

형식개념분석을 이용한 폭소노미 마이닝 기법과 지원도구의 개발

강유경^{1*}, 황석형¹, 양해술²
¹선문대학교 컴퓨터공학부, ²호서대학교 벤처전문대학원

On development of supporting tool for Folksonomy Mining based on Formal Concept Analysis

Yu-Kyung Kang^{1*}, Suk-Hyung Hwang¹ and Hae-Sool Yang²

¹Department of Computer Science & Engineering, Sun Moon University,

²Graduate School of Venture, Hoseo University

요약 폭소노미(folksonomy)는 웹에 존재하는 리소스에 대해 사용자가 자유롭게 선택한 태그(tag)를 붙여서 정보를 체계화하는 새로운 분류 체계이다. 폭소노미 기반의 시스템에서는 사용자들의 협력태깅에 의해 사용자, 태그, 리소스 사이의 관계를 나타내는 3항원소데이터가 생성된다. 이와 같은 폭소노미 데이터는 웹 리소스에 대한 정보체계화를 위한 메타데이터로서 시맨틱 웹과 웹2.0 분야에 활용되고 있다. 본 논문에서는 다중다양한 폭소노미 데이터를 다양한 관점으로 분석하여 유용한 정보를 추출하기 위한 형식개념분석 기반의 폭소노미 데이터 마이닝 기법을 제안하고, 이를 지원하기 위한 분석도구 FMT를 개발하였다. 또한, 제안한 기법과 FMT의 유용성을 검증하기 위하여, 폭소노미 기반 시스템인 del.icio.us의 데이터를 대상으로 실험을 수행하고, 그 결과를 보고한다.

Abstract Folksonomy is a user-generated taxonomy to organize information by which a user assigns tags to resources published on the web. Triadic datas that indicate relations of between users, tags, and resources, are created by collaborative tagging from many users in folksonomy-based system. Such the folksonomy data has been utilized in the field of the semantic web and web2.0 as metadata about web resources. In this paper, we propose FCA-based folksonomy data mining approach in order to extract the useful information from folksonomy data with various points of view. And we developed tool for supporting our approach. In order to verify the usefulness of our proposed approach and FMT, we have done some experiments for data of del.icio.us, which is a popular folksonomy-based bookmarking system. And we report about result of our experiments.

Key Words : Folksonomy mining, Formal Concept Analysis, Fuzzy Formal Concept Analysis, Rough Formal Concept Analysis

1. 서론

웹 2.0시대의 대표적인 특징인 폭소노미(folksonomy)는 일반적인 대중을 의미하는 folk와 분류법을 의미하는 taxonomy의 합성어로서, 그림1과 같이 웹에 존재하는 리소스에 대해 사용자가 자유롭게 선택한 태그(tag)를 붙여서 정보를 체계화하는 새로운 분류 체계이다. 폭소노미는 흔히 협력 태깅(Collaborative tagging) 또는 소셜 태깅

(Social tagging)으로 알려져 있으며, 사용자들이 자유롭게 참여하는 bottom-up 방식의 분류를 기반으로 하고 있어서, 시시각각으로 변하는 웹의 리소스를 분류하는데 적합한 방법으로 WWW 분야에 널리 사용되고 있다.

근래에 널리 사용되고 있는 폭소노미 기반 시스템으로는 del.icio.us, Flickr, Youtube, BibSonomy 등이 있다. del.icio.us[1]는 자신의 즐겨찾기를 저장하고 타인들과 공유하는 소셜 북마킹(Social bookmarking) 서비스를 제공

*교신저자 : 강유경(yukyung.kang@gmail.com)

접수일 09년 04월 10일

수정일 09년 06월 12일

재제확정일 09년 08월 19일

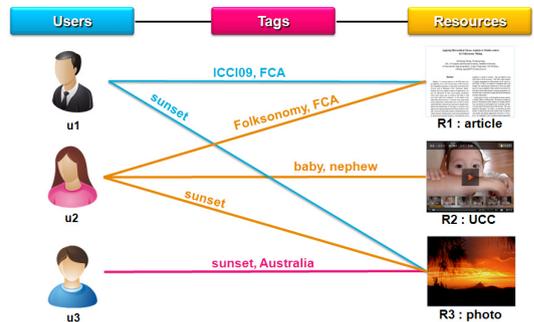
한다. 사용자는 웹에 공개된 리소스를 북마크할 때 해당 리소스를 설명하기 위한 태그를 함께 부착하여 저장할 수 있으며, 동일한 리소스를 북마크할 때 다른 사용자들이 사용한 태그를 추천받을 수 있다. Flickr[2]는 온라인 사진 관리 및 공유를 위한 커뮤니티 사이트로 전 세계적으로 큰 인기를 얻고 있다. 사용자는 자신의 이미지를 업로드하면서 이를 설명하는 태그를 부착할 수 있으며 자신이 찾고자 하는 이미지를 태그를 활용하여 검색할 수 있다. 웹2.0 시대가 도래하면서 사용자 제작 콘텐츠(UCC : User Created Contents)들이 급증하게 되었고, Youtube[3]는 사용자가 직접 제작한 동영상들을 공유하기 위한 폭소노미 기반 커뮤니티로서 각광받고 있다. BibSonomy[4]는 북마크뿐만 아니라 논문 등과 웹사이트에 공개된 리소스에 태깅을 하고 공유할 수 있다. 또한 연구자, 과학자들을 위한 학술적 의도로 만들어진 Conntea[5]와 CiteULike[6] 등이 있다. 이외에도 폭소노미 기반의 다양한 웹 애플리케이션들이 존재하며, 이와 같은 시스템들은 기본적으로 사용자들이 태깅한 정보를 토대로 웹 리소스를 검색하는 기능을 제공하고 있다.

폭소노미 기반의 시스템에서는 사용자(user), 태그(tag), 그리고 리소스(resource) 사이의 관계를 나타내는 3항원소데이터(triadic data)를 기반으로 하고 있다. 즉, 대부분의 폭소노미 기반 웹 애플리케이션에서는 사용자가 웹에 존재하는 다양한 리소스(논문, 사진, 동영상, 등)에 태그를 붙여서 3항원소데이터를 구성하고 공유함으로써, 이를 기반으로 사용자들에게 다양한 서비스를 제공한다. 이와 같이 사용자들의 협력 태깅으로 구축된 폭소노미는 웹 리소스들에 대한 훌륭한 메타데이터로 활용될 수 있다.

그러나, 현재 폭소노미 기반 웹 애플리케이션들은 태그의 사용 빈도수에 비례하여 글자 크기를 할당하는 태그 클라우드(Tag cloud)의 제공과 co-occurrence를 토대로 하는 관련 태그의 표시 등 기본적인 통계적 방법을 기반으로 하는 단순한 정보만을 제공하고 있으며, 협력 태깅에 의해 축적된 메타데이터를 충분히 활용하지 못하고 있다. 따라서 본 논문에서는, 방대한 양의 폭소노미 메타데이터를 다양한 관점으로 분석하여 유용한 정보를 추출하기 위해 형식개념분석 기반의 폭소노미 데이터 마이닝 기법을 제안한다. 또한, 본 연구에서 제안한 기법을 지원하기 위한 도구를 개발하여 del.icio.us의 데이터에 적용하여 분석한 실험 결과를 보고한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 데이터 마이닝 분야의 관련 연구를 소개하고, 3장에서는, 본 연구에서 제안한 형식개념분석기법 기반 폭소노미 데이터 마이닝 기법 및 알고리즘과 지원도구를 소개한다. 4

장에서는 본 연구에서 제안한 기법을 실제 폭소노미 데이터에 적용하여 실시한 실험결과에 대해 보고하고, 결론 및 향후 연구과제에 대해서 설명한다.



[그림 1] 폭소노미의 예

2. 관련 연구

2.1 데이터 마이닝

데이터마이닝은 대용량의 데이터로부터, 이전에 잘 알려지지 않은 패턴, 규칙 등을 찾아내고 이를 모형화하여 유용한 지식을 추출하는 일련의 과정들이다. 즉, 데이터마이닝은 대용량의 데이터로부터 새롭고 의미 있는 정보를 추출하여 의사 결정에 활용하는 작업을 의미한다. IT 기술이 점차 발달함에 따라 데이터마이닝은 통계학(Statistics), 데이터베이스(Database), 패턴인식(Pattern Recognition), 기계학습(Machine Learning), 의학(Medical Science), 생물정보학(Bioinformatics), 인공지능(Artificial Intelligence) 등 다양한 분야에 활용되고 있다[7-9].

데이터마이닝을 통해 얻을 수 있는 정보의 형태는 매우 다양하며, 이에 따라 다양한 기법들이 존재한다. 데이터의 특성과 목적에 따라 기법들을 선택하고 조합하여 유용한 정보를 추출해 낼 수 있다. 데이터마이닝의 대표적인 기법으로는 데이터 사이의 관련성을 표현하는 연관 규칙(Associations rule), 과거에 수집된 데이터들을 분석하여 이들 사이에 존재하는 패턴을 트리형태로 만들어서 새로운 데이터를 분류하고 예측하는 의사결정나무(Decision trees)가 있다. 또한, 인간 두뇌의 신경세포를 모방한 개념으로 인간의 경험으로부터 학습해 가는 두뇌의 신경망 활동을 흉내 내어 자신이 가진 데이터로부터 반복적인 학습과정을 거쳐 패턴을 찾아내고 이를 일반화함으로써 향후를 예측하고자 하는 신경망모형(Neural networks)과 같은 기법들이 있다. 이외에도 전통적인 통

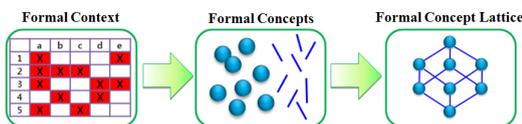
계기법 및 분류(Classification), 군집화(Clustering), 요약(Summarization) 등 다양한 분석기법들이 있다[7-10].

2.2 웹 데이터 마이닝

최근에는 WWW를 이용하는 전 세계의 수많은 사용자들의 다양한 데이터가 축적되고 있으며, 이러한 웹 데이터는 계속적으로 증가확장체제화되고, 시시각각으로 변화하는 특성을 가지고 있다. 이와 같은 웹 데이터를 분석하여 유용한 지식정보들을 추출하기 위한 다양한 연구들[11-15]이 활발하게 진행되고 있다. Ching-man Au Yeung et al.[11]는 애매모호한 태그의 의미를 파악하기 위하여 폭소노미의 구성요소들을 2부그래프(bipartite graph)를 기반으로 하는 클러스터분석기법을 제안하였다. Robert Jaschke et al.[12]는 tri-concept 데이터 마이닝 기법을 기반으로 폭소노미에 내재되어 있는 암묵적인 사용자 공유 정보를 추출하기 위한 기법을 제안하였다. Suk-Hyung Hwang et al.[13]는 폭소노미의 3항원소정보에 대한 계층화된 동치류 분석기법을 제안하였다. 한편, Kim et al.[14]은 폭소노미의 3항원소정보를 형식개념분석기법[16,17]을 이용하여 분석하고, 사용자들의 공통태그를 추출하기 위한 기법을 제안하였다. 이강표[15]는 협력태깅으로 이루어진 태그들 사이의 의미적인 상하위 관계를 밝혀내기 위해 del.icio.us의 태그들을 대상으로 Wikipedia 텍스트를 이용한 태그들간 상하위 관계 산출 기법을 제안하였다.

2.3 형식개념분석

형식개념분석기법(FCA: Formal Concept Analysis)은 주어진 데이터로부터 공통속성을 갖는 객체들을 클러스터링 하여 정보의 최소단위로서 개념(Concept)들을 추출하고 그들 사이의 관계를 토대로 계층화하여 데이터에 내재된 개념들의 구조를 가시화 해주는 데이터분석기법으로서, 형식개념분석기법의 처리과정은 그림2와 같다. 형식개념분석기법을 사용함으로써 실세계의 데이터에 함축된 개념들에 대한 계층구조를 효과적으로 구축할 수 있기 때문에, 현재 의학, 정보과학, 소프트웨어 공학, 등 다양한 분야에서 적용하여 활용되고 있다[16,17].



[그림 2] 형식개념분석기법 처리 과정

형식개념분석기법에서는 분석 대상이 되는 데이터가

Formal context라는 일종의 이진데이터 테이블 형태로 주어진다. Formal context $K=(G, M, I)$ 는 객체들(Objects)의 집합 G 와 속성들(Attributes)의 집합 M , 그리고 G 와 M 사이의 이항관계 $I \subseteq G \times M$ 로 구성된다. 즉, 어떤 객체 g 가 속성 m 을 가지고 있을 경우, gIm 또는 $(g, m) \in I$ 로 나타내며, g 는 m 을 갖는다는 것을 의미한다. 예를 들어, 표 1에서 객체 $o1$ 은 속성 $a1$ 과 $a2$ 를 갖는다.

[표 1] Formal Context의 예

Attributes Objects	a1	a2	a3	a4
o1	X	X		
o2	X	X	X	
o3			X	
o4				X

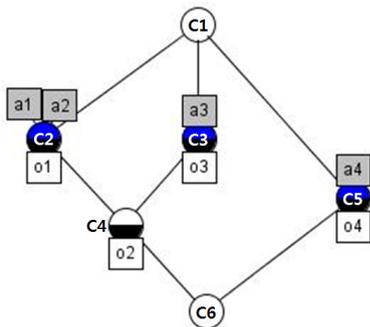
Formal context $K=(G, M, I)$ 에 대하여, $O \subseteq G, A \subseteq M$ 일 때, $O' = A \wedge A' = O$ 를 만족하는 (O, A) 를 개념(formal concept)이라고 한다. 단, $O' := \{a \in M \mid \forall o \in O: (o, a) \in I\}$, $A' := \{o \in G \mid \forall a \in A: (o, a) \in I\}$. 즉, formal concept (O, A) 는 O 의 모든 객체들이 공통적으로 갖는 속성들의 집합이 A 와 같고, A 의 모든 속성들을 공통적으로 갖는 객체들의 집합이 O 와 같은 객체집합과 속성집합으로 구성된다. 표1로부터 총 6개의 개념들이 표2와 같이 추출하였다. 또한, 임의의 개념 $(O1, A1), (O2, A2)$ 에 대하여, $O1 \subseteq O2 (\Leftrightarrow A1 \supseteq A2)$ 라면, $(O1, A1)$ 은 $(O2, A2)$ 의 상위개념(또는, $(O2, A2)$ 는 $(O1, A1)$ 의 하위개념)이며, $(O1, A1) \leq (O2, A2)$ 와 같이 표현한다. Formal context $K=(G, M, I)$ 로부터 만들어진 모든 개념들의 집합 C 와 그들 사이의 상하위개념관계로 이루어진 계층구조 $L := (C, \leq)$ 을 개념격자(Concept Lattice 또는 Galois Lattice)라고 부른다.

[표 2] 표1의 context로부터 추출된 모든 개념들

ID	Extents	Intents
C1	{o1, o2, o3, o4}	{ }
C2	{o1, o2}	{a1, a2}
C3	{o2, o3}	{a3}
C4	{o2}	{a1, a2, a3}
C5	{o4}	{a4}
C6	{ }	{a1, a2, a3, a4}

개념격자는 Hasse Diagram을 사용하여 계층적인 개념 구조형태로 가시화할 수 있다. Hasse Diagram은 정점들과 링크들로 구성되며, 각 정점은 개념격자의 개념들에 대응하고, 각 정점에는 2종류에 레이블(Extension과

Intension)이 각각 하단과 상단에 표시된다. 각 개념들 사이를 연결하는 링크들은 개념들 사이의 “상·하위개념관계”를 나타낸다. 특히, 개념격자를 나타낸 Hasse Diagram에서는 상위부분의 정점들은 보다 추상적이고 일반화된 개념들을, 반면에 하위부분에는 보다 구체화된 개념들이 위치하게 된다. 표1의 context에서 추출한 모든 개념들 사이에 상·하위개념관계를 파악하여 개념격자를 구축하면 그림3과 같다.



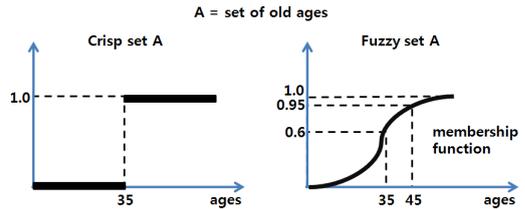
[그림 3] 표1의 context에 대한 개념격자

개념격자를 나타낸 Hasse Diagram에서는, 각 개념들과 이들 사이의 상·하위관계가 정점과 링크에 의해 각각 표시되며, 특히, 개념들 간의 링크에 의해 만들어지는 경로에 의해 상위개념으로부터 하위개념으로 속성들이 상속되며, 하위개념으로부터 상위개념으로 해당 객체들이 전파된다. 예를 들면, 그림3에서 개념C2는 고유의 객체 o1과 하위개념C4로부터 전파된 객체 o2, 그리고 속성 a1과 a2로 구성되어있다. 즉, C2는 o1과 o2가 공통적으로 갖는 속성이 a1과 a2임을 나타내는 개념이다. 또한, 개념 C4는 객체 o2와 상위개념C2와 C3으로부터 각각 계승된 “a1, a2”와 “a3”으로 구성되며, “o2는 속성 a1, r2, r3를 갖는다.”고 해석한다. 이와 같이 형식개념분석기법을 사용함으로써, 주어진 문제영역의 객체들과 이들이 갖는 속성들을 context형태로 파악하여, 개념을 추출하며 개념격자 형태로 나타냄으로써, 도메인 내의 개념들을 공통속성이 추출된 형태로 분류할 수 있는 계층적 개념구조를 수월하게 구축할 수 있다.

2.4 퍼지 집합 이론

퍼지 집합 이론(FST : Fuzzy Set Theory)은 ‘키가 큰 사람’, ‘나이가 많은 사람’, ‘7에 가까운 수’ 등과 같이 인간의 언어와 사고에 관련된 “애매함”과 “모호성”을 정량적으로 다루기 위해 1965년 L. A. Zadeh에 의해 제안되었다[18,19].

실세계에 존재하는 애매모호한 데이터를 다루기 위해 퍼지집합이론[20-22]이 제안되었다. 퍼지집합은 집합의 각 원소가 그 집합에 귀속되어지는 정도를 [0, 1]내의 수로 나타낸 귀속도(歸屬度, membership degree)를 갖는 원소들의 집합이다.



[그림 4] 일반적인 집합과 퍼지 집합의 예

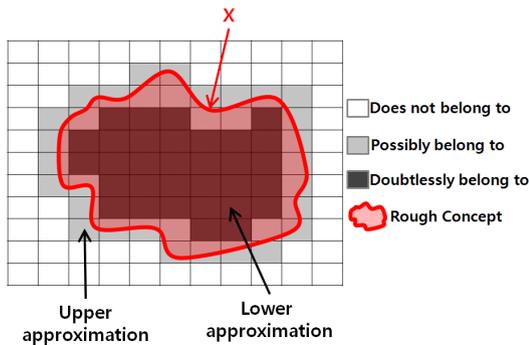
그림4에서는 노인을 나타내는 일반집합과 퍼지집합의 예를 나타내고 있다. 일반 집합의 경우 노인을 나타내는 집합 A는 $\{A|A \geq 35\}$ 와 같이 정의할 수 있다. 이 때, 집합 A에서는 원소들 사이에 명확한 경계를 가지고 있다. 따라서, 35살 보다 크거나 같은 나이를 갖는 사람들은 집합 A의 멤버가 되며, 이러한 조건을 만족하지 못하는 경우에는 집합 A의 멤버가 되지 못한다. 따라서, 34.5살(34살 6개월)인 사람은 집합 A의 멤버가 될 수 없으며, 35살과 85살인 사람들은 같은 노인 집합 A의 원소가 된다. 한편, 퍼지 집합 A의 경우, 각 원소들은 0.0(non membership)부터 1.0(full membership)까지의 멤버십 함수값(귀속도)을 가질 수 있다(그러나, 일반집합의 경우에는 멤버십 함수값으로서 0.0 또는 1.0이 설정되어 있다). 예를 들면, 위의 그림4에서 퍼지집합A의 35살인 사람과 45살인 사람의 멤버십 함수값은 각각 0.6과 0.95이다. 즉, 퍼지집합A에서는 각 사람에 대한 “노인집합의 귀속정도”를 멤버십 함수값으로 설정하거나 가능해 볼 수 있다. 이와 같은 특성을 갖는 퍼지집합은, 주어진 도메인에 관한 불확실/불명확한 데이터로부터 지식을 추출하여 다양한 정보처리를 수행하는 경우에 매우 유용하다.

2.5 러프 집합 이론

데이터의 홍수 속에서 유용한 정보를 추출하기 위해 데이터마이닝이 다양한 분야에서 사용되고 있으며, 2000년대에 들어서면서 데이터의 모호함과 불확실성에 대한 처리를 위하여 러프집합이론(RST : Rough Set Theory[23-25])을 적용한 다양한 기법들도 소개되고 있으며, 데이터마이닝과 같은 지식발견에 있어서 효과적인 기법중의 하나로 인정받고 있다[26].

러프집합이론은 1980년대 초에 Zdzislaw Pawlak에 의

해 불완전성(Incompleteness) 또는 불확실성(Uncertainty)과 같은 특징을 갖는 데이터를 다루기 위해 제안되었다. RST는 어떤 집합에 확실하게 분류되는 하한근사(Lower Approximation)와 불확실하게 분류되는 상한근사(Upper Approximation) 집합 이론을 통해서 나타낸다[27]. 러프 집합이론의 상한근사와 하한근사에 대한 개념은 그림5와 같다. 최근 데이터베이스의 지식 발견, 패턴 인식, 정보 처리, 기업과 재정, 공업과 환경공학, 의료 진단과 의료데이터 분석, 시스템 결합 및 감시와 같은 여러 분야에서 응용되고 있으며, 퍼지와 러프집합을 결합한 연구도 활발하게 이루어지고 있다[28].



[그림 5] 상한근사와 하한근사

3. 형식개념분석기법 기반의 폭소노미 데이터 마이닝

웹2.0의 주요한 특징은 사용자가 능동적으로 참여하여 정보를 생산하고, 이러한 정보를 다른 사용자들과 공유한다는 것이다. 사용자는 자신이 직접 제작한 콘텐츠(UCC : User Created Contents) 또는 다른 사용자가 웹에 공개한 리소스에 태그를 붙여서 정보를 분류하고 공유하는 폭소노미 기반의 웹 애플리케이션이 최근에 다수 등장하고 있다[1-6]. 이와 같은 애플리케이션에는 3항원소(사용자, 태그, 리소스)와 그들 사이의 관계를 나타내는 폭소노미 데이터가 축적되어 있고, 이에 대한 다양한 데이터 마이닝 기법을 적용함으로써 보다 유용한 정보와 지식을 추출하여 활용할 수 있다.

그러나, 다종다양한 성향을 가진 사람들에 의해 만들어지는 폭소노미 데이터는 “애매함”과 “모호성”을 포함할 뿐만 아니라, “불완전성”과 “불확실성”등과 같은 특징을 갖는다[27,29]. 따라서, 본 연구에서는, 폭소노미 데이터들을 다양한 관점으로 분석하여 유용한 정보를 추출하

기 위해 퍼지집합이론(Fuzzy set theory)과 러프집합이론(Rough set theory)에 각각 형식개념분석기법을 융합시킨 퍼지개념분석 및 러프개념분석을 제안하고 이를 토대로 하는 폭소노미 데이터 마이닝 지원도구를 개발하였다.

3.1 퍼지개념분석기법

여기에서는 FST에 관한 기본 개념과 이를 토대로 하는 퍼지개념분석기법(FCA : Fuzzy Formal Concept Analysis)의 제반정의와 지원도구에 대해서 설명한다.

2.4에서 설명한 퍼지집합의 특성을 형식개념분석기법에 적용함으로써, 주어진 데이터와 그 귀속도로부터 숨겨져 있는 지식을 개념단위로 추출하여 구조화 할 수 있다. 퍼지개념분석기법에서는 주어진 도메인의 퍼지데이터(어떤 객체가 갖는 속성과 해당객체의 속성에 대한 멤버십 함수값)를 대상으로 Fuzzy context를 사용하여 입력데이터 테이블로 정의한다.

[정의 1] Fuzzy context

Fuzzy context $K := (G, M, I = \square(G \times M))$ 는 객체들의 집합 G 와 속성들의 집합 M , 그리고 G 와 M 사이의 관계를 나타내는 I 로 구성된다. 단, $(g, m) \in I$ 는 0과 1사이의 귀속도 $\square(g, m)$ 를 갖는다.

[표 3] Fuzzy context

Users \ Tags	art	design	photograph	tutorial	web2.0
bellehammond	0.124	0.118	0	0.118	0
JonFlood	0	0.129	0.038	0	0.105
joonanooj	0	0.012	0.115	0.048	0.006

표3은 Fuzzy context의 예로써, 3명의 사용자(bellehammond, JonFlood, joonanooj)가 사용한 5개의 태그 “art”, “design”, “photograph”, “tutorial”, “web2.0”과 각 태그의 사용 빈도수를 나타낸 것이다. 예를 들어, 사용자 bellehammond는 “art”태그를 12.4%(멤버십 함수값 : 0.124)정도 사용하고 있다. 관심데이터에 대해서만 퍼지개념분석을 수행하기 위해 도메인의 전문가가 임계값(Threshold)을 설정할 수 있으며, 임계값 T 를 기준으로 입력된 데이터를 필터링(filtering)하여 간략화 된 Fuzzy context를 도출해 낼 수 있다. 예를 들어, 사용 빈도수가 낮은 데이터를 제거하기 위해 표3에 임계값 $T=0.1$ 을 적용하여 표4와 같이 임계값에 의해 필터링 된 context를 도출하였다.

Algorithm 1. 퍼지개념 추출 알고리즘

```

//Input: Fuzzy Context  $K := (G, M, I = \square(G \times M))$ 
//      with a confidence threshold T
//Output: The set of C of all fuzzy concepts of K
 $C := \{(FE(M), M)\}$ 
setMembershipValue((FE(M), M))
for each  $g \in G$ 
  for each  $(X, Y) \in C$ 
    Inters:= $Y \cap FI(\{g\})$ 
    if Inters different from any concept intent in C
then
   $C := C \cup \{(FI(Inters), Inters)\}$ 
  setMembershipValue((FI(Inters), Inters))
end if
end for
end for
end for

function setMembershipValue((X, Y))
//Input: A fuzzy concept (X, Y)
for each  $x \in X$ 
  min = 0.0
  for each  $y \in Y$ 
    if(  $(x, y) < min$  ) then
      min =  $(x, y)$ 
      x.membershipValue = min;
    end if
  end for
end for
end for
    
```

주어진 Fuzzy context로부터 퍼지개념들을 추출하고 상·하위개념관계를 파악하여 개념격자구조를 구축하기 위해 다음과 같은 알고리즘들을 개발하였다.

Algorithm 2. 퍼지개념격자 구축 알고리즘

```

//Input: Fuzzy context  $K=(G, M, I = \square(G \times M))$  and  $C=B(K)$ 
//      that is a set of all fuzzy concepts in K
//Output: Fuzzy Concept Lattice  $B(K) = (C, E)$ 
Find C with the algorithm 1
XCoveringEdges(C, K)

function XCoveringEdges(C, K)
//Input: A set of fuzzy concepts C
    
```

```

//      and a fuzzy formal context K
for each  $(X, Y) \in C$ 
  Set count of any concept in C to 0
  for each  $m \in M \setminus Y$ 
    inters :=  $X \cap FE(\{m\})$ 
    Find  $(X1, Y1) \in C$  such that  $X1 =$ 
inters
    count(X1, Y1):= count(X1, Y1) + 1
    if(|Y1|-|Y|) = count(X1, Y1) then
      Add edge  $(X1, Y1) (X, Y)$  to E
    end if
  end for
end for
end for
    
```

3.2 러프개념분석기법

러프개념분석기법(RFCA : Rough Formal Concept Analysis)은 애매모호한 데이터를 다루기 위한 러프집합이론에 형식개념분석기법을 접목하여 구별하기 애매모호한 원소들을 포함하여 하나의 개념으로 추출해서 구조화함으로써, 주어진 데이터를 분류하기 위한 개념분석 기법이다.

러프개념분석기법의 입력 데이터는 표6과 같은 Formal context로 나타내고, 애매모호한 데이터를 다루기 위한 러프집합이론에 형식개념분석기법을 적용하기 위해 러프집합이론의 상한근사와 하한근사 함수를 개념분석에 적합하도록 다음과 같이 정의하였다.

[표 6] Formal context

Tags Users	art	design	photograph	tutorial	web2.0
bellehammond	X	X		X	
JonFlood		X			X
joonanoj			X		

[정의 4] 상한근사와 하한근사

임의의 Formal context $K := (G, M, I)$ 에 대해서, $X \subseteq G, Y \subseteq M$ 일 때, 하한근사(LA)와 상한근사(UA)는 다음과 같이 정의한다.

$$\begin{aligned}
 LA(X) &= \{a \in M \mid OS(a) \subseteq X\}, \\
 UA(X) &= \{a \in M \mid OS(a) \cap X \neq \emptyset\}, \\
 LA(Y) &= \{o \in G \mid AS(o) \subseteq Y\}, \\
 UA(Y) &= \{o \in G \mid AS(o) \cap Y \neq \emptyset\}.
 \end{aligned}$$

단, $OS(a) := \{o \in G \mid (o, a) \in I\}$, $AS(o) := \{a \in M \mid (o, a) \in I\}$.

$LA(X)$ 는 어떤 집합 X 에 확실하게 포함되는 원소들이 갖는 배타적 속성의 집합, 즉, X 에 완전하게 포함되는 객

체들이 배타적으로 갖는 속성들의 집합이다. $UA(X)$ 는 어떤 집합 X 에 불확실하게(또는 불완전하게) 속하는 원소들이 갖는 총체적인 속성의 집합, 즉, X 에 총체적으로 속하는 객체들이 갖는 속성들의 집합이다. 위와 마찬가지로, 어떤 집합 Y 에 대해서, LA 함수와 UA 함수를 적용하면 각각 집합 Y 에 확실하게 포함되는 속성들을 배타적으로 갖는 객체집합과 Y 에 불확실하게 속하는 속성들을 총체적으로 갖는 객체집합을 구할 수 있다. 예를 들어, 표6에 대해서, $X = \{JonFlood, joonanooj\}$, $Y = \{photography, web2.0\}$ 일 때, $LA(X) = \{photography, web2.0\}$ 이고, $UA(X) = \{design, photography, web2.0\}$ 이며, $LA(Y) = \{joonanooj\}$ 이고, $UA(Y) = \{JonFlood, joonanooj\}$ 이다.

LA 와 UA 함수를 사용하여 주어진 Formal context로부터 러프개념을 추출할 수 있다.

[정의 5] 러프개념(Rough Concept)

임의의 Formal context $K := (G, M, I)$ 에 대해서, $X \subseteq G, Y \subseteq M$ 일 때, $X = UA(Y) \wedge Y = LA(X)$ 를 만족하는 (X, Y) 를 러프개념이라고 한다.

예를 들어, 표6에 대해서, $X = \{JonFlood, joonanooj\}$ 일 때, $LA(X) = \{photography, web2.0\}$ 이고, $UA(LA(X)) = UA(\{photography, web2.0\}) = \{JonFlood, joonanooj\}$ 이다. 즉, $X = UA(Y) \wedge Y = LA(X)$ 이므로, $(\{JonFlood, joonanooj\}, \{photography, web2.0\})$ 는 러프개념이다. 이와 같은 방법으로 표6으로부터 8개의 러프개념들이 표7과 같이 추출되었다.

[표 7] 표6으로부터 추출된 러프개념들

ID	Extents	Intents
C1	{ bellehammond, JonFlood, joonanooj }	{ art, design, web2.0, photography, tutorial }
C2	{ bellehammond }	{ art, tutorial }
C3	{ bellehammond, JonFlood }	{ art, design, tutorial, web2.0 }
C4	{ bellehammond, joonanooj }	{ art, photograph, tutorial }
C5	{ joonanooj }	{ photography }
C6	{ JonFlood, joonanooj }	{ photography, web2.0 }
C7	{ JonFlood }	{ web2.0 }
C8	{ }	{ }

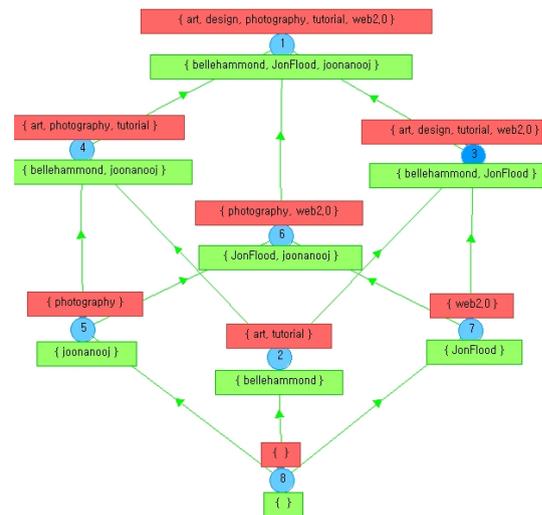
정의5에 의해서 추출된 러프개념들 사이에는 상·하위 개념관계(\leq)가 존재한다.

[정의 6] 상·하위개념관계

임의의 러프개념 $C1 = (X1, Y1), C2 = (X2, Y2)$ 에 대하여, 상·하위개념관계 $(X1, Y1) \leq (X2, Y2)$ 는 다음과 같이 정의된다.

$$(X1, Y1) \leq (X2, Y2) \Leftrightarrow X1 \subseteq X2 (\Leftrightarrow Y1 \supseteq Y2).$$

러프개념 $C5 = (\{joonanooj\}, \{photography\})$ 와 $C6 = (\{JonFlood, joonanooj\}, \{photography, web2.0\})$ 에 대해서, $\{joonanooj\} \subseteq \{JonFlood, joonanooj\} (\Leftrightarrow \{photography\} \supseteq \{photography, web2.0\})$ 이므로, $C5$ 는 $C6$ 의 하위개념(Sub concept)이며, $C5 \leq C6$ 과 같이 표현한다.



[그림 7] 표6에 대한 러프개념격자

위의 정의들을 토대로 구축된 러프개념격자구조는 그림7과 같다. 그림7은 최상위 개념 C1은 주어진 context의 객체와 속성의 전체집합을 나타낸다. 다른 객체들과는 다르게 배타적으로 갖는 속성들에 의해 클러스터링 되며, 각 클러스터들은 객체들과 그들이 배타적으로 갖는 속성들의 쌍으로 구성되고 이와 같은 클러스터들을 러프개념이라고 부른다. C1은 C3, C4, C6로 분류되며, 3개의 러프개념들은 각각 2개의 러프개념들로 분류된다. 예를 들어, C4는 사용자 bellehammond와 joonanooj가 사용자 JonFlood와는 배타적으로 사용한 태그가 “art”, “photograph”와 “tutorial”임을 나타내는 러프개념이다. C4는 C2와 C5로 분류되며, C2 bellehammond가 다른 사용자들과 배타적으로 사용한 태그가 “art”와 “tutorial”임을 나타내고, C5는 joonanooj가 배타적으로 사용한 태그가 “photograph”임을 나타낸다. 이와 같은 제반 정의를 토대로 러프개념 추출과 러프개념격자를 구축하기 위한 알고리즘을 다음과 같이 개발하

였다.

Algorithm 3. 러프개념 추출 알고리즘

```
// Input : a formal context K = (G, M, I)
// Output : a set C of all rough concepts in K
for each a ∈ M
  for each b ∈ M
    if (a == b) then
      X = UA({a});
      Y = LA(X);
    else
      X = UA({a, b});
      Y = LA(X);
    end if
    if (UA(Y) == X)
      for each x ∈ X
        for each y ∈ Y
          Z = ZU{(x, y, μ(x, y))};
        end for
      end for
      C = C U {(X, Y, Z)};
    end if
  end for
end for
```

Algorithm 4. 러프개념격자 구축 알고리즘

```
//Input : a formal context K = (G, M, I) and
//      C that is a set of all rough concepts in K
//Output : the rough concept lattice B(K)=(C, ≤, ∧, ∨)
//      of K
for each c1 ∈ C
  for each c2 ∈ C
    if (c1 != c2) then
      M = meet(c1, c2);
      J = join(c1, c2);
      if(M == c1 && J == c2) then
        for each c3 ∈ C
          if(c1 ≤ c3 && c3 ≤ c2) then
            add an edge from c2 to c1
          end if
        end for
      end if
    end if
  end for
end if
end for
end for
```

3.3 폭소노미 마이닝 지원도구

앞 절의 제반 정의들을 토대로, 각 분석기법을 지원하기 위한 폭소노미 마이닝 지원도구(FMT : Folksonomy Miming Tool)를 개발하였다. FMT의 전체적인 아키텍처

는 그림8과 같이 3계층구조로 구성되어 있으며, 각 계층의 컴포넌트들에 대한 설명은 다음과 같다.

3.3.1 Core Components

컨텍스트 처리 컴포넌트(Context processing component) : 이 컴포넌트는 CSV파일 형태로 표현된 폭소노미 데이터를 입력받아서 분석하고자 하는 관점에 따라 Fuzzy context 또는 Formal context 형태로 표현한다.

개념추출 컴포넌트(Concepts extracting component) : 이 컴포넌트는 주어진 context로부터 개념 및 그들 사이의 관계들을 추출하는 컴포넌트로서, 퍼지개념추출(Fuzzy concept extracting)과 러프개념추출(Rough concept extracting) 모듈들을 포함하고 있다.

개념격자구축 컴포넌트(Concept lattice constructing component) : 개념추출 컴포넌트로부터 추출된 개념들과 그 개념들 사이의 상·하위 관계를 파악하여 개념격자를 구축하는 컴포넌트로서, 퍼지개념격자(Fuzzy Concept Lattice)와 러프개념격자(Rough Concept Lattice) 모듈들을 포함하고 있다.

3.3.2 Internal Data Model

Context : 주어진 도메인으로부터 객체와 속성들을 추출하여 cross table 형태로 표현하고 객체와 속성 사이의 관계를 해당 셀에 “X”표시한다.

개념(Concepts) : 객체와 속성의 쌍(pair)으로써, 공통 속성을 갖는 객체들의 클러스터이다.

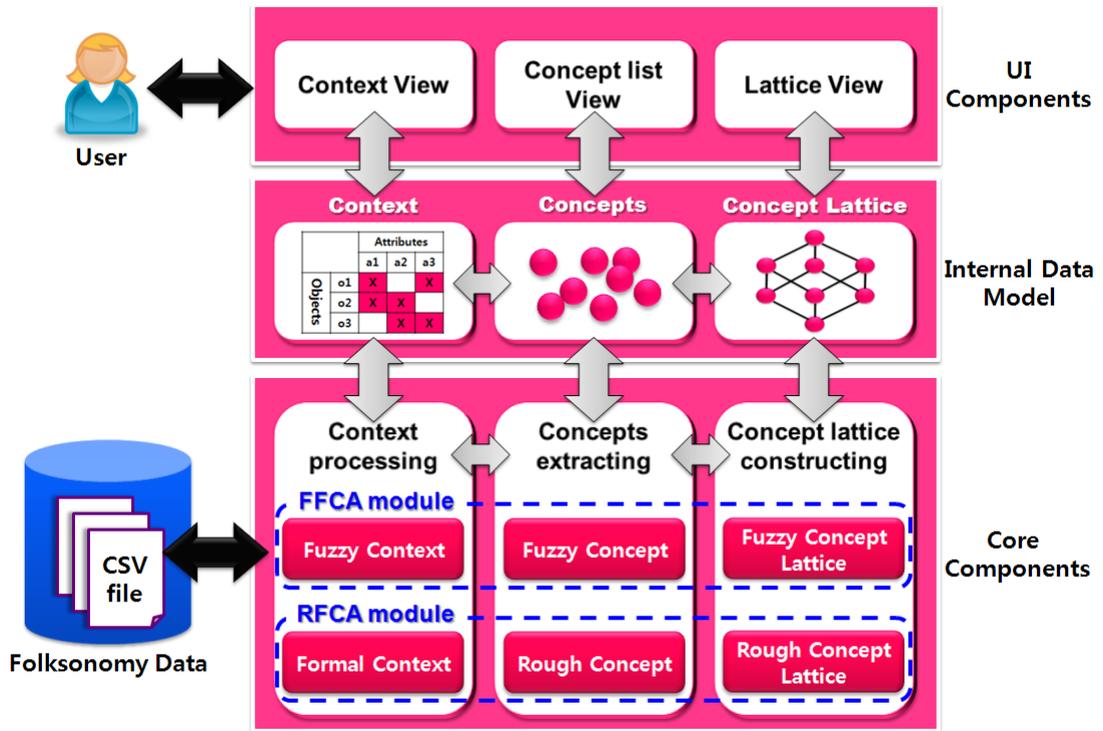
개념격자(Concept Lattice) : 추출된 개념들과 그들 사이의 상·하위 관계를 파악하여 구축된 계층적 개념구조(Conceptual hierarchy)이다.

3.3.3 UI Components

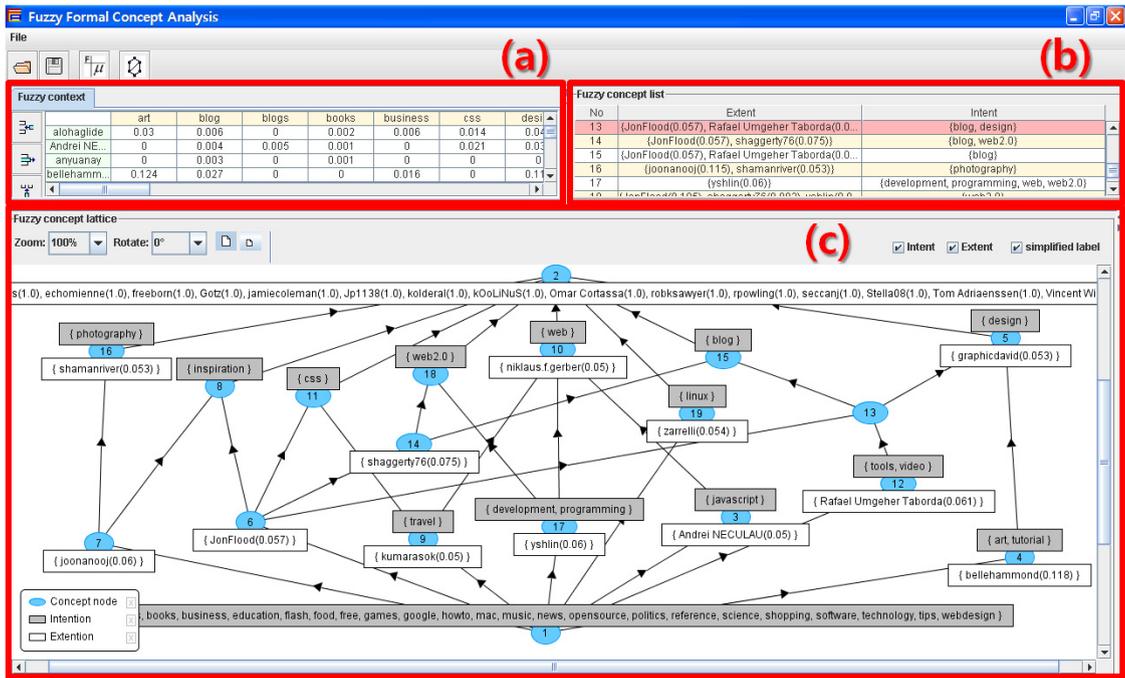
컨텍스트 가시화 컴포넌트(Context View component) : CSV 파일 형태로 입력된 폭소노미 데이터를 컨텍스트 처리 컴포넌트를 적용하여 사용자에게 context 형태로 나타낸다. 또한, context 생성 및 편집 기능도 제공한다.

개념리스트 가시화 컴포넌트(Concept list View component) : 이 컴포넌트는 개념추출 컴포넌트에서 추출된 모든 개념들에 대한 정보를 사용자에게 표7과 같은 형태로 보여준다.

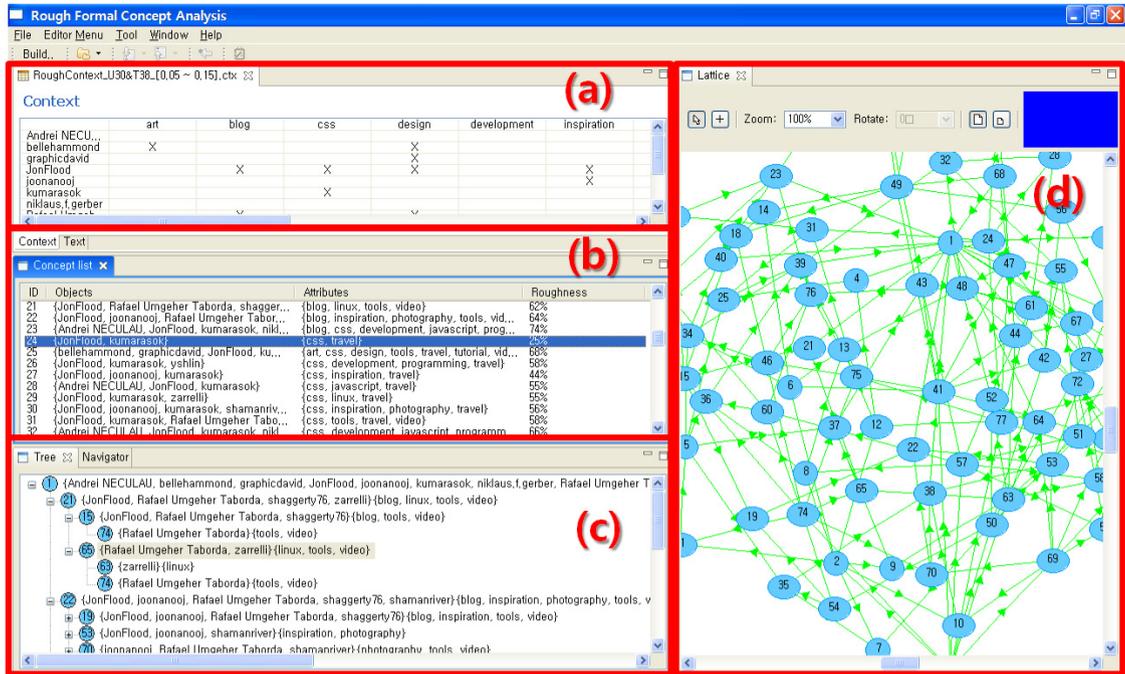
개념격자 가시화 컴포넌트(Lattice View



[그림 8] 폭소노미 마이닝 지원도구(FMT)의 아키텍처



[그림 9] 퍼지개념분석을 지원하는 FFCA 모듈의 실행



[그림 10] 러프개념분석을 지원하는 RFCA 모듈의 실행

component) : 이 컴포넌트는 개념격자구축 컴포넌트에 의해 생성된 개념격자를 가시화하는 컴포넌트이다. 사용자가 데이터를 분석하고자 하는 관점에 따라 입력된 데이터의 분석 결과로써 퍼지개념격자 또는 러프개념격자를 제공한다.

3.3.4 FMT의 사용

FMT는 퍼지개념분석기법을 지원하는 FFCA 모듈과 러프개념분석기법을 지원하는 RFCA 모듈로 구성되어 있다. FFCA 모듈은 그림9과 같이 CSV(Comma Separated Values)파일을 입력하여 Fuzzy context로 변환하거나 Fuzzy context를 생성 및 수정할 수 있도록 컨텍스트처리(Context processing) 기능을 제공하고 있다. 또한, Fuzzy context에 [0, 1]사이의 임계값을 적용하여 불필요한 데이터들을 삭제한 후 임계값이 적용된 그림9의 (a)의 Fuzzy context로부터 모든 퍼지개념들을 추출하여 개념 리스트를 그림9의 (b)와 같이 가시화하고 그들 사이의 관계를 파악하는 퍼지개념 및 관계 추출 기능을 제공한다. 뿐만 아니라, 퍼지개념들과 그들 사이의 관계들을 토대로 퍼지개념격자를 구축하고 그림9의 (c)와 같이 가시화하는 기능도 제공한다.

한편, 러프개념분석을 지원하는 RFCA 모듈의 컨텍스트처리 기능은 CSV 파일을 입력하여 러프개념분석기법

을 적용할 수 있도록 formal context 형태로 변환하는 기능을 하며, context를 생성하거나 수정할 수 있는 기능을 제공한다. 러프개념 및 관계추출 기능은 그림10의 (a)와 같이 입력된 formal context로부터 모든 러프개념들을 추출하여 개념들 사이의 상·하위개념관계를 파악하고 그림10의 (b)와 같이 개념 리스트를 가시화한다. 또한, RFCA 모듈에서는 formal context로부터 추출된 러프개념들과 그들 사이의 관계들을 토대로 그림10의 (c)와 같은 트리형태와 그림10의 (d)와 같은 러프개념격자 형태로 가시화해주는 기능들을 제공하고 있다.

4. 실험 및 결론

본 장에서는, 실제 폭소노미 데이터에 본 연구에서 제안한 형식개념분석기법 기반의 폭소노미 데이터 마이닝 기법을 적용하여 실시한 실험결과에 대해 보고한다.

4.1 퍼지개념분석기법을 이용한 태깅정보 분석 실험

delicious의 전체 태그 클라우드에서 가장 많이 사용된 38개(design, blog, video, ... games)의 태그들을 임의로

[표 8] delicious로부터 추출한 태그 관련 데이터

	art	blog	blogs	books	business	css	design	development	education	flash	food	free	games	google	howto	inspiration	javascript	linux	mac	music	news	opensource
alohaglide	111	22	0	10	23	53	180	83	4	18	4	46	0	6	128	31	11	10	45	37	15	7
Andrei NECULAU	0	5	6	2	0	25	45	6	6	3	0	7	3	0	18	6	58	0	0	0	5	0
anyuanay	0	5	0	2	0	0	0	4	23	0	0	0	0	6	0	0	4	0	0	3	3	0
bellehammond	23	5	0	0	3	0	22	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
dyers	0	8	5	0	11	4	7	14	8	0	4	14	6	2	18	2	2	7	6	10	9	2
echomienne	7	35	10	3	2	0	4	0	23	3	0	15	0	7	0	0	0	0	0	9	4	5
freeborn	26	55	20	4	19	23	81	39	7	6	0	29	10	3	42	33	11	8	28	0	13	6
Gotz	0	346	113	0	45	66	40	0	0	33	0	0	0	140	0	0	69	34	35	34	36	92
graphicdavid	635	216	86	67	107	31	872	0	107	92	26	129	23	25	93	340	0	71	89	74	25	25
jamiecoleman	39	82	28	11	30	25	102	17	8	17	0	33	9	21	26	20	9	0	39	169	24	12
JonFlood	0	12	0	0	0	12	27	0	0	0	2	0	0	0	0	18	0	0	0	5	0	0
joonanoj	0	6	0	0	0	0	2	0	0	0	0	0	0	0	4	10	4	0	0	1	0	0
jp1138	35	148	82	18	0	0	80	14	11	0	0	54	67	9	10	0	30	0	38	31	0	57
kolderal	15	18	9	5	3	2	21	4	2	2	0	2	4	9	4	0	2	2	0	9	4	2
kOoLiNuS	0	121	0	15	18	43	93	88	0	0	0	35	0	38	0	60	0	197	315	14	0	106
kumarasok	0	2	0	0	0	9	3	0	0	0	0	2	0	0	0	1	0	0	0	0	1	0
niklaus.fgerber	0	1	0	0	0	0	3	0	2	6	0	2	0	0	0	7	0	0	2	4	1	0
Omar Cortassa	18	78	23	27	32	62	19	172	14	24	0	114	12	66	55	23	48	103	0	97	22	166
Rafael Umgeher Taborda	28	88	0	0	13	68	0	0	0	3	0	0	0	0	0	0	7	3	0	23	14	8
robksawyer	18	87	19	0	13	116	103	4	105	0	26	0	4	29	50	17	0	27	24	4	49	
rpowling	63	68	34	97	0	0	74	0	150	18	0	82	12	26	21	27	0	0	0	11	14	11
seccanj	0	0	0	2	0	0	10	12	0	1	0	2	1	3	3	0	5	0	0	0	0	3
shaggerty76	0	39	0	0	14	0	6	0	5	0	0	0	0	0	5	0	0	0	0	9	0	0
shamanriver	87	98	45	7	9	96	180	13	0	79	0	61	0	7	34	90	37	0	7	27	8	0
Stella08	13	54	0	8	0	0	27	0	10	20	0	77	18	0	24	0	0	0	59	64	25	0
Tom Adriaenssen	8	93	17	0	8	43	95	141	0	19	0	58	13	12	47	18	57	10	55	18	8	19
Vincent Wlekenkamp	0	0	5	0	0	0	0	0	3	7	0	0	0	1	0	0	0	0	0	0	0	3
Vinod Chandran	4	31	0	10	0	9	30	0	0	7	0	7	6	4	43	0	16	60	26	9	0	3
yshlin	1	11	0	0	0	6	10	25	0	2	1	0	0	2	0	8	10	0	0	0	0	0
zarrelli	0	9	3	5	0	8	11	6	0	0	0	9	0	4	80	0	123	0	0	0	0	15
	photography	politics	programming	reference	science	shopping	software	technology	tips	tools	travel	tutorial	video	web	web2.0	webdesign	Total_frequ	All Tags				
alohaglide	31	8	92	161	15	0	30	6	5	14	0	16	34	145	7	16	3657	559				
Andrei NECULAU	7	0	15	8	0	0	10	0	12	20	0	13	0	59	20	18	1150	388				
anyuanay	0	0	4	2	2	0	4	4	0	0	0	6	0	7	5	0	1308	505				
bellehammond	0	0	1	3	0	0	0	0	0	0	0	22	0	2	0	0	185	29				
dyers	2	0	15	16	2	5	24	10	6	22	10	19	4	17	21	13	1361	550				
echomienne	8	0	0	8	0	2	6	0	0	25	0	8	19	5	63	6	1543	534				
freeborn	37	0	84	48	5	0	31	12	27	27	0	67	21	42	19	48	2917	762				
Gotz	0	0	29	0	0	0	335	109	0	60	0	74	123	1140	164	152	23539	3312				
graphicdavid	268	75	29	270	84	67	118	144	59	133	0	67	150	191	105	154	16313	781				
jamiecoleman	23	19	0	59	56	8	76	35	44	74	13	65	110	77	52	75	6061	1418				
JonFlood	8	0	0	0	0	0	0	2	0	4	0	0	0	0	3	0	22	208	49			
joonanoj	19	0	2	0	0	0	1	0	0	0	0	8	0	1	1	1	165	55				
jp1138	24	0	16	69	13	0	72	0	12	42	0	58	21	15	9	6	4046	714				
kolderal	29	0	8	0	3	3	9	8	2	17	2	11	24	14	37	9	1185	461				
kOoLiNuS	19	0	30	120	0	0	290	33	78	238	0	0	37	73	77	137	11493	1772				
kumarasok	0	0	0	1	0	0	0	0	3	1	0	0	0	8	1	3	158	89				
niklaus.fgerber	3	0	3	5	0	1	4	1	3	9	1	3	8	14	10	12	277	132				
Omar Cortassa	31	0	247	97	11	0	289	37	39	465	23	93	85	251	386	216	10414	1771				
Rafael Umgeher Taborda	0	0	6	10	0	0	26	0	0	75	0	9	102	0	0	15	1105	88				
robksawyer	8	0	135	87	5	0	60	17	12	59	0	84	52	22	0	39	3630	471				
rpowling	43	0	0	80	17	0	38	81	14	149	0	12	44	90	175	40	5366	943				
seccanj	0	0	12	0	0	0	3	0	0	1	0	6	0	3	3	1	264	129				
shaggerty76	0	3	0	3	0	0	0	8	3	6	0	1	2	0	43	3	518	112				
shamanriver	289	6	10	25	0	6	39	7	22	62	0	74	41	58	64	114	5353	1128				
Stella08	9	7	0	53	0	8	9	20	25	41	0	0	53	13	16	0	2015	178				
Tom Adriaenssen	47	0	100	26	0	0	100	11	13	75	0	25	47	58	19	55	5440	1255				
Vincent Wlekenkamp	0	0	0	1	3	0	13	3	0	0	0	0	0	0	27	0	695	164				
Vinod Chandran	15	0	25	8	2	0	5	0	8	12	3	2	5	8	2	26	1388	392				
yshlin	4	0	33	2	0	0	3	1	0	0	0	0	0	24	26	19	400	112				
zarrelli	0	0	11	45	0	3	31	3	26	35	0	38	9	19	0	8	2264	708				



[그림 11] 멤버십 함수값의 분포도와 임계값 설정

추출한 30명의 사용자들이 어느 정도 사용하고 있는지에 대한 정보를 토대로 태그의 사용횟수, 사용한 모든 태그들 그리고 사용한 모든 태그들의 사용빈도수를 추출한다. 이때 추출된 데이터들은 임의로 추출한 30명의 사용자들 객체집합으로, 사용 빈도수가 높은 38개의 태그들을 속성집합으로, 각 해당 셀에는 사용자가 해당 태그를 사용한 횟수로 구성된다. 이렇게 구성된 표8에 다음과 같은 멤버십 함수를 적용하여 표9와 같은 Fuzzy context를 생성하였다.

$$\mu(g, m) = \frac{\text{freq}(g, m)}{\text{total_freq}(g)}$$

위의 멤버십 함수는 각 태그가 사용된 빈도수를 사용한 모든 태그들의 사용빈도수로 나눔으로서, 사용자가 각 태그를 어느 정도 사용하고 있는지에 대한 멤버십 함수값을 0.0부터 1.0사이의 값으로 표현한다.

그림11은 표9에 대한 멤버십 함수값의 분포도이다. 그림11에서 보는 바와 같이 대부분의 사용자들은 특정 태그를 빈번하게 사용하기 보다는 다양한 태그들을 적은 횟수로 사용하고 있음을 알 수 있다. 이와 같은 멤버십 함수값의 분포도를 토대로 사용 빈도수가 적은 태그들을 제거하기 위해 임계값 $T(0.05 \leq T \leq 0.15)$ 를 설정하고, 본 연구에서 개발된 FMT를 사용하여 표9에 대해 퍼지개

[표 9] 멤버십 함수를 적용하여 생성된 Fuzzy context

	art	blog	blogs	books	business	css	design	development	education	flash	food	free	games	google	howto	inspiration	javascript	linux	mac	music
alohaglide	0.03	0.006	0	0.002	0.006	0.014	0.049	0.022	0.001	0.004	0.012	0	0.001	0.001	0.035	0.008	0.003	0.002	0.012	0.001
Andrei NECULAU	0	0.004	0.005	0.001	0	0.021	0.039	0.005	0.005	0.002	0	0.006	0.002	0	0.015	0.005	0.005	0	0	0
anyuanay	0	0.003	0	0.001	0	0	0	0.003	0.017	0	0	0	0	0.004	0	0	0.003	0	0	0.002
bellehammond	0.124	0.027	0	0	0.016	0	0.118	0	0	0.005	0	0	0	0	0	0	0	0	0	0
dyers	0	0.005	0.003	0	0.008	0.002	0.005	0.01	0.005	0	0.002	0.01	0.004	0.001	0.013	0.001	0.001	0.005	0.004	0.007
echomienne	0.004	0.022	0.006	0.001	0.001	0	0.002	0	0.014	0.001	0	0.009	0	0.004	0	0	0	0	0	0.005
freeborn	0.008	0.018	0.006	0.001	0.006	0.007	0.027	0.013	0.002	0.002	0	0.009	0.003	0.001	0.014	0.011	0.003	0.002	0.009	0
Gotz	0	0.014	0.004	0	0.001	0.002	0.001	0	0	0.001	0	0	0	0.005	0	0	0.002	0.001	0.001	0.001
graphicdavid	0.038	0.013	0.005	0.004	0.006	0.001	0.053	0	0.006	0.005	0.001	0.007	0.001	0.001	0.005	0.02	0	0	0.004	0.005
jamiecoleman	0.006	0.013	0.004	0.001	0.004	0.004	0.016	0.002	0.001	0.002	0	0.005	0.001	0.003	0.004	0.003	0.001	0	0.006	0.027
JonFlood	0	0.057	0	0	0	0.057	0.129	0	0	0	0.009	0	0	0	0	0	0.086	0	0	0.024
joonanoj	0	0.036	0	0	0	0	0.012	0	0	0	0	0	0	0	0.024	0.06	0.024	0	0	0.006
Jp1138	0.008	0.036	0.02	0.004	0	0	0.019	0.003	0.002	0	0	0.013	0.016	0.002	0.002	0	0	0.007	0	0.009
kolderal	0.012	0.015	0.007	0.004	0.002	0.001	0.017	0.003	0.001	0.001	0	0.001	0.003	0.007	0.003	0	0.001	0.001	0	0.007
kOoLiNuS	0	0.01	0	0.001	0.001	0.003	0.008	0.007	0	0	0.003	0	0.003	0	0.005	0	0.017	0.027	0.001	0.001
kumarasok	0	0.012	0	0	0.056	0.018	0	0	0	0	0.012	0	0	0	0.006	0	0	0	0	0
niklaus.fgerber	0	0.003	0	0	0	0	0.01	0	0.007	0.021	0	0.007	0	0	0.025	0	0	0	0.007	0.014
Omar Cortassa	0.001	0.007	0.002	0.002	0.003	0.005	0.018	0.016	0.001	0.002	0	0.01	0.001	0.006	0.005	0.002	0.004	0.009	0	0.009
Rafael Umgeher Taborda	0.025	0.079	0	0	0.011	0.061	0	0	0.002	0	0	0	0	0	0	0	0.006	0.002	0	0.02
robksawyer	0.004	0.023	0.005	0	0.003	0.031	0.028	0.028	0.001	0.028	0	0.007	0	0.001	0.007	0.013	0.004	0	0.007	0.006
rpowling	0.011	0.012	0.006	0.018	0	0	0.013	0	0.027	0.003	0	0.015	0.002	0.004	0.003	0.005	0	0	0	0.002
seccan	0	0	0.007	0	0	0.037	0.045	0	0.003	0	0.007	0.003	0.011	0.011	0	0.018	0	0	0	0
shaggerty76	0	0.075	0	0	0.027	0	0.011	0	0.009	0	0	0	0	0	0.009	0	0	0	0	0
shamanriver	0.016	0.018	0.008	0.001	0.001	0.017	0.033	0.002	0	0.014	0	0.011	0	0.001	0.006	0.016	0.006	0	0.001	0.005
Stella08	0.006	0.026	0	0.003	0	0	0.013	0	0.004	0.009	0	0.038	0.008	0	0.011	0	0	0	0.029	0.031
Tom Adriaenssen	0.001	0.017	0.003	0	0.001	0.007	0.017	0.025	0	0.003	0	0.01	0.002	0.002	0.008	0.003	0.01	0.001	0.01	0.003
Vincent Wiekenkamp	0	0	0.007	0.007	0	0	0	0	0.004	0	0	0	0	0.001	0	0	0	0	0	0
Vinod Chandran	0.002	0.022	0	0.007	0	0.006	0.021	0	0.005	0	0.005	0.004	0.002	0.03	0	0.011	0.043	0.018	0.006	0.006
yshlin	0.002	0.027	0	0	0.015	0.025	0.062	0	0.005	0.002	0	0.005	0	0	0.02	0.02	0	0	0	0
zarrelli	0	0.003	0.001	0.002	0	0.003	0.004	0.002	0	0	0.003	0	0.001	0.001	0.035	0	0	0.054	0	0
	news	opensource	photography	politics	programming	reference	science	shopping	software	technology	tips	tools	travel	tutorial	video	web	web2.0	webdesign		
alohaglide	0.004	0.001	0.008	0.002	0.025	0.044	0.004	0	0.008	0.001	0.001	0.003	0	0.004	0.009	0.039	0.001	0.004		
Andrei NECULAU	0.004	0	0.006	0	0.013	0.006	0	0	0.008	0	0.01	0.017	0	0.011	0	0.051	0.017	0.015		
anyuanay	0.002	0	0	0	0.003	0.001	0.001	0	0.003	0.003	0	0	0	0.004	0	0.005	0.003	0		
bellehammond	0	0	0	0	0.005	0.016	0	0	0	0	0	0	0	0	0.118	0	0.01	0		
dyers	0.006	0.001	0.001	0	0.011	0.011	0.001	0.003	0.017	0.007	0.004	0.016	0.007	0.013	0.002	0.012	0.015	0.009		
echomienne	0.002	0.003	0.005	0	0	0.005	0	0.001	0.003	0	0	0.016	0	0.005	0.012	0.003	0.04	0.003		
freeborn	0.004	0.002	0.012	0	0.028	0.016	0.001	0	0.01	0.004	0.009	0.009	0	0.022	0.007	0.014	0.006	0.016		
Gotz	0.001	0.003	0	0	0.001	0	0	0	0.014	0.004	0	0.002	0	0.003	0.005	0.048	0.006	0.006		
graphicdavid	0.004	0.001	0.016	0.004	0.001	0.016	0.005	0.004	0.007	0.008	0.003	0.008	0	0.004	0.009	0.011	0.006	0.009		
jamiecoleman	0.003	0.001	0.003	0.003	0	0.009	0.009	0.001	0.012	0.005	0.007	0.012	0.002	0.01	0.018	0.012	0.008	0.012		
JonFlood	0	0	0.038	0	0	0	0	0	0	0.009	0	0.019	0	0	0.014	0.105	0			
joonanoj	0	0	0.115	0	0.012	0	0	0	0.006	0	0	0	0	0.048	0.006	0.006	0.006	0.006		
Jp1138	0.007	0.014	0.005	0	0.003	0.017	0.003	0	0.017	0	0.002	0.01	0	0.014	0.005	0.003	0.002	0.001		
kolderal	0.003	0.001	0.024	0	0.006	0	0.002	0.002	0.007	0.006	0.001	0.014	0.001	0.009	0.02	0.011	0.031	0.007		
kOoLiNuS	0	0.009	0.001	0	0.002	0.01	0	0.025	0.002	0.006	0.02	0	0	0.003	0.006	0.006	0.011			
kumarasok	0.006	0	0	0	0	0.006	0	0	0	0.018	0.006	0.05	0	0	0.05	0.006	0.018			
niklaus.fgerber	0.003	0	0.01	0	0.01	0.018	0	0.003	0.014	0.003	0.01	0.032	0.003	0.01	0.028	0.05	0.036	0.043		
Omar Cortassa	0.002	0.015	0.002	0	0.023	0.009	0.001	0	0.027	0.003	0.003	0.044	0.002	0.008	0.008	0.024	0.037	0.02		
Rafael Umgeher Taborda	0.012	0.007	0	0	0.005	0.009	0	0	0.032	0	0.067	0	0.008	0.008	0.092	0	0	0.013		
robksawyer	0.001	0.013	0.002	0	0.037	0.023	0.001	0	0.016	0.004	0.003	0.016	0	0.023	0.014	0.006	0	0.01		
rpowling	0.002	0.002	0.008	0	0	0.014	0.003	0	0.007	0.015	0.002	0.027	0	0.002	0.008	0.016	0.032	0.007		
seccan	0	0.011	0	0	0.045	0	0	0	0.011	0	0	0.003	0	0.022	0	0.011	0.011	0.003		
shaggerty76	0.017	0	0	0.005	0	0.005	0	0	0	0.015	0.005	0.011	0	0.001	0.003	0	0.083	0.005		
shamanriver	0.001	0	0.053	0.001	0.001	0.004	0	0.001	0.007	0.001	0.004	0.011	0	0.013	0.007	0.01	0.011	0.021		
Stella08	0.012	0	0.004	0.003	0	0.026	0	0.003	0.004	0.009	0.012	0.02	0	0	0.026	0.006	0.007	0		
Tom Adriaenssen	0.001	0.003	0.008	0	0.018	0.004	0	0	0.018	0.002	0.002	0.013	0	0.004	0.008	0.01	0.003	0.01		
Vincent Wiekenkamp	0	0.004	0	0	0	0.001	0.004	0	0.018	0.004	0	0	0	0	0	0	0.038	0		
Vinod Chandran	0	0.002	0.01	0	0.018	0.005	0.001	0	0.003	0	0.005	0.008	0.002	0.001	0.003	0.005	0.001	0.018		
yshlin	0.01	0	0.01	0	0.082	0.005	0	0	0.007	0.002	0	0	0	0	0.06	0.065	0.047			
zarrelli	0	0.0																		

[표 10] 표9로부터 변환된 Formal context

	art	blog	css	design	development	inspiration	javascript	linux	photography	programming	tools	travel	tutorial	video	web	web 2.0
Andrei NECULAU							X								X	
bellehammond	X			X									X			
graphicdavid				X												
JonFlood		X	X	X		X										X
joonanoj						X			X							
kumarasok			X									X				X
niklaus.f.gerber																X
Rafael mgeher Taborda	X			X							X			X		
shaggerty76	X															X
shamanriver									X							
yshlin					X					X					X	X
zarrelli								X								

념분석기법을 적용하는 실험을 수행하였다.

임계값 T가 적용된 context로부터 19개의 퍼지개념들이 추출되었고, 추출된 퍼지개념들 사이에 순서관계를 파악하여 그림9의 (c)와 같이 퍼지개념격자를 구축하였다. 퍼지개념격자는 FCA의 개념격자구조에서와 같이, 각 정점은 퍼지개념들을 나타내고 정점들 사이의 링크들은 개념들 사이의 상·하위개념관계를 나타내며, 객체는 아래서 위로 전파되고 속성은 위에서 아래로 계승되어진다. 이와 같은 퍼지개념격자로부터 사용자들이 공통적으로 사용하고 있는 태그들을 파악할 수 있을 뿐만 아니라, 멤버십 함수값을 토대로 각 사용자가 공통 태그를 어느 정도 사용하고 있는지에 대한 유용한 정보도 수월하게 파악해 낼 수 있다.

예를 들어, 그림12와 같은 퍼지개념격자에서 퍼지개념 C18은 3명의 사용자(JonFlood, shaggerty76, syhlin), 그리고 이들이 공통적으로 사용하고 있는 “web2.0” 태그로 구성되어 있다. 이때, 사용자 JonFlood, shaggerty76, syhlin는 각각 10.5%, 8.3%, 6.5%(멤버십 함수값 : 0.105, 0.083, 0.065)의 태그사용빈도로 “web2.0” 태그를 사용하고 있다. 또한, 퍼지개념C14는 상위개념C15와 C18로부터 계승된 태그 “blog”와 “web2.0”과 하위개념C6로부터 전파된 사용자 JonFlood와 퍼지개념C14 고유의 사용자 shaggerty76으로 구성되어 있다. 즉, C14는 사용자가 JonFlood와 shaggerty76이 공통으로 사용한 태그는 “blog”와 “web2.0”임을 나타내는 퍼지개념이다. JonFlood는 “blog”와 “web2.0” 태그들을 최소 5.7%(멤버십 함수값 : 0.057) 이상 사용하고 있으며, shaggerty76은 최소 7.5%(멤버십 함수값 : 0.075) 이상 사용하고 있음을 파악해 낼 수 있다. 퍼지개념C13도 같은 방법으로 “blog”와 “design”을 공통적으로 사용하고 있는 사용자는 JonFlood와 Rafael Umgeher Taborda이고 각각 최소 5.7%와 6.1% 이상 공통태그를 사용하고 있음을 알 수 있다. 뿐만 아니

라, 퍼지개념격자로부터 「태그 “css”, “design”과 “inspiration”을 사용한 사용자는 태그 “blog”와 “web2.0” 역시 사용한다」 라는 사용자의 태깅에 관한 함의관계(implication)를 추가적으로 추출해 낼 수 있다.

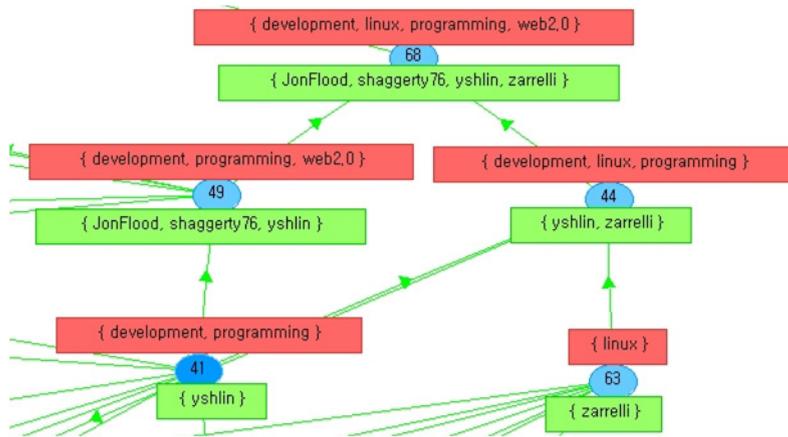
이와 같은 방법에 따라서, 폭소노미 데이터에 퍼지개념분석기법을 적용하여 구축된 퍼지개념격자로부터, 사용자들의 공통태그와 최소 사용빈도에 대한 유용한 정보를 추출할 수 있을 뿐만 아니라, 사용자의 태깅에 관한 함의관계도 추출해 낼 수 있다. 이와 같은 제반정보를 토대로, 폭소노미 기반 애플리케이션 사용자들의 태깅 성향을 파악해 볼 수 있다.

4.2 러프개념분석기법을 이용한 사용자 분류 실험

러프개념분석기법의 입력 데이터는 Formal context이므로, 4.1에서 사용한 같은 데이터에 대해서 러프개념분석기법을 적용하기 위해서는 Fuzzy context를 Formal context로 변환 할 필요가 있다.

표9에 임계값 $T(0.05 \leq T \leq 0.15)$ 를 적용하여 임계값 범위 외부에 존재하는 멤버십 관계를 모두 제거한 후 Formal context로 변환하였다. 또한, 이와 같이 변환된 context로부터 38개의 태그 중 한 개의 태그도 사용하고 있지 않는 사용자들과 30명의 사용자에게 의해 한 번도 사용되지 않는 태그들을 삭제하면 표10과 같다. 표10은 12명의 사용자들이 16개의 태그들 중 어떤 태그를 사용하고 있는지에 대한 정보를 나타낸다. 예를 들어, 사용자 Andrei NECULAU는 “javascript”와 “web” 태그를 사용한다. 표10과 같은 context에 러프개념분석기법을 적용한 결과, 79개의 러프개념들이 추출되었으며, 개념들 사이에 포함관계를 파악하여 그림10의 (d)와 같이 러프개념격자 구조를 구축하였다.

그림13에서, C68은 사용자 JonFlood, shaggerty76,



[그림 13] 표10에 대한 러프개념격자(일부분)

yshlin, zarrelli가 다른 사용자와는 구별되게 "development", "linux", "programming", "web2.0" 태그를 배타적으로 사용하고 있음을 나타내는 개념이다. C68은 하위 개념C49와 C44로 분류되고, C49는 "development", "programming", "web2.0" 태그를 배타적으로 사용하는 JonFlood, shaggerty76과 yshlin로 구성된 사용자 그룹을 나타내는 개념이고, C44는 "development", "linux", "programming" 태그를 배타적으로 사용하는 사용자 yshlin과 zarrelli로 구성된 사용자 그룹을 나타내는 개념이다. C49는 C41로 분류되고, C44는 C41과 C63으로 다시 분류된다. C49와 C44로부터 추출된 C41은 사용자 yshlin가 다른 사용자들과 구별되게 "development"와 "programming" 태그를 배타적으로 사용하고 있음을 나타내며, C63은 사용자 zarrelli가 배타적으로 사용하고 있는 태그는 "linux"임을 나타내고 있다.

이와 같이 del.icio.us로부터 추출한 사용자와 태그 정보에 대해 러프개념분석기법을 적용하여 구축된 러프개념격자로부터 사용자가 배타적으로 사용하고 있는 태그들을 수월하게 파악해 낼 수 있다. 뿐만 아니라, 다른 사용자와 구별되게 배타적으로 사용된 태그에 의해서 사용자들을 그룹핑하여 클러스터링함으로써, 전체 사용자들의 태그사용에 관한 성향을 기반으로 하는 분류 및 체계화가 가능하다.

4.3 결론 및 향후 연구 과제

본 연구에서는 폭소노미 기반의 웹 애플리케이션에서 발생되고 있는 태깅 데이터를 다양한 관점에서 분석하여 유용한 정보를 추출하기 위한 형식개념분석 기반의 폭소노미 데이터 마이닝 기법을 제안하였으며, 이를 지원하기 위한 폭소노미 마이닝 지원도구(FMT)를 개발하였다. 또

한, 제안한 기법과 FMT의 유용성을 검증하기 위하여, 폭소노미 기반 시스템으로 널리 사용되고 있는 del.icio.us의 데이터를 대상으로 실험을 수행하고, 그 결과를 보고하였다.

delicious로부터 추출한 30명의 사용자들과 38개의 태그 데이터를 대상으로 퍼지개념분석기법을 적용한 실험 결과 총 19개의 퍼지개념들을 추출하여 퍼지개념격자를 구축하였다. 한편, delicious로부터 추출된 12명의 사용자들과 16개의 태그를 대상으로 러프개념분석기법을 적용한 실험결과 총 79개의 러프개념들을 추출하여 러프개념격자를 구축하였다.

퍼지개념분석기법을 적용하여 구축된 퍼지개념격자들로 토대로 공통적으로 사용한 태그에 의해 분류된 사용자들과 각 사용자가 공통태그를 어느 정도 사용하고 있는지에 관한 수치정보를 수월하게 분석해 낼 수 있다. 한편, 러프개념분석기법은 퍼지개념분석기법의 분류체계와 다르게 다른 사용자들과 배타적으로 사용한 태그에 의해 사용자들을 분류하는 새로운 분류체계를 제공하며, 퍼지개념분석기법에 의해서는 추출해 낼 수 없는 유용한 개념들을 추출하여 분류해 낼 수 있다.

퍼지개념분석기법은 수치값을 갖는 데이터를 처리하는데 적합하며, 입력된 Fuzzy context에 대해서 도메인 전문가가 임계값을 설정할 수 있다. 임계값에 의해서 필터링 된 Fuzzy context로부터 퍼지개념들을 추출하고 그들 사이에 상·하위개념관계를 파악하여 퍼지개념격자를 구축할 수 있다. 구축된 개념격자로부터 객체들이 공통적으로 갖는 속성을 추출해내고 그러한 공통속성을 각 객체가 어느 정도 함유하고 있는지에 대한 귀속도 정보까지도 분석해 낼 수 있다. 이와 같은 퍼지개념분석기법을 대량의 폭소노미 데이터에 적용함으로써, 웹 사용자의 선

호도 및 성향에 대해서 분석해 낼 수 있을 뿐만 아니라 Social Network를 분석하는 기반이 될 수 있다.

한편, 러프개념분석기법은 러프집합이론의 특징인 상하근사와 하하근사를 토대로 주어진 데이터로부터 러프 개념들을 추출해내고 러프개념격자를 구축하는 일련의 과정을 제공한다. 러프개념분석기법은 배타적으로 갖는 속성에 의해 객체들을 하나의 그룹으로 클러스터링하고 분류하는 새로운 분류체계를 제공한다. 이러한 특징을 기반으로, 의료데이터 또는 웹 데이터와 같은 대량의 데이터를 분류할 때 새로운 분류체계로써 널리 사용될 수 있으며, 배타적 속성에 의한 새로운 분류체계를 기반으로 하는 검색엔진에서 유용하게 활용될 수 있다.

참고문헌

[1] <http://delicious.com>
 [2] <http://www.flickr.com>
 [3] <http://www.youtube.com>
 [4] <http://www.bibsonomy.org>
 [5] <http://www.connotea.org>
 [6] <http://www.citeulike.org>
 [7] P. N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, Pearson, 2006.
 [8] M. Kamber and J. Han, Data Mining(Concepts and Techniques), Morgan Kaufmann Pub, 2006.
 [9] 최영희, 데이터마이닝을 이용한 데이터 활용에 관한 연구, 조선대학교, 전산통계학과, 석사학위논문, 2001.
 [10] 김은수, 웹 마이닝을 이용한 개인 광고 시스템 설계 및 구현, 한남대학교대학원, 컴퓨터공학과, 박사학위논문, 2004.
 [11] 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 at ISWC/ASWC 2007, pp. 108~121, 2007.
 [12] Robert J'aschke, Andreas Hotho, Christoph Schmitz, Bernhard Ganter, and Gerd Stumme, "Discovering shared conceptualizations in folksonomies", Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 6, Pages 38-53, 2008.
 [13] Suk-Hyung Hwang, Yu-Kyung Kang, "Applying Hierarchical Classes Analysis to Triadic context for Folsonomy Mining", 2007 International Conference on Convergence Informati on Technology(ICCIT'07), pp.103-109, 2007.
 [14] Hong-Gee Kim, Suk-Hyung Hwang, Yu-Kyung Kang,

Hak-Lae Kim, and Hae-Sool Yang, "An Agent Environment for Contextualizing Folksonomies in a Triadic Context", First KES International Symposium, KES-AMSTA2007, LNAI4496, pp.728-737, 2007. 5.
 [15] 이강표, 김현우, 장충수, 김형주, "FolksoViz: Wikipedia 본문을 이용한 상하위 관계 기반 폭소노미 시각화 기법", 정보과학회논문지:컴퓨팅의 실제 및 레터 제14권 제4호, 2008. 6.
 [16] B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer, 1999.
 [17] C. Carpineto, G. Romano, Concept Data Analysis: Theory and Applications, Wiley, September, 2004.
 [18] 안수민, 퍼지 집합 이론을 이용한 아파트입지특성 평가에 관한 연구, 중앙대학교, 석사학위논문, 2003.
 [19] 최재만, 퍼지이론과 확률론의 불확실성에 대한 고찰, 수원대학교 교육대학원 석사학위논문, 2007.
 [20] R. Lowen, Fuzzy Set Theory: Basic Concepts, Techniques and Bibliography, Springer, 1996.
 [21] H.J. Zimmermann, Fuzzy Set Theory and Its Applications, Germany : Kluwer Academic, 1996.
 [22] J. Jantzen, Foundations of Fuzzy Control, WileyBlackwell; New title edition, 2007.
 [23] Z. Pawlak, Rough Sets : Theoretical Aspects of Reasoning about Data, Springer, 1991.
 [24] B. Walczak and D. L. Massart, "Tutorial Rough sets theory", Chemometrics and Intelligent Laboratory Systems, Vol. 47, pp. 1-16, 1999.
 [25] J. Komorowski, L. Polkowsik and A. Skowron, Rough Sets : A Tutorial, in:S.K. Pal, A. New Trend in Decision Making, Springer, Singapore, pp. 1-98, 1999.
 [26] S. D. Jitender, V. V. Raghavan, A. Sarkar and H. Sever, "Data Mining Trends in Research and Development", Rough Sets and Data Mining analysis of imprecise data, Kluwer Academic publishers, pp. 9-45, 1997.
 [27] 유재진, 김재련, "러프 집합을 이용한 의사결정나무의 노드 선택 방법", 한국산업경영시스템학회 2003 추계 학술대회 논문집, pp.41-46, 2003.
 [28] 김형수, 김흥기, 이상부, "확률적 러프집합 이론에 기반한 퍼지 정보의 검색", 정보과학회논문지(B), 제25 권 제9호, pp.1431-1441, 1998.
 [29] 서완석, 김재련, "러프집합이론과 SOM을 이용한 연속형 속성의 이산화", Journal of the Society of Korea Industrial and Systems Engineering, Vol. 28, No. 1, pp.1-7, 2005.

강 유 경(Yu-Kyung Kang)

[정회원]



- 2003년 2월 : 선문대학교 컴퓨터 정보학부 (이학사)
- 2005년 8월 : 선문대학교 일반대학원 전자계산학과 (이학석사)
- 2009년 2월 : 선문대학교 일반대학원 컴퓨터정보학과 (이학박사)
- 2009년 3월 ~ 현재 : 선문대학교 컴퓨터공학부 BK21 박사후 과정생

<관심분야>

Formal Concept Analysis, Semantic Web Mining, 시맨틱 웹, 소프트웨어공학 등

황 석 형(Suk-Hyung Hwang)

[정회원]



- 1991년 8월 : 강원대학교 전자계산학과 조기졸업 (이학사)
- 1993년 4월 : 일본 오사카대학교 대학원 정보공학과 (공학석사)
- 1997년 4월 : 일본 오사카대학교 대학원 정보공학과 (공학박사)
- 1997년 3월 ~ 현재 : 선문대학교 컴퓨터공학부 전임강사, 조교수, 부교수

- 2007년 1월 ~ 2008년 1월 : 아일랜드국립대학교 DERI 객원연구원

<관심분야>

소프트웨어공학, 객체지향, 온톨로지공학, 시맨틱 웹, Formal Concept Analysis, Semantic Web Mining

양 해 술(Hea Sool Yang)

[정회원]



- 1975년 2월 : 홍익대학교 전기공학과 졸업(학사)
- 1878년 2월 : 성균관대학교 정보처리학과(석사)
- 1991년 2월 : 日本 오사카대학 정보공학과 SW공학전공(공학박사)
- 1975년 3월 ~ 1979년 12월 : 육군중앙경리단 전산장교

- 1980년 3월 ~ 1995년 2월 : 강원대학교 전자계산학과 교수
- 1986년 1월 ~ 1987년 12월 : 日本 오사카대학 객원연구원
- 1995년 3월 ~ 2002년 12월 : 한국S/W품질연구소 소장
- 1999년 3월 ~ 현재 : 호서대학교 벤처전문대학원 교수

<관심분야>

소프트웨어공학(특히, S/W 품질보증과 평가 및 품질관리, 프로젝트관리, CBD기반기술, IT품질경영