

# FOAF 및 소셜 네트워크 분석을 이용한 핫 이슈 추출 기법

왕청\*, 손종수\*, 권경락\*, 정인정\*

\*고려대학교 컴퓨터정보학과

e-mail:{wangqing, mis026, helpnara, chung}@korea.ac.kr

## Hot issue extraction method using FOAF and Social Network Analysis

Wang Qing\*, Sohn Jongsoo\*, Kwon Kyunglak\*, InJeong Chung\*

\*Dept of Computer and Information Science, Korea University

### 요 약

웹 2.0의 적극적인 도입에 따라 소셜 네트워크 기반 커뮤니티 사이트에서는 관련된 콘텐츠를 적절하게 추천하는 것은 중요한 문제로 부각되고 있으며 이로 인해 사용자들의 동향 및 이슈 추출 기법이 중요하게 작용하고 있다. 이르기 위해서 지금까지의 연구에서는 콘텐츠에 포함된 키워드 매칭 방법을 이용하고 있으나 사용자들 간의 연결 관계와 키워드의 중요도를 고려하지 못하고 있다.

본 논문에서는 FOAF 기반의 소셜 네트워크와 del.icio.us에서 제공하는 소셜 북마크 데이터를 기초로 소셜네트워크 분석을 보이며 이를 통한 사용자들 사이에서 중요하게 부각되는 핫 이슈를 추출하는 방법을 제안한다. 본 논문에서 제안하는 핫 이슈 추출 방법을 활용하면 사용자들의 관심 분야 동향과약을 효율적으로 수행할 수 있으며 이를 통해 맞춤형 마케팅 및 콘텐츠 추천이 가능해 진다.

### 1. 서론

최근 웹 2.0은 웹의 진화에 있어 중요한 키워드로 부각되고 있으며 많은 소셜 네트워크 기반 커뮤니티 사이트들은 웹 2.0의 개념을 도입하고 있다[1]. 웹 2.0을 통한 인터넷의 변화는 콘텐츠의 측면에서 바라볼 때 콘텐츠 생산 주체의 변화를 들 수 있다. 전통적인 웹에서는 콘텐츠를 서비스 제공자가 생산했던 반면 최근의 웹에서는 콘텐츠를 서비스 수요자가 생산하고 있다[2]. 이에 따라 좋은 콘텐츠를 더 많이 생산하는 사용자가 더 중요한 비중을 차지하게 되었다. 그리고 사용자들은 서로 '친구'관계를 맺어 상호간에 콘텐츠를 공유하면서 커뮤니티를 키워나가는 경향이 생기게 되었다[2]. 이런 웹의 변화에 따라서 사용자들의 성향과 사용자들 사이에 이슈가 되는 주제를 추출하는 것이 중요한 문제로 대두되고 있다[3].

연관 콘텐츠를 추출하는 방법은 본 논문에서 제시한 방법 외에도 다양한 방법들이 제안되었는데 대부분의 방법들은 다음과 같은 문제를 안고 있다[3][4].

- 태그는 사용자의 성향을 직접적으로 묘사하지 못함
- 방대한 콘텐츠들에서 사용자들의 이슈가 되는

태그를 요약할 수 없음

- 소셜 네트워크에서 이슈가 되는 관심 키워드를 구별할 수 없음

본 논문에서 제안하는 방법은 사용자들의 성향을 직접적으로 묘사하기 위하여 온톨로지 기반 사용자 프로파일 기술언어인 FOAF[5]를 사용한다. 그리고 방대한 콘텐츠들에서 사용자들의 핫 이슈가 되는 관심 태그를 요약하기 위하여 mybloglog[6]를 통해 콘텐츠를 수집하고 del.icio.us[7]를 통해 추천 정보를 수집한다. 마지막으로 소셜 네트워크에서 이슈가 되는 관심 키워드 추출하기 위해 FOAF와 콘텐츠 데이터를 대상으로 소셜 네트워크 분석을 한다. 본 논문에서 제안하는 방법을 활용하면 소셜 네트워크에서 이슈가 되는 키워드를 방대한 콘텐츠들에서 추출하는 것이 가능하며 이를 통해 보다 사용자의 성향을 반영한 연관 콘텐츠를 추천하는 것이 가능하다.

### 2. 관련 연구

상업적 웹 사이트들과 비 상업적 커뮤니티를 포함한 많은 사이트들에서 이슈를 추출하는 시스템은 해당 사이트의 경쟁력을 결정하는 요소가 되어가고 있

다[8]. 이에 따라 다양한 연구를 통해 다양한 콘텐츠 추천 방법이 제시되었다[9].

소셜 네트워크 분석을 활용하여 이슈가 되는 태그를 추출하는데 있어서는 사용자 중심(User-centric)과 객체 중심(Object-centric) 방법이 사용되고 있다. 사용자 중심 소셜 네트워크 분석 방법은 사용자들의 사용자들의 관계를 분석하여 공통의 관심 분야를 찾아내는 것이며 객체 중심 소셜 네트워크 분석 방법은 소셜 네트워크 커뮤니티에서 특정 객체에 해당하는 사용자들의 집단을 찾아내는 방법이다[3].

사용자 중심 소셜네트워크 분석을 활용한 연구에는 대표적으로 [10]과 [11]이 있다. [10]과 [11]에서는 특정한 관심 분야를 사용자들의 네트워크의 분석을 통해 찾아내었다. 그러나 [10]과 [11]에서 사용한 방법은 특정한 사이트에 국한되어 사용된다는 단점이 있으므로 다양한 사이트에서 커뮤니티를 이루고 있는 사용자들에게 범용적으로 활용되기가 어렵다.

객체 중심 소셜 네트워크 분석 방법을 활용한 연구에는 [12]와 [13]이 있다. [12]와 [13]에서는 사용자가 입력한 키워드에 대한 소셜 네트워크를 생성하지만 콘텐츠에 대한 중요도와 정확성을 판단하기에 어렵다.

### 3. FOAF와 소셜 네트워크 분석을 이용한 연관 콘텐츠 추천 방법

본 논문에서 제안하는 FOAF와 소셜 네트워크 분석을 이용한 연관 콘텐츠 추천방법은 크게 4개 단계로 구분할 수 있다.

- 1단계 : FOAF를 이용한 콘텐츠의 수집
- 2단계 : 소셜 네트워크의 생성
- 3단계 : 소셜 네트워크 분석
- 4단계 : 핫 이슈 추출 및 관련 콘텐츠 추천

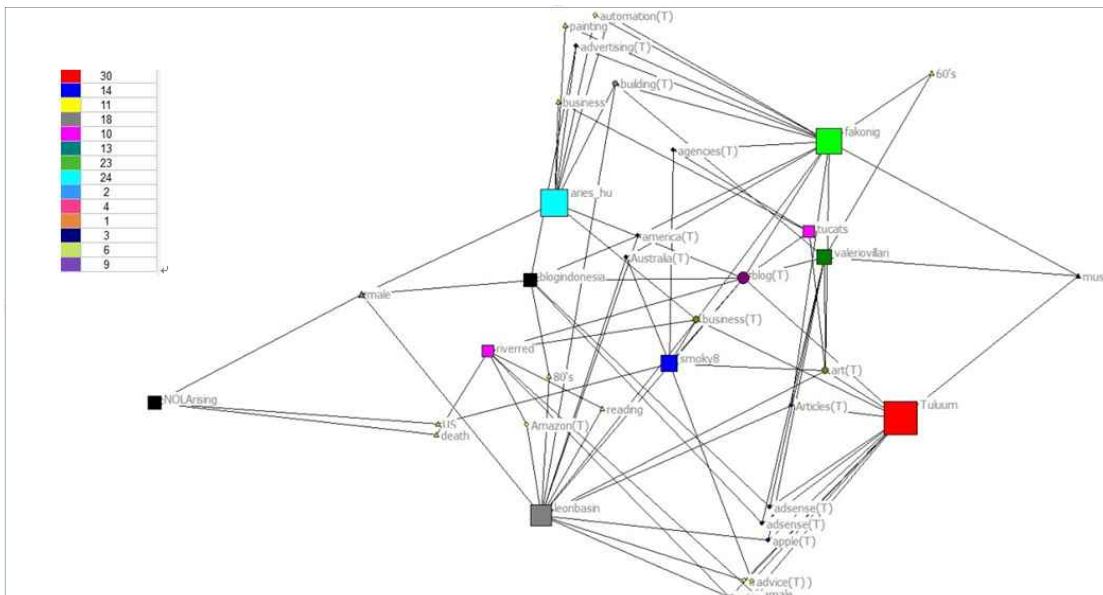
#### 1단계 : FOAF를 이용한 콘텐츠의 수집

많은 웹 2.0 기반 웹 사이트에서는 개방형 API 인터페이스를 통해 FOAF를 제공하고 있다. FOAF는 사용자의 관심 분야를 포함한 웹 사이트 및 성별 등에 대한 정보를 제공한다. 본 논문에서는 mybloglog에서 제공하는 FOAF를 통해 사용자 정보를 수집한다. 그리고 소셜 북마크 사이트인 del.icio.us를 통해 콘텐츠 추천 태그 정보를 수집하며 livejournal[14]을 통해 사용자들이 작성한 콘텐츠에 대한 정보를 수집한다.

#### 2단계 : 소셜 네트워크의 생성

소셜 네트워크를 생성하기 위하여 2단계에서는 세 개의 집합을 활용한다. 첫 번째 집합은 사용자 아이디 집합이며 두 번째 집합은 각 사용자들의 특징 집합이다. 그리고 마지막 집합은 태그 집합이다. 여기서 사용하는 세 개의 집합은 1단계에서 수집한 사용자 정보와 태그 및 콘텐츠 정보를 재구성하여 만들어진다. 이 세 개의 집합은 각각 다음과 같이 정의된다.

- 사용자 아이디 집합  $A = \{a_1, \dots, a_n\}$
  - 사용자 관심 키워드 집합  $C = \{c_1, \dots, c_p\}$
  - 태그 집합  $T = \{t_1, \dots, t_q\}$
- 위의 세 집합을 이용하여 본 논문에서는  $AT$ 와



(그림 1) 사용자-키워드-태그 그래프

AC, 이렇게 두 개의 행렬을 만든다. AT 행렬은 사용자 a가 태그 t를 사용했을 때 1을 기입하고 아닐 경우 0을 기입한 2차원 테이블이다. 그리고 AC 행과 C×T로 이루어진 2차원 테이블로 변형하면 사용자와 사용자 관심 키워드나 태그가 존재할 때 1이 기입되고 그게 아닐 경우 0이 기입된 행렬이 된다. 이 행렬에서 원소의 값이 1인 행과 열은 각각 그래프에서 꼭지점이 되며 두 꼭지점은 연결을 갖게 된다. 그리고 원소의 값이 0일 경우 두 꼭지점은 서로 연결을 갖지 않는다.

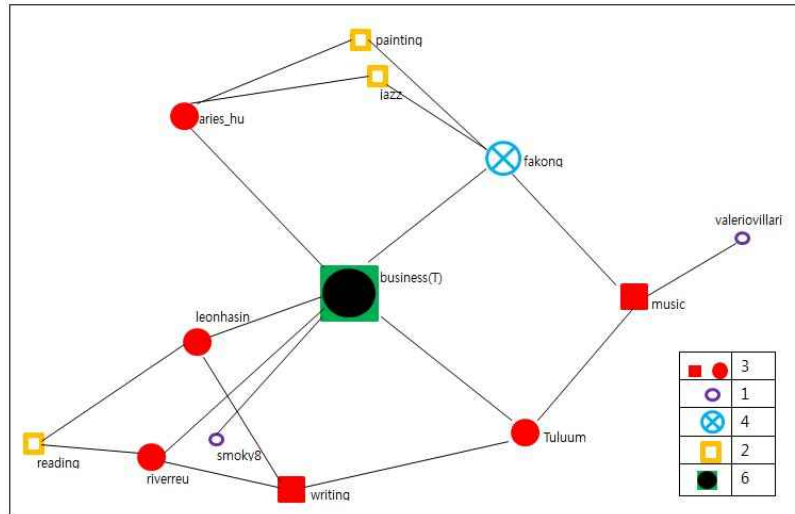
**3단계 : 소셜 네트워크 분석**

본 논문에서 사용한 소셜 네트워크 분석 방법은 중심성(Measure of degree centrality)이다. 중심성은 그래프에서 특정 꼭지점에 대해 진입/진출 차수의 개수를 수치화하여 중요도가 높은 꼭지점을 분석하

는 사용자 관심 키워드와 사용자가 매칭될 때 1을 기입하고 그렇지 않을 경우 0을 기입한 2차원 테이블이다. 이렇게 만들어진 AC와 AT 행렬을 다시 A한 콘텐츠를 추천하고자 한다면 사용자가 속한 네트워크에서 가장 중심성이 높은 키워드를 가진 콘텐츠를 선택하여 출력한다. 중심성이 높은 키워드를 선택할 때 불필요한 연산을 줄이기 위하여 본 논문에서는 상위 40%의 높은 값을 가진 꼭지점과 노드만으로 그래프를 재 생성한다. 그리고 재 생성된 그래프에서 사용자에게 추천할 연관 태그와 콘텐츠를 추출한다.

**4. 구현 및 실험**

본 논문에서 제안한 방법을 구현하기 위하여 우리는 UCINET[16]이라는 소셜 네트워크 분석 도구를 활용하였다. UCINET은 다양한 소셜 네트워크 분석



(그림 2) 사용자-중요 키워드 그래프

는 방법이다[15]. 그래프에서 어떤 꼭지점 v가 있고 꼭지점들의 총 개수가 n개 일 때, 꼭지점 v의 중심성  $C_D(v)$ 은  $(n-1)$ 을 와 v연결된 노드의 개수로 나눈 값이며 다음 식 1과 같이 정의 된다. 3단계에서는 중심성을 각 꼭지점에 대하여 계산하고 그 결과를 출력한다.

$$C_D(v) = \frac{\text{deg}(v)}{n-1} \quad (\text{식 1})$$

**4단계 : 핫 이슈 추출 및 관련 콘텐츠 추천**

핫 이슈 추출 및 관련 콘텐츠 추천 단계에서는 위 3단계에서 분석한 중심성 값을 이용하여 각 키워드들의 중요도를 반영한다. 만약 어떤 사용자에게 적합

한 도구를 제공하고 있으며 본 논문에서 제안하는 행렬 구조를 지원하고 있다.

본 논문에서 제안하는 방법을 적용하기 위하여 (그림 1)와 같이 10명의 사용자와 34개의 태그, 72개의 사용자 관심 키워드를 가진 그래프를 생성하였다. 그리고 (그림 1)에서 보인 그래프를 대상으로 중심성을 계산하고 상위 40%에 해당하는 꼭지점과 노드로 그래프를 재구성 하면 (그림 2)과 같은 그래프가 생성된다. (그림 2)에서 valerovilan이라는 사용자는 'music' 태그를 사용하였으며 다른 6명의 사용자는 모두 'business'태그를 사용하였고 그 외에 'painting', 'jazz', 'writing' 등과 함께 'music'태그를 사용하였음을 보인다. 이 때 valerovilan 사용자는 자신이 속한 소셜 네트워크에서 'business'태그가 핫

이슈로 추출되었음을 확인할 수 있다.

## 5. 결론

사용자들이 적극적으로 콘텐츠를 생산하고 공유하는 최근의 웹의 경향으로 인하여 사용자들의 활동 정도와 콘텐츠의 추천 빈도는 콘텐츠의 질을 결정하는데 중요한 요소로 부각되고 있다. 따라서 본 논문에서는 콘텐츠의 추천 횟수와 사용된 태그의 개수만으로 핫 이슈를 추출하는 전통적인 방식과는 달리 사용자의 네트워크와 태그 사용 빈도를 활용한 추천 방법을 제안하였다.

본 논문에서 제안한 방법은 FOAF와 소셜 네트워크 분석 방법을 활용함으로써 다양한 웹 사이트에서 형성된 소셜 네트워크의 핫 이슈 추출이 가능하도록 하였으며 소셜 네트워크에 속한 사용자에게 관련이 있는 태그와 콘텐츠를 추천할 수 있도록 하였다. 이 방법이 사용자의 성향뿐만 아니라 추천된 주제의 특징도 고려하기 때문에 사용자에게 성향에 맞는 주제를 추천해 줄 수 있다. 따라서 이 방법을 이용함으로써 기존의 방법 보다 향상된 효율성을 기대할 수 있다.

향후 우리는 본 논문에서 제안한 방법을 보다 범용적으로 활용할 수 있도록 웹 기반으로 시스템을 구현할 예정이며 추출한 핫 이슈의 정확도에 대하여 분석하는 연구를 진행하고자 한다.

## 참고문헌

- [1] Feng Fu, Lianghuan Liu, Long Wang: Empirical analysis of online social networks in the age of Web 2.0. In: Physica A 387 675 - 684 (2008).
- [2] Engene Barsky, Michelle Purdon: Introducing Web 2.0: social networking and social bookmarking for health librarians. In: JCHLA/JABSC 27: 65-67 (2006).
- [3] Xin Li., Lei Guo., Yihong(Eric) Zhao.: Tag-based Social Interest Discovery. In: International World Wide Web Conference Committee (IW3C2). ACM 978-1-60558-085-2/08/04(2008).
- [4] Anon Plangprasopchok., Kristina Lerman.: Exploiting social annotation for automatic resource discovery. In : AAAI workshop on Information Integration from the Web, (2007).
- [5] <http://xmlns.com/foaf/0.1/>
- [6] <http://www.mybloglog.com/buzz/>
- [7] <http://del.iciou.us>
- [8] Eric Butow, Kathleen Taylor. In: How to Succeed in Business Using Linked: Making Connections and Capturing Opportunities on the World's #1 Business Networking Site.(2009)
- [9] lorenTerveen,., Will Hill., BrianAmento., David McDonald., JoshCreter.: PHOAKS: A system for sharing recommendations. Communications. In: .ACM 40(3), 59 - 62 (1997).
- [10]Noor Ali-Hasan., Lada AAdamic.: Expressing social relationships on the blog through links and comments. In: Proc. of International Conference on Weblogs and Social Media, (2007).
- [11]Michael F. Schwartz., David C. M. Wood.: Discovering shared interests using graph analysis. In: Communications of the ACM, 36(8):78 - 89 (1993).
- [12]Kunwadee Sripanidkulchai, Bruce Maggs., Hui Zhang.: Efficient content location using interest-based locality. In: peer-to-peer systems. In Proc. of INFOCOMM, (2003).
- [13]Lei Guo, Song Jiang, Li Xiao., Xiaodong Zhang.: Fast and low-cost search schemes by exploiting localities in p2p networks. In: Journal of Parallel and Distributed Computing, 65(6):729 - 742 (2005).
- [14] <http://www.livejournal.com/site/about.bml>
- [15] Stephen P. Borgatti: Centrality and network flow. In: Social Networks 27 55-71 (2005).
- [16] <http://www.analytictech.com/ucinet/>