태그  $s$ 를 사용하는 사용자태그데이터를 수집하여 구성된 동시사용태그 데이터테이블  $K_s := (U_s, T_s, I_s)$ 에 대하여, 알고리즘 (Build\_WordHierarchyGraph)을 사용하여 동시사용태그 ( $s, t_i$ )에 관한 다양한 형태의 공통어휘계층구조  $WHG_i := (V, E)$ 가 구축될 수 있다. 즉, 워드넷에 어휘  $s$ 와  $t_i$  사이에 어떠한 유형의 상하위어관계 또는 동의어관계가 성립하는가에 따라서, 그림6의 (a)~(e)와 같은 5종류의 패턴으로 분류하고, 이를 토대로 공통어휘계층구조가 구축될 수 있으며, 사용자가 선택한 공통어휘수준 ( $l$ )에 따라서, 다음과 같은 변환을 수행하여  $\hat{K}_s := (\hat{U}_s, \hat{T}_s, \hat{I}_s)$ 를 구성한다.

- (a) 공통어휘계층구조  $WHG_i := (V, E)$ 에서 어휘  $s$ 와  $t_i$ 의 공통상위어  $t$ 가 존재하는 경우, 어휘  $t$ 의 상위어들 중에

서 사용자가 선택한 공통어휘수준  $l$ 의 어휘  $t_i$ 를 이용하여 태그  $t$ 를  $(s, t_i) < t_i$ 로 대체하여  $\widehat{K}_s := (\widehat{U}_s, \widehat{T}_s, \widehat{I}_s)$ 를 구성한다.

(b) 공통어휘계층구조  $WHG_i := (V, E)$ 에서 어휘  $s$ 가  $t_i$ 의 상위어인 경우(즉,  $t_i < +s$ ),  $s$ 의 상위어들 중에서 사용자가 선택한 공통어휘수준  $l$ 의 어휘  $t_l$ 를 사용하여 태그  $t$ 를  $t_i < s < t_l$ 로 대체하여  $\widehat{K}_s := (\widehat{U}_s, \widehat{T}_s, \widehat{I}_s)$ 를 구성한다.

(c) 공통어휘계층구조  $WHG_i := (V, E)$ 에서 어휘  $s$ 가  $t_i$ 의 하위어인 경우(즉,  $s < +t_i$ ),  $t_i$ 의 상위어들 중에서 사용자가 선택한 공통어휘수준  $l$ 의 어휘  $t_l$ 를 사용하여 태그  $t$ 를  $s < t_i < t_l$ 로 대체하여  $\widehat{K}_s := (\widehat{U}_s, \widehat{T}_s, \widehat{I}_s)$ 를 구성한다.

(d) 공통어휘계층구조  $WHG_i := (V, E)$ 에서 어휘  $s$ 와  $t_i$ 가 동의어인 경우(즉,  $s \equiv t_i$ ), 태그  $t$ 를  $s = t_i$ 로 대체하여  $\widehat{K}_s := (\widehat{U}_s, \widehat{T}_s, \widehat{I}_s)$ 를 구성한다.

(e) 그 밖의 경우, 즉, 어휘  $s$ 와  $t_i$ 가 서로 무관한 경우, 아무런 변환을 수행하지 않는다. 즉,  $\widehat{K}_s = K_s$ .

이상에서 설명한 전체 변환과정은 아래의 알고리즘3(사용자태그데이터의 워드넷기반 형식개념분석)으로 정형화하여 정리할 수 있다.

### IV. 지원도구의 개발 및 실험

제3장의 각 정의들과 알고리즘들을 토대로, 본 연구에서는 폭소노미 사용자태그데이터의 워드넷기반 형식개념분석을 수행하기 위한 지원 도구(TagLighter, 한국저작권위원회 프로그램등록번호: 제C-2013-015062호)를 개발하였다. 여기서는, TagLighter의 구성요소들에 대하여 소개하고, TagLighter의 유용성을 검증하기 위한 실험결과에 대하여 설명한다.

#### 1. TagLighter의 구성

##### 1.1 폭소노미 사용자태그데이터 수집모듈

폭소노미 사용자태그데이터 수집모듈(그림7의 실행화면(1))에서는, 사용자가 입력한 관심있는 태그들(1개의 Seed Tag와 여러 개의 Associated Tag)을 토대로, Open API를 제공하고 있는 Bibsonomy.org사이트로부터 사용자태그

#### Algorithm 3 사용자태그데이터의 워드넷기반 형식개념분석

1. Bibsonomy로부터 태그  $s$ 를 사용하는 사용자태그데이터를 수집하여 동사사용태그 데이터테이블  $K_s := (U_s, T_s, I_s)$ 를 구성.
2. 사용자가 변환대상태그  $t_i \in T_s - \{s\}$ 를 선택.
3. 워드넷을 기반으로 공통사용태그  $(s, t_i)$ 에 대한 공통어휘계층구조  $WHG_i := (V, E)$ 를 구축. 즉,  $WHG_i := \text{Build\_WordHierarchyGraph}((s, t_i), \text{WordNet})$ ;
4. 다음의 각 조건에 따라서  $K_s := (U_s, T_s, I_s)$ 로부터  $\widehat{K}_s := (\widehat{U}_s, \widehat{T}_s, \widehat{I}_s)$ 으로 변환작업수행.
  - 4.1 **if**  $(\exists t \in V[(s < +t) \wedge (t_i < +t)])$  **then** // (a)
 

//  $t_l$ 은  $s$ 와  $t_i$ 의 공통상위어  $t$ 의 상위어들 중에서 사용//자가 선택한 공통어휘수준  $l$ 의 어휘

$$\begin{aligned} \widehat{U}_s &\leftarrow U_s; & \widehat{T}_s &\leftarrow (T_s - \{t_i\}) \cup \{(s, t_i) < t_l\}; \\ \widehat{I}_s &\leftarrow (I_s - \bigcup_{\forall u \in U_s} \{(u, t_i) | \exists (u, t_i) \in I_s\}) \\ &\quad \cup \bigcup_{\forall u \in U_s} \{(u, (s, t_i) < t_l) | \exists (u, t_i) \in I_s\}; \end{aligned}$$
  - 4.2 **else if**  $(t_i < +s)$  **then** // (b)
 

//  $t_l$ 은  $s$ 의 상위어들 중에서 사용자가 선택한 공통어휘수준  $l$ 의 어휘

$$\begin{aligned} \widehat{U}_s &\leftarrow U_s; & \widehat{T}_s &\leftarrow (T_s - \{t_i\}) \cup \{t_i < s < t_l\}; \\ \widehat{I}_s &\leftarrow (I_s - \bigcup_{\forall u \in U_s} \{(u, t_i) | \exists (u, t_i) \in I_s\}) \\ &\quad \cup \bigcup_{\forall u \in U_s} \{(u, t_i < s < t_l) | \exists (u, t_i) \in I_s\}; \end{aligned}$$
  - 4.3 **else if**  $(s < +t_i)$  **then** // (c)
 

//  $t_l$ 은  $t_i$ 의 상위어들 중에서 사용자가 선택한 공통어휘수준  $l$ 의 어휘

$$\begin{aligned} \widehat{U}_s &\leftarrow U_s; & \widehat{T}_s &\leftarrow (T_s - \{t_i\}) \cup \{s < t_i < t_l\}; \\ \widehat{I}_s &\leftarrow (I_s - \bigcup_{\forall u \in U_s} \{(u, t_i) | \exists (u, t_i) \in I_s\}) \\ &\quad \cup \bigcup_{\forall u \in U_s} \{(u, s < t_i < t_l) | \exists (u, t_i) \in I_s\}; \end{aligned}$$
  - 4.4 **else if**  $(s \equiv t_i)$  **then** // (d)
 
$$\begin{aligned} \widehat{U}_s &\leftarrow U_s; & \widehat{T}_s &\leftarrow (T_s - \{t_i\}) \cup \{s = t_i\}; \\ \widehat{I}_s &\leftarrow (I_s - \bigcup_{\forall u \in U_s} \{(u, t_i) | \exists (u, t_i) \in I_s\}) \\ &\quad \cup \bigcup_{\forall u \in U_s} \{(u, s = t_i) | \exists (u, t_i) \in I_s\}; \end{aligned}$$
  - 4.5 **else** // (e)
 
$$\widehat{K}_s \leftarrow K_s;$$
5. 형식개념분석과정 **FCA**( $\widehat{K}_s$ )를 실행.
6. 결과(분류체계)를 화면에 표시.



데이터를 수집한다. 수집된 사용자태그데이터는 내부적으로 이진데이터테이블( $K_s := (U_s, T_s, I_s)$ )형태로 저장되며, 여기에는 관심 있는 태그(seed tag)  $s$ 와 동시 사용된 태그들(Associated Tags)  $T_s = \{t_1, t_2, \dots, t_n\}$ 을 어떤 사용자들( $U_s = \{u_1, u_2, \dots, u_m\}$ )이 사용했는지에 관한 정보( $I_s \subseteq U_s \times T_s$ )가 표현되어 있다.

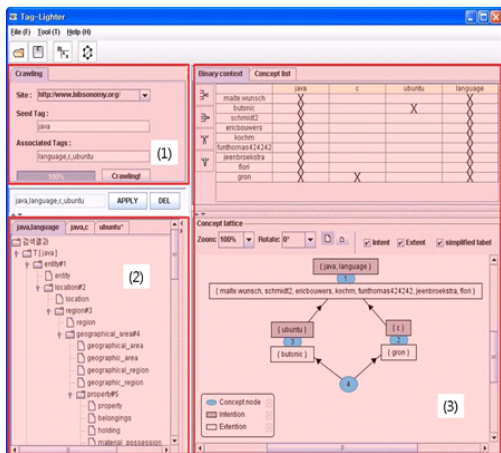
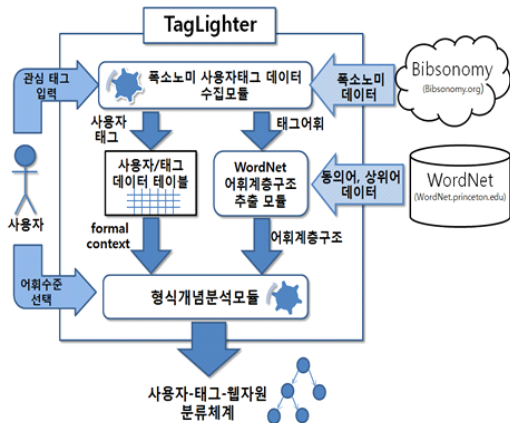


그림 7 TagLighter의 구성과 실행화면 모습  
Fig. 7. Architecture and Screenshot of TagLighter

### 1.2 WordNet 어휘계층구조 추출모듈

WordNet어휘계층구조 추출 모듈(그림7의 실행화면(2))에서는, Bibsonomy.org사이트에서 수집된 사용자태그데이터와 워드넷에서 제공하는 동어 및 상위어 데이터 등을 기반으로, 사용자들의 태그들에 관한 트리형태의 어휘계층구조를 구성한다. 즉, Seed tag  $s$ 와 Associated tags  $\{t_1, t_2, \dots, t_n\}$ 에 대하여, 알고리즘

**Build\_WordHierarchyGraph**를 적용하여 어휘계층구조  $WHG_1, WHG_2, \dots, WHG_n$ 를 구성한다.

### 1.3 형식개념분석모듈

형식개념분석모듈에서는, Bibsonomy.org사이트로부터 수집된 사용자태그데이터  $K_s := (U_s, T_s, I_s)$ 와 워드넷으로부터 구성된 어휘계층구조  $WHG_1, WHG_2, \dots, WHG_n$ 를 토대로, 사용자가 선택한 어휘수준에 맞추어서 동시사용태그에 관한 데이터테이블( $\hat{K}_s$ )을 구성하고(그림7의 실행화면(3)의 상단부), 앞절에서 정의한 알고리즘 **FCA**를 적용하여 사용자태그분류체계(개념격자  $L(\hat{K}_s)$ )를 구축하여 그 결과를 Hasse Diagram형태로 화면에 표시한다(그림7의 실행화면(3)의 하단부).

## 2. 실험

연구관련논문 및 북마킹 서비스를 제공하는 대표적인 폭소노미기반 시스템인 Bibsonomy.org을 대상으로, 본 연구에서 개발된 TagLighter의 유용성을 확인하기 위하여 다음과 같은 간단한 실험을 수행하였다.

먼저, Bibsonomy사용자들이 가장 많이 사용하고 있는 상위10개의 태그들 중에서, WordNet에 등록되어 있는 8개의 태그들(internet, web, design, data, information, software, mobile, science)을 실험대상으로 선택하였다. 선택된 8개의 태그들을 조합하여 seed tag와 associated tag로 설정하여 TagLighter의 폭소노미 사용자태그데이터 수집모듈을 실행시켜서 Bibsonomy.org로부터 사용자태그데이터를 수집하였다. 또한, 수집된 사용자태그데이터에 대하여, WordNet어휘계층구조 추출모듈을 실행시켜서, 공통어휘계층구조의 유형을 파악하였다(표3).

표3의 공통태그들중에서 상위4개의 태그들(internet, web, design, data)에 대하여, 1개의 Seed Tag와 1~3개의 Associated Tag들로 구성되는 공통태그에 관한 가능한 모든 조합들을 각각 TagLighter에 입력하여 형식개념분석모듈을 실행시킨 결과, 표4~표5와 같은 다양한 사용자태그분류체계를 구축할 수 있다(단, 공통어휘수준은 2~4중에서 사용자가 임의로 선택하도록 하였다. 또한, 지면관계상, 본 실험에서는 seed tag = "internet"과 associated tag = {web, design, data}의 모든 조합들(총7가지)에 대한 실험결과만을 수록하였다). 이상의 실험에서는 TagLighter에서 제공하는 제반기능(사용자태그데이터의 수집, WordNet 어휘계층구조의 구성, 형식개념분석기법의 적용 등)을 사용하여, Bibsonomy.org에 축적되어 있는 사용자태그

표 3. Bibsonomy.org에서 추출된 동시사용태그들의 어휘계층구조의 유형 및 리소스의 개수

Table 3. Categories of Word Hierarchies and number of related resources for co-tags driven from Bibsonomy.org

| Seed Tag | Associated Tag | 1        | 2      | 3      | 4      | 5           | 6        | 7      | 8       |
|----------|----------------|----------|--------|--------|--------|-------------|----------|--------|---------|
|          |                | internet | web    | design | data   | information | software | mobile | science |
| 1        | internet       |          | a(70)  | a(2)   | a(9)   | a(40)       | a(20)    | a(23)  | a(3)    |
| 2        | web            | a(70)    |        | a(72)  | a(159) | a(284)      | a(142)   | a(43)  | a(34)   |
| 3        | design         | a(2)     | a(72)  |        | a(31)  | a(70)       | a(253)   | a(66)  | a(68)   |
| 4        | data           | a(9)     | a(159) | a(31)  |        | d(169)      | e(74)    | e(29)  | a(11)   |
| 5        | information    | a(40)    | a(284) | a(70)  | d(169) |             | a(69)    | a(40)  | a(115)  |
| 6        | software       | a(20)    | a(142) | a(253) | e(74)  | a(69)       |          | a(33)  | e(18)   |
| 7        | mobile         | a(23)    | a(43)  | a(66)  | e(29)  | a(40)       | a(33)    |        | e(4)    |
| 8        | science        | a(3)     | a(34)  | a(68)  | a(11)  | a(115)      | e(18)    | e(4)   |         |

표 4. 실험결과(1/2)

Table 4. Experiment results(1/2)

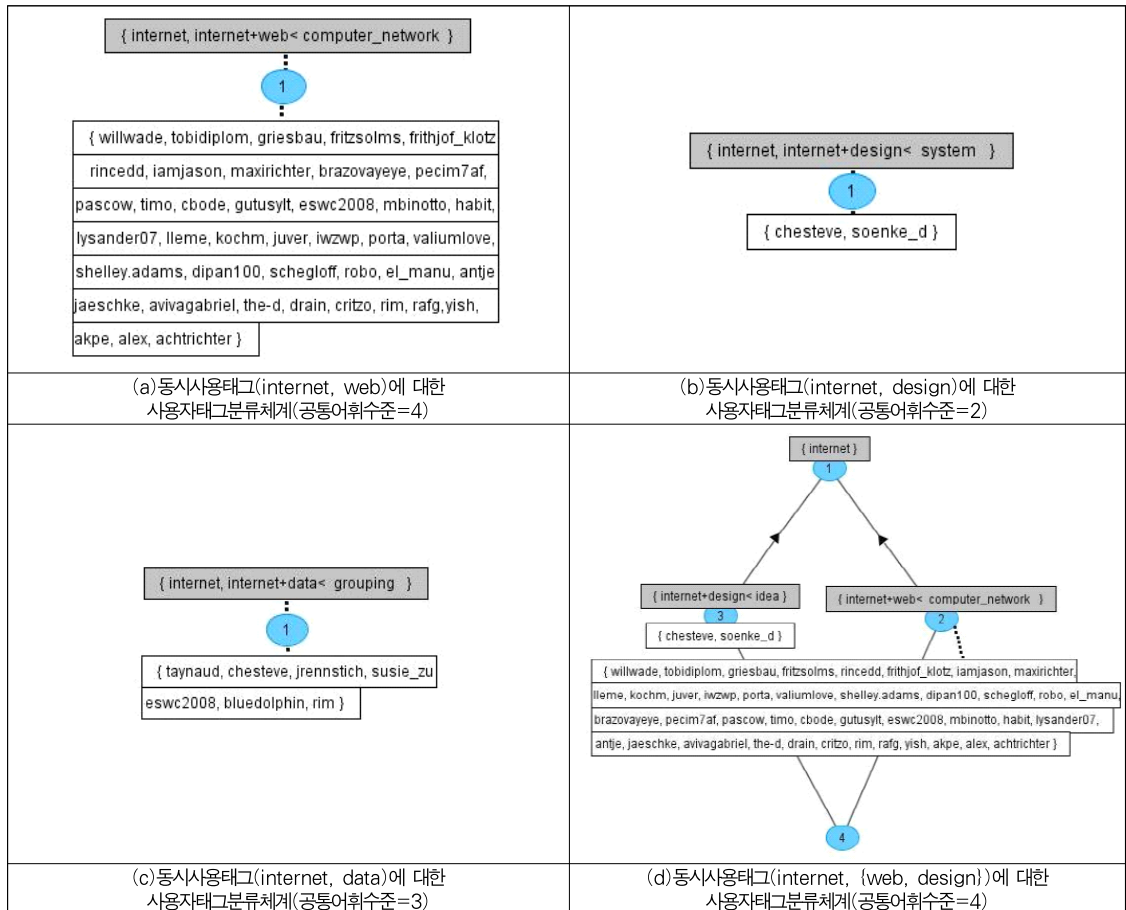
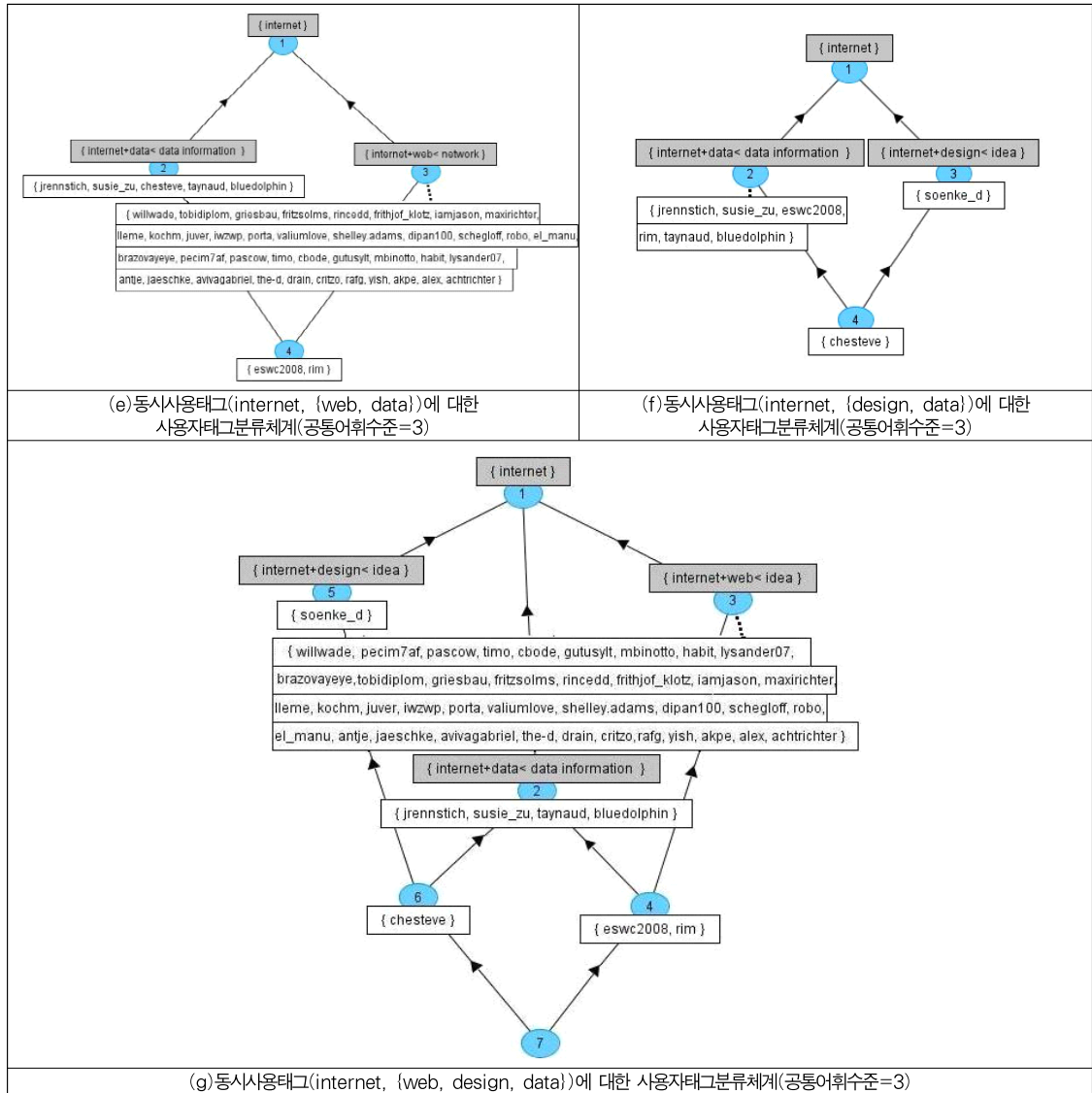


표 5. 실험결과(2/2)  
Table 5. Experiment results(2/2)



데이터를 수집하고 WordNet기반의 형식개념분석을 수행하여 다양한 사용자태그분류체계를 구축할 수 있음을 확인할 수 있다.

### V. 결론 및 향후 연구과제

본 논문에서는, 관심있는 웹자원에 대하여 자유형태의 키

워드인 태그를 사용자들이 자유롭게 직접 부착하고 공유함으로써 자발적으로 구축되는 웹자원 분류체계인 폭소노미기반 시스템에 축적되어 있는 사용자태그데이터를 표준화된 어휘 정보체계인 워드넷을 기반으로 형식개념분석기법을 적용하여, 다양한 사용자태그분류체계를 구축하기 위한 기법을 제안하였다. 또한, 본 연구에서 제안한 기법을 지원하는 도구(TagLighter)를 개발하였으며, TagLighter를 사용하여 대표적인 폭소노미기반시스템인 Bibsonomy.org의 사용자태그데이

터를 대상으로 본 연구에서 제안한 기법을 적용한 실험결과를 보고하였다.

본 연구에서 제안한 기법을 토대로 TagLighter에 의해 구축된 사용자태그에 대한 분류계층구조는 사용자태그에 대한 보다 상세한 이해를 제공하기 때문에 폭소노미기반 시스템에서의 정보검색과 데이터마이닝에 유용하다. 본 논문의 연구결과는, 폭소노미기반의 웹서비스와 소셜네트워크시스템, 시맨틱 웹 어플리케이션 등을 대상으로 하는 웹데이터 마이닝분야에 응용할 수 있다.

본 논문에서는, 주로 워드넷에 수록된 명사들 사이의 상하 위어관계를 중심으로 어휘계층구조를 구성하여 사용자태그분류체계 구축에 활용하였으나, 현실상황에 적합한 보다 다양한 태그분류체계를 구축하기 위하여 동의어 및 반의어, 그리고 부분어 등을 토대로 다양한 품사의 어휘계층구조를 구성하여 적용하기 위한 후속연구가 진행될 예정이다.

## 참고문헌

- [1] Y. Keshet, "Classification systems in the light of sociology of knowledge", *Journal of Documentation*, Vol. 67, Issue 1, pp.144-158, September 2011.
- [2] I. Peters, *Folksonomies. Indexing and Retrieval in Web 2.0*, Walter de Gruyter & Co., pp.153-262, 2009.
- [3] M. Strohmaier, C. Körner, R. Kern, "Understanding why users tag: A survey of tagging motivation literature and results from an empirical study", *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol.17, pp.1-11, December 2012.
- [4] D. Kwon, S. Cho, "Calculation of Semantic Relatedness of Tags on Web Images using On-line Lexical Dictionaries", *Proceedings on the 2009 Spring Conference, Korea Multimedia Society*, Vol.12, No.1, pp.277-280, ChonNam National University, May 2009.
- [5] W. G. Stock, "Folksonomies and science communication : A mach-up of professional science databases and Web 2.0 services", *Information Services & Use*, Vol. 27, pp.97-103, August 2007.
- [6] S. Anfinnsen, G. Ghinea, S. de Cesare, "Web 2.0 and folksonomies in a library context", *International Journal of Information Management*, Vol. 31, Issue 1, pp.63-70, February 2011.
- [7] S. Pi, H. Liao, S. Liu, C. Lin, "Content Classification by folksonomies: Framework of Social Bookmarking System", *Journal of Software*: Vol.7, No.4, pp.745-748, April 2012.
- [8] C. Kiu, E. Tsui, "TaxoFolk : A hybrid taxonomy-folksonomy classification for enhanced knowledge navigation", *Knowledge Management Research & Practice*, Vol. 8, Issue 1, pp.24-32, March 2010.
- [9] D. Yoo, G. Kim, K. Choi, Y. Suh, "CTKOS : Categorized Tag-based Knowledge Organization System", *Journal of Intelligence Information System*, Vol. 17, No. 4, pp.59-74, December 2011.
- [10] Y. Kang, S. Hwang, "An development of framework and a supporting tool for organizing Grouped Folksonomy", *Journal of the Korea Society of Computer and Information*, Vol. 16, Issue 5, pp.109-125, 2011.
- [11] H. Kim, D. Kim, J. Jo, "A Hybrid Music Recommendation system Combining Listening Habits and Tag Information", *Journal of the Korea Society of Computer and Information*, Vol. 18, Issue 2, pp.107-116, 2013.
- [12] S. A. Golder, B. A. Huberman, "Usage patterns of collaborative tagging systems", *Journal of Information Science*, Vol.32, No.2, pp.198-208, April 2006.
- [13] F. Limpens, F. Gandon, M. Buffa, "Helping online communities to semantically enrich folksonomies", *Proceedings of the Web Science Conference*, pp.1-8, Raleigh, NC, USA April 2010.
- [14] WordNet, <http://wordnet.princeton.edu/>
- [15] B. Ganter, R. Wille, "Formal Concept Analysis : Mathematical Foundations", Springer, pp.17-58, 1999.

- [16] A. Hotho, "Data Mining on Folksonomies", Intelligent Information Access, Vol. 301, Springer, pp.57-82, 2010.

## 저 자 소 개



### 황 석 형

1991: 강원대학교 전자계산학과 이학사.

1993: 일본 오사카대학교  
정보공학과 공학석사.

1997: 일본 오사카대학교  
정보공학과 공학박사

현 재: 선문대학교 컴퓨터공학과 교수

관심분야: 데이터마이닝, 시맨틱웹,

온톨로지공학, 소프트웨어공학

Email : shwang@sunmoon.ac.kr