

# FOAF와 SNA를 이용한 개선된 인터넷 자원 추천 방법

Qing Wang<sup>†</sup> · 손 종 수<sup>\*\*</sup> · 정 인 정<sup>\*\*\*</sup>

## 요 약

최근 사용자들이 생성한 콘텐츠들이 크게 늘어나고 커뮤니티 기반 웹 사이트가 발전함으로 인하여 사용자들에게 인터넷 자원을 추천하는 시스템이 큰 각광을 받고 있다. 그러나 대부분의 인터넷 자원 추천 시스템들은 사용자의 특징을 충분히 반영하지 못하는 한계를 가지고 있다. 이에 따라 본 논문에서는 사용자의 특징이 충분히 반영되는 자원의 추천을 위하여 FOAF와 SNA를 사용한 추천 방법을 제안한다. 제안하는 방법은 1) FOAF를 통해 사용자의 특징 데이터와 태그 데이터를 취득한다. 2) 취득한 데이터를 세 종류의 행렬에 삽입하고 통합한 후 사용자, 사용자의 특징, 태그를 나타내는 그래프를 생성한다. 3) 소셜 네트워크 분석을 통해 추천 항목의 일반 특징과 핫태그(Hot tag)를 선정하여 인터넷 자원을 추천한다. 본 논문의 검증은 위하여 우리는 실험을 통해 본 논문에서 제안한 방법과 아이템 기반 추천 방법을 비교하였다. 이를 통해 보다 많은 사용자가 참여할수록 아이템 기반 추천 방법보다 본 논문에서 제안한 방법에 의한 추천 결과의 품질이 우수함을 확인하였다. 본 논문에서 제안하는 방법을 활용하면 사용자들에게 보다 적합한 자원을 추천하는 것이 가능하다. 그리고 제안하는 방법은 폭발적으로 늘어나는 인터넷 자원을 검색하는데 있어 효율적으로 활용될 수 있다.

키워드 : 소셜네트워크, 소셜네트워크 분석, 인터넷 자원 추천 방법, FOAF

## Improved Internet Resource Recommendation Method using FOAF and SNA

Qing Wang<sup>†</sup> · Jong-Soo Sohn<sup>\*\*</sup> · In-Jeong Chung<sup>\*\*\*</sup>

## ABSTRACT

In recent years, due to rapidly increasing user-created internet contents coupled with the development of community-based websites, the internet resource recommendation systems are attracting attentions of the users. However, most of the systems have failed in properly reflecting users' characteristics and thus they have difficulty in recommending appropriate resources to users. In this paper, we propose an internet resource recommendation method using FOAF and SNA which fully reflects the characteristics of users. In our method, 1) we extract the data about user characteristics and tags using FOAF; 2) we generate graphs representing users, user characteristics and tags after inserting data into 3 matrixes and integrating them; 3) we recommend the appropriate internet resources after selecting common characteristics of the recommended items and Hot tags by analyzing social network. For verification of our proposed method, we implemented our method to establish and analyze an experimental social group. We verified through our experiments that the more users added in the social network, the higher quality of recommendation result we got than the item-based recommendation method. By using the suggested idea in this paper, we can make a more appropriate recommendation of resources to users while effectively retrieving explosively increasing internet resources.

Keywords : Social Network, Social Network Analysis, Internet Resource Recommendation Method, FOAF

## 1. 서 론

최근에 들어 웹2.0은 인터넷에 있어서 중요한 역할을 수행하고 있으며 빠르게 보급되고 있다. 그리고 웹2.0의 보급

으로 인하여 사용자의 참여와 자원의 공유가 중요한 이슈로 대두되었다. 이로 인하여 페이스북(Facebook, <http://www.facebook.com>), 트위터(Twitter, <http://www.twitter.com>), 델리셔스(Del.icio.us, <http://del.icio.us>), 플리커(Flickr, <http://www.flickr.com>) 등과 같은 커뮤니티 기반의 웹 사이트들이 빠르게 보급되고 있다. 사용자들은 서로 인터넷을 통해 관계를 맺고 즐겨찾기, 블로그, 사진, 음악, 비디오 등의 자원을 커뮤니티 기반 웹사이트에서 공유한다. 한 예로, 2009년의 통계에 따르면, 미국의 청소년 12세에서 17세 사이의 청

<sup>†</sup> 준회원: 중국 운남성 전력실험 연구원  
<sup>\*\*</sup> 준회원: 고려대학교 전산학과 박사과정  
<sup>\*\*\*</sup> 종신회원: 고려대학교 컴퓨터정보학과 교수  
논문접수: 2011년 10월 25일  
수정일: 1차 2012년 1월 5일  
심사완료: 2012년 1월 27일

소년들 중에서 약 73%가 온라인 소셜 네트워크를 사용하고 있으며 성인의 47%가 온라인 소셜 네트워크를 이용하고 있다고 보고되었다[17]. 그러나 많은 커뮤니티 기반 웹 사이트가 성공을 하고 더 많은 사람들이 웹을 통해 콘텐츠를 생성 및 공유하는 최근의 경향에도 불구하고 각 웹 사이트들에서 제공하고 있는 자원 추천은 사용자의 특징이 완벽하게 반영되지 못하고 있다. 이에 따라 적합한 자원을 추천하기 위한 방법은 중요한 문제로 대두되고 있다. 인터넷 자원을 추천하는데 있어 많이 쓰이는 추천 방법은 협업적 인터넷 자원 추천 방법이다[30]. 협업적 추천 방법은 다수의 사용자들이 직접 콘텐츠를 선별하고 공유하도록 하기 때문에 웹 서버의 계산량이 적고 구현이 간단한 장점이 있다. 따라서 이를 활용하기 위한 다양한 연구가 이루어져 왔다[16][17][18][19]. 그러나 협업적 인터넷 자원 추천 방법은 다음과 같은 문제를 가지고 있다[27][28][29].

- 협업적 인터넷 자원 추천 방법은 사용자 특징을 동적으로 생성하는 것을 지원하지 못한다.
- 참여자의 수가 적을 때 협업적 인터넷 자원 추천 방법을 활용하기 어렵다.
- 사용자간의 관계와 자원의 내용이 추천에 반영되기가 쉽지 않다.

한편, 협업적 인터넷 자원 추천 방법과는 다르게 콘텐츠의 내용을 기반으로 하는 콘텐츠 기반 추천 방법 또한 많은 연구가 이루어져 왔다[2] [8] [10] [11] [12] [13][14] [15]. 그러나 콘텐츠 기반 추천 방법은 다음과 같은 한계점이 있다[28][29].

- 사용자의 프로파일이 좋은 품질을 가지고 있지 않으면 인터넷 자원의 추천 품질 또한 잠재적으로 좋지 않을 수 있다.
- 아이템의 유사도만을 측정하여 사용자의 관심사와 다른 사용자와의 관계를 이용한 추천이 어렵다.
- 사용자의 관심사와 콘텐츠의 정보를 모두 반영한 추천이 어렵다.

따라서 적절한 자원을 추천하기 위해서는 사용자의 특징을 반영하지 못하는 협업적 추천 방법과 대중적 인기도를 평가하지 못하는 콘텐츠 기반 추천 방법의 한계를 극복하여야 한다.

본 논문에서는 사용자의 특징과 콘텐츠의 대중적 인기도를 동시에 반영하기 위하여 사용자들의 태그와 FOAF 프로파일을 동시에 사용하는 방법을 제안한다. 제안하는 방법은 XML/OWL 기반 사용자 프로파일 표현 도구인 FOAF (Friend-Of-A-Friend)[4]와 SNA (Social Network Analysis)[5]를 활용한 추천 방법이다. 우리는 커뮤니티 기반 웹 사이트인 마이블로그로그, 딜리셔스, 라이브저널에서 제공하는 FOAF를 활용하여 태그 데이터와 사용자 특징 데이터를 취득한다. 그리고 FOAF를 이용한 소셜 네트워크를

생성하기 위하여 우리는 세 종류의 행렬에 데이터를 삽입 및 통합하고 그래프를 생성한다. 생성된 그래프는 사용자와 특징, 사용자와 태그의 관계를 표현하는데 이를 활용하여 우리는 소셜 네트워크 분석을 수행한다. 그리고 소셜 네트워크 분석을 통해 우리는 사용자의 핫태그와 일반 특징을 추출한다. 마지막으로 핫태그와 일반 특징을 이용하여, 우리는 공통적인 일반특징을 가지고 있는 사용자에게 핫태그가 가리키는 추천 항목을 사용자들에게 추천한다.

본 논문에서 제안한 방법의 검증을 위하여, 우리는 본 논문에서 제안하는 인터넷 자원 추천 방법을 실험하였다. 우리는 실험 평가를 통해 적중률, 재현율에 대하여 제안한 방법과 아이템 기반 방법을 비교한다. 평가 결과, 본 논문에서 제안한 방법은 아이템 기반 추천 방법에 비해 적중률이 0.02에서 0.103으로 향상되었다. 그리고 아이템 추천 방법은 참여한 사용자의 수에 큰 영향을 받지 않았지만 본 논문에서 제안한 방법은 참여한 사용자가 많아질수록 더 좋은 결과가 산출되었다. 따라서 본 논문에서 제안하는 방법을 활용하면 서비스 제공자는 사용자들에게 보다 적합한 자원을 추천해 줄 수 있다. 그리고 사용자의 수가 많아 질수록 추천 결과의 품질이 높아지므로 폭발적으로 늘어나는 인터넷 자원을 검색하는데 있어 효율적으로 활용될 수 있다.

본 논문의 구조는 다음과 같다. 2장에서는 자원 추천과 관련된 연구를 보인다. 3장에서는 우리가 제안하는 시스템의 구조와 자원 추천을 위한 4가지 단계를 보인다. 4장에서는 본 논문에서 제안하는 방법의 예제를 보이며 아이템 기반 추천방법과의 비교 평가를 보인다. 마지막으로 5장에서는 본 논문의 결과를 정리하고 향후 연구에 대하여 토의한다.

## 2. 관련 연구

전자상거래 및 커뮤니티 웹 사이트 등에서의 추천은 중요한 역할을 수행하고 있다. 웹 사이트에서의 자원 추천은 사용자들에게 어떤 아이템을 구매할 것인지, 혹은 어떤 뉴스를 읽고 어떤 음악을 들을 것인지를 도와준다[6]. 최근에 들어 인터넷에서는 사용자들이 직접 콘텐츠를 생성하고 공유하는 것이 대중화되어 이전에 비해 폭발적으로 인터넷 자원이 늘어나고 있다. 따라서 적절한 자원을 선별하고 추천하는 것은 점차 중요하게 여겨지고 있으며 이에 따라 추천 방법에 대한 연구는 꾸준히 이루어지고 있다. 인터넷 자원의 추천 방법으로는 대표적으로 협업적 추천 방법[16], [17], [18], [19]과 콘텐츠 기반 방법[8], [10], [11], [2], [12], [13], [14], [15] 그리고 그 외에 하이브리드 추천 방법[20], [21] 등이 있다.

협업적 추천 방법은 다른 사용자들에 의해 점수가 매겨진 자원을 추천하는 것을 목표로 한다. 자원에 점수를 매긴 사용자들은 그와 관련된 자원을 추천 받는다. 이와 관련된 연구로써[16]에서는 다양한 측정 방법을 통한 추천 시스템을 제안하였으며 [17]에서는 적은 정보를 가진 항목에서 차상위 정보 항목을 이용한 방법을 제안하였다. 그리고 [18]에서는 반복적 투표 방법을 제안하였다. 마지막으로 대부분의 협업

적 추천 방법과는 다르게 [19]에서는 링크드 데이터(Linked-data)를 활용한 개방적 추천 방법을 제안하였다. 협업적 추천 방법에 대한 연구들은 사용자의 경험과 지식을 자원 추천에 활용하는 면에서 최근의 인터넷 동향에 잘 부합되지만 사용자 간의 관계와 자원의 내용이 추천에 반영되지 못한다는 단점을 가지고 있다[28]. 그리고 협업적 추천 방법은 참여자의 수가 적을 때 추천의 품질이 떨어지며 사용자 특징을 동적으로 생성하는 것을 지원하지 못하는 단점을 가지고 있다[27][29].

콘텐츠 기반 추천 방법은 크게 아이템 기반 추천과 사용자 기반 추천으로 구분될 수 있다. 아이템 기반 추천 방법은 사용자들이 일반적으로 좋아하는 것이나 예전에 좋아했던 것을 추천에 활용하며 사용자 기반 추천 방법은 소셜 커뮤니티에서 비슷한 유형의 사용자들의 관심사를 기반으로 추천한다[1]. 아이템 기반 추천 방법의 연구로써 [8]의 연구에서는 아이템과 아이템간의 유사도를 기반으로 자원을 추천하는 방법을 제안하였다. [8]에서 보인 추천 방법은 아이템의 유사도만을 측정하여 사용자의 관심사와 다른 사용자와의 관계를 고려하지 않은 단점이 있다. 그리고 [10], [11]의 연구에서는 아이템 기반 추천 방법을 활용하여 사용자들의 일반적 관심사를 추출하였으나 오직 P2P 시스템에서 사용중인 아이템을 대상으로 하였으며 사용자들에게 정보를 제공하는 방법을 불명확하게 정리하였다. 이와 유사한 연구로는 [2]의 연구를 들 수 있다. [2]에서는 사용자들의 관심태그를 추출하였다. 그러나 [2]에서도 사용자와 사용자 간의 관계를 추천에 반영하지 못한 단점을 가지고 있다. 아이템 기반 추천 방법의 대부분은 추천할 자원에 초점이 맞춰져 있으므로 사용자의 관심사와 콘텐츠의 정보를 모두 반영한 추천이 어려운 단점이 있다[28].

아이템 기반 추천 방법과는 다르게 사용자 기반 추천 방법은 대체로 사용자들의 관계를 기반으로 사회적 관심사를 발견하는데 초점이 맞추어져 있다. 사용자 기반 추천 방법에는 대표적으로 [12]와 [13]이 있다. 이들은 소셜 네트워크에서 사용자들의 공통된 관심사를 찾는 방법을 제안하였다. [14]에서는 사용자의 아이템에 점수를 매겨 다른 자원의 점수를 예측하는 연구를 보였다. 그리고 보다 이전의 연구로써 [15]에서는 사용자 기반의 추천 방법을 제안하였다. 그러나 [12][13][14][15]에서 보인 사용자 기반 추천 방법은 인터넷 자원의 내용 및 항목의 특징을 활용하지 않은 단점을 공통적으로 가지고 있다. 그리고 사용자의 프로파일이 좋은 품질을 가지고 있지 않으면 인터넷 자원의 추천 품질 또한 잠재적으로 좋지 않을 수 있다[29].

하이브리드 콘텐츠 추천 방법은 콘텐츠 기반 추천 방법과 협업적 추천 방법을 조합하여 각각의 단점을 보완하는데 그 목적이 있다. 그의 한 예로써 [20]에서는 퀵스텝(Quickstep)과 폭스트로트(Foxtrot)라는 이름의 연구 주제 온톨로지를 기반으로 온라인 연구 논문을 추천하는 시스템을 제안하였다. 그리고 [21]에서는 음식점과 관련된 도메인 지식을 기반으로 적절한 식당을 추천하는 시스템을 제안하였다. 그러나 [20]과 [21]에서 보인 추천 방법은 아주 특수한 분야의 자원

을 추천하는 것으로써 다른 분야에 적용하기 어려운 문제를 가지고 있다.

### 3. SNA와 FOAF를 활용한 자원 추천 방법

#### 3.1 용어의 정의

사용자(Actor): 일반적으로 사용자는 웹이나 소셜 네트워크 서비스에서의 사용자를 의미하며 본 논문에서는 마이블로그로그 웹 사이트의 사용자 및 딜리셔스와 라이브저널의 사용자인 모든 사용자를 집합 A라고 정의한다.

특징(Characteristic): 특징은 사용자의 국가, 나이, 관심사 등을 의미하며 본 논문에서는 라이브저널을 통해 수집한 FOAF 안에 포함된 *foaf:interests*, *foaf:age*, *foaf:gender*, *foaf:country* 등의 항목들을 말한다. 본 논문에서 우리는 특징들의 집합을 집합 C로 정의한다.

태그(Tag): 태그는 웹 문서나 콘텐츠를 표현하는 키워드의 역할을 하는 것으로서, 우리는 딜리셔스에서 사용자들이 자신의 콘텐츠에 입력한 키워드들을 태그라고 정의하며 이 태그들의 집합을 집합 T로 정의한다.

핫태그(Hot Tag): 핫태그는 전체 태그들 중에서 많은 사용자들에 의해 사용되고 있는 태그를 의미한다. 사용자들이 입력한 태그 중에서 핫태그를 선정하는데 그 상세한 과정은 3.5절에서 보인다.

일반 특징(Common Characteristic): 일반 특징은 사용자들의 FOAF에 포함된 항목들 중에서 많은 사용자들에 의해 사용되고 있는 특징을 의미하며 일반 특징의 선정 과정 역시 3.5절에서 보인다.

#### 3.2 SNA와 FOAF를 활용한 자원 추천 절차

본 절에서는 본 논문에서 제안하는 방법의 전체적인 윤곽을 보이며 자원추천의 절차를 요약한다. 제안하는 방법은 태그와 사용자 특징을 수집하고 이를 이용하여 3개의 행렬을 생성하여 그래프 변환 과정을 거친다. 그리고 변환된 그래프의 꼭지점을 대상으로 소셜 네트워크 분석을 수행하여 자원을 추천한다. 다음 1), 2), 3)는 본 논문에서 제안하는 방법의 순서와 요약이다. 1)에서 3)의 순서는 각각 3.3절, 3.4절 3.5절에서 상세하게 설명한다.

##### (1) FOAF 프로파일 데이터의 취득

커뮤니티 기반 웹 사이트에서 제공하고 있는 사용자 FOAF를 사용하여 사용자의 특징과 태그를 수집한다. 본 논문에서 사용한 웹 사이트는 마이블로그로그(Mybloglog, <http://www.mybloglog.com/>), 라이브저널(Livejournal, <http://www.livejournal.com>) 그리고 딜리셔스(Del.icio.us, <http://del.icio.us>)이다.

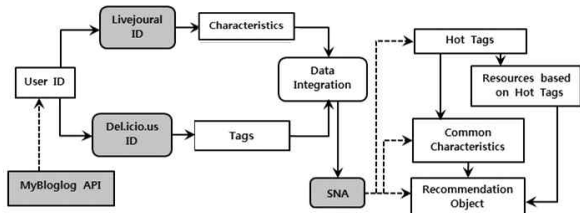
##### (2) 데이터의 통합 및 그래프 생성

취득된 데이터는 사용자 특징 데이터와 태그 데이터로 구분되며 이 데이터들은 다음 3.3절에서 소개한 3개의 행렬

( $N_{acs}$ ,  $M_{at}$ ,  $Q_{act}$ )에 통합된다. 그리고 이 행렬들은 각각 2차원 행렬이며 꼭지점과 간선으로 구성된 그래프로 변환된다.

(3) SNA 분석 및 자원의 추천

변환된 그래프에서 각 꼭지점들의 중심성 측정을 통해 일반 특징과 핫태그를 추출한다. 그리고 핫태그를 포함하고 있는 추천 항목을 추천한다.



(그림 1) 자원 추천 수행 절차

(그림 1)은 본 논문에서 제안하는 자원 추천 방법의 수행 절차를 보인다. 제안하는 방법은 먼저 마이블로그로그 API를 사용하여 사용자의 FOAF를 받아온다. 그리고 받아온 FOAF를 이용하여 딜리셔스에서 태그 데이터를 추출하며 라이브저널에서 특징 데이터를 추출한다. 그리고 이들은 'Data Integration' 모듈을 통해 3개 행렬로 통합된다. 이 행렬들은 각각 꼭지점과 아크로 구성된 그래프로 표현되며 이 그래프를 활용하여 SNA를 수행한다. SNA 과정을 거치면 다음으로는 필터링 과정을 거치게 되는데 이는 (그림 1)의 오른쪽 부분에 해당한다. 'Hot tag' 필터는 분석된 결과를 기반으로 일반 특징과 추천 항목을 선별한다. 그리고 선별된 일반 특징과 추천 항목은 각각 SNA 분석 결과를 통해 사용자에게 자원을 추천한다.

3.3 데이터의 취득

대중적으로 사용되고 있는 대부분의 커뮤니티 기반 웹 사이트는 개방형 API를 제공하고 있다. FOAF 프로파일은 개방형 API를 통해 제공 받을 수 있으며 우리는 이를 통해 충분한 사용자 정보를 제공 받는다. 본 논문에서 사용하는 데이터는 사용자의 관심사나 취향 등을 나타내는 사용자 특징 데이터와 인터넷 자원의 키워드 역할을 하는 태그 데이터이다. 이 두 가지 데이터의 상세한 내용은 다음 목록과 같다.

• 사용자 특징 데이터

사용자 특징데이터는 FOAF에 기술된 것을 기반으로 하며 사용자의 나이, 국가, 성별, 관심사 등으로 이루어져 있다. 사용자 정보는 사용자의 특징을 충분히 표현하고 있으므로 우리는 서로 다른 특징을 가진 사용자들을 각기 다른 그룹으로 구별한다.

• 태그 데이터

디지털 자원과 관련된 메타데이터의 한가지인 태그는 사용자들이 키워드를 표현할 때 사용되거나 웹 자원의 분류를

위한 어휘로 사용된다. 본 논문에서는 태그 정보를 사용자의 특징과 맵핑하여 SNA를 수행하는데 사용한다.

사용자 특징 데이터와 태그 데이터를 기반으로 본 논문에서 제안하는 방법을 수행하기 위하여 우리는 한 명의 사용자에 대해서 3개 종류의 웹 커뮤니티 기반 사이트(마이블로그로그, 라이브저널, 딜리셔스)의 API를 통해 3개의 서로 다른 FOAF 프로파일을 수집한다. 수집한 FOAF는 각기 다른 용도로 활용이 되는데 그 중에서 첫 번째로, 우리는 마이블로그로그에서 FOAF 프로파일을 수집한다. 마이블로그로그는 블로그와 블로그를 연결하는 위젯(Widget) 기반의 커뮤니티 기반 웹 사이트이다[22]. 마이블로그로그를 통해 수집된 FOAF는 딜리셔스와 라이브저널 사용자들의 ID를 보여주므로 우리는 이를 통해 딜리셔스와 라이브저널 사용자 아이디를 파악한다. 본 논문에서는 마이블로그로그의 FOAF를 기초로 딜리셔스와 라이브저널의 FOAF 프로파일을 수집하며 이를 본 논문에서 제안하는 방법의 데이터로 활용한다.

두 번째로, 우리는 딜리셔스를 통해 수집한 FOAF들에서 태그 데이터를 추출한다. 딜리셔스는 웹 북마크를 저장하고 공유하고 찾을 수 있도록 해주는 웹 기반 소셜 북마킹 서비스이다. 이는 사용자들이 입력한 북마크에 직접 자유롭게 키워드를 달아서 생성된 비계층적인 분류 시스템(Folksonomy)을 사용한다[23].

그리고 마지막으로 우리는 라이브저널을 통해 사용자의 개인특징을 수집한다. 라이브저널 웹 사이트는 자기표현을 위한 블로그 및 온라인 커뮤니티이다[24]. 우리는 라이브저널에서 제공하는 FOAF를 수집하며 이를 통해 나이, 성별, 국가, 관심사 등의 풍부한 개인 정보를 수집한다.

3.4 자료 통합 및 그래프 생성

3.3절에서 보인 바와 같이, 태그 데이터와 사용자의 특징 데이터를 각각 딜리셔스와 라이브저널에서 수집하면 본 논문에서 제안하는 시스템은 두 데이터를 다음과 같이 2차원 행렬  $N_{acs}$ ,  $M_{at}$ ,  $Q_{act}$ 를 생성한다.  $N_{ac}$  행렬은 사용자( $a$ )와 사용자의 특징( $c$ )을 각각 한 개 차원으로 설정하여 생성한 2차원 행렬이다.  $M_{at}$  행렬은 사용자( $a$ )와 태그( $t$ )를 각각 한 개의 차원으로 설정하여 생성한 2차원 행렬이다. 그리고  $Q_{act}$  행렬은  $N_{ac}$  행렬과  $M_{at}$  행렬을 하나로 합친 2차원 행렬이다.  $N_{acs}$ ,  $M_{at}$  그리고  $Q_{act}$  행렬은 사용자, 특징 그리고 태그간의 관계를 나타내는 그래프이다. 본 절에서는  $N_{acs}$ ,  $M_{at}$  그리고  $Q_{act}$  행렬을 설명한다.

(1) 사용자(Actor), 특징(Characteristic), 태그(Tag)

사용자는 웹이나 소셜 네트워크에서의 사용자를 의미하며 본 논문에서는 마이블로그로그 웹 사이트의 사용자이면서 딜리셔스와 라이브저널의 사용자인 모든 사용자를 집합  $A$ 라고 정의한다. 그리고 집합  $A$ 에서  $i$  번째 원소를  $a_i$ 로 표현한다. 따라서  $n$ 명의 사용자가 있다고 했을 때 사용자 집합  $A$ 는  $A = \{a_1, a_2, \dots, a_i, \dots, a_n\}$  ( $1 \leq i \leq n$ )으로 표현된다. 예를 들어 예제 소셜 네트워크에 5명의 사용자( $John, Amy,$

Mike, Lucy, Mary)가 있다고 할 때 집합  $A$ 는  $A = \{John, Amy, Mike, Lucy, Mary\}$ 로 표현된다.

특징은 사용자의 관심사나 취향 등을 의미하며 우리는 라이브저널을 통해 수집한 FOAF 안에 포함된 *foaf:interests*, *foaf:age*, *foaf:gender*, *foaf:country*의 항목을 특징으로 사용한다. 다시 말해, 라이브저널을 통해 수집한 FOAF 파일에서 추출한 사용자 특징들을 집합  $C$ 로 정의한다. 그리고 우리는 집합  $C$ 에서  $j$  번째 원소를  $c_j$ 로 표현한다. 따라서 집합  $C$ 는 전체 원소의 개수가  $p$ 개라고 했을 때  $C = \{c_1, c_2, \dots, c_j, \dots, c_p\}$  ( $1 \leq j \leq p$ )로 표현된다. 예를 들어 예제 소셜 네트워크에 4개의 특징(USA, UK, Male, 25)이 있다고 했을 때  $C$ 는  $C = \{USA, UK, Male, 25\}$ 으로 표현된다.

태그는 웹 문서나 콘텐츠의 특징을 표현하는 키워드의 역할을 하는 것으로서, 우리는 딜리셔스에서 수집한 태그들을 집합  $T$ 로 정의하며 집합  $T$ 에서  $k$ 번째 원소를  $t_k$ 로 표현한다. 따라서 집합  $T$ 에 총  $q$ 개의 원소가 있다고 할 때 집합  $T$ 는  $T = \{t_1, t_2, \dots, t_k, \dots, t_q\}$  ( $1 \leq k \leq q$ )로 표현된다. 예를 들어, 예제 소셜 네트워크에 6개의 태그(*google*, *music*, *movie*, *photo*, *tv*, *web*)가 있다면 집합  $T$ 는  $T = \{google, music, movie, photo, tv, web\}$ 로 표현된다.

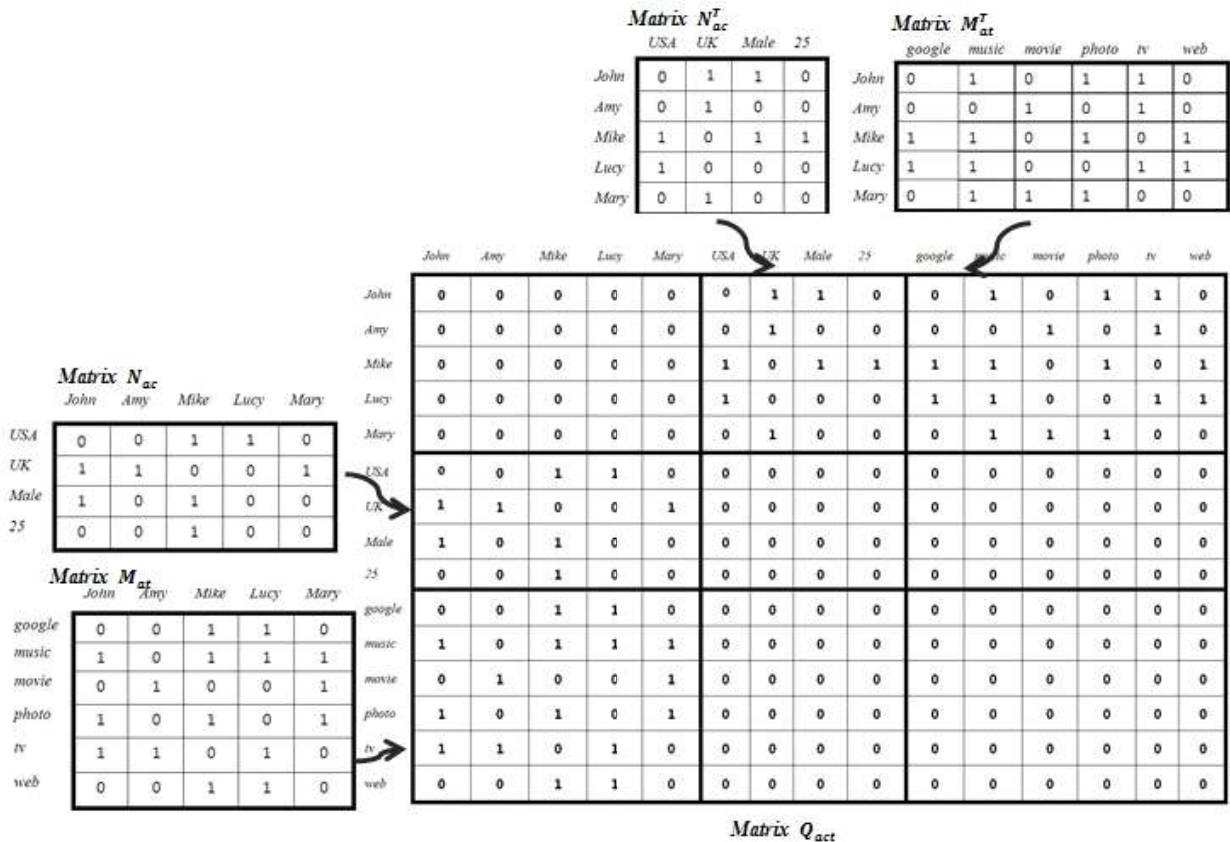
(2) 행렬  $M_{at}$ , 행렬  $N_{ac}$ , 행렬  $Q_{act}$

행렬  $M_{at}$ 는 사용자  $a$ 와 태그  $t$ 로 구성된 행렬이며  $M_{at} = \{m_{ij}\}$ 으로 표현된다. 만약  $i$ 번째 사용자  $a_i$ 와  $j$ 번째 태그  $t_j$

가 있다고 할 때  $a_i$ 와  $t_j$ 의 관계  $m$ 은  $m_{ij}$ 로 표현하며  $m_{ij}$ 는 행렬  $M_{at}$ 의 원소가 된다.  $m_{ij}$ 는 0과 1을 그 값으로 가질 수 있는데 만약  $a_i$ 와  $t_j$  사이에 연결이 있다면  $m_{ij}$ 는  $m_{ij} = 1$ 로 표현되며 그렇지 않은 경우에는  $m_{ij} = 0$ 으로 표현된다. 예를 들어 소셜 네트워크  $SN$ 안의 'John'이라는 사용자가 3개의 태그(*music*, *photo*, *tv*)를 사용하고 있고 'Amy'라는 사용자가 'movie', 'tv'를, 'Mike' 사용자가 'music', 'movie', 'photo'를 사용하고 있다면  $M_{at}$ 는 (그림 2)에서 좌측 하단과 같이 표현된다.

행렬  $N_{ac}$ 는 사용자  $a$ 와 특징  $c$ 로 구성된 행렬이며  $N_{ac} = \{n_{xy}\}$ 으로 표현된다. 만약  $x$ 번째 사용자  $a_x$ 와  $y$ 번째 특징  $c_y$ 가 있다고 할 때  $a_x$ 와  $c_y$ 의 관계  $n$ 은  $n_{xy}$ 로 표현하며  $n_{xy}$ 는 행렬  $N_{ac}$ 의 원소가 된다. 그리고  $a_x$ 와  $c_y$ 의 연결이 있을 때  $n_{xy} = 1$ 이 되며 아닐 경우  $n_{xy} = 0$ 이 된다. 만약, 예제 소셜 네트워크  $SN$ 에 'John'이라는 사용자가 'USA', 'Male'를 가지고 있고 'Amy' 사용자가 'UK'를, 'Mike' 사용자가 'USA', 'Male', '25'를, 'Lucy' 사용자가 'USA'를 각각 가지고 있다면 행렬  $N_{ac}$ 는 (그림 2)에서 좌측 상단과 같이 표현된다.

행렬  $Q_{act}$ 는  $M_{at}$ 와  $N_{ac}$ 의 특징을 모두 가지고 있는 통합적 형태의 행렬이다. 그리고 이 행렬  $Q_{act}$ 는 소셜 네트워크 그래프를 생성하는데 기초적인 자료로 활용이 된다. 행렬  $Q_{act}$ 는 위에서 보인  $M_{at}$ 와  $N_{ac}$ 를 활용하여 생성하며 다음과 같이 표현되며 그 예제는 (그림 2)에서 중앙부에서 보이는 바와 같다.



(그림 2) 행렬  $M_{at}$ , 행렬  $N_{ac}$ , 행렬  $Q_{act}$

$$\text{Define matrix } Q_{act} = \begin{pmatrix} 0 & N_{ac}^T & M_{at}^T \\ N_{ac} & 0 & 0 \\ M_{at} & 0 & 0 \end{pmatrix}$$

행렬  $Q_{act}$ 는 위에서 정의되었듯이  $N_{ac}$ ,  $M_{at}$ ,  $N_{ac}^T$ ,  $M_{at}^T$  및 기타 0들로 구성이 된다. 그리고  $Q_{act}$ 는 대칭행렬이므로 본 논문에서 제안한 방법을 이용하여 연산할 때 실질적으로 사용되는 것은  $N_{ac}$ ,  $M_{at}$  뿐이다. 다시 말해, 데이터의 크기가 커짐으로 인하여  $Q_{act}$ 의 크기 또한 급격히 커지지만 실제로 연산에 참여하는 부분은  $N_{ac}$  및  $M_{at}$  뿐이므로 연산 시간이 지수적으로 증가하지는 않는다.

(3) 그래프 AT, 그래프 AC, 그래프 ACT

위에서 보인  $M_{at}$ ,  $N_{ac}$ ,  $Q_{act}$  모두 2차원 행렬이고 연결이 있을 때 1로, 연결이 없을 때 0으로 그 값을 넣었으므로 그래프로 표현이 가능하다. 그리고  $Q_{act}$ 는  $M_{at}$ ,  $N_{ac}$ 의 연결 관계를 모두 통합한 행렬이므로  $Q_{act}$ 를 그래프로 생성하면  $M_{at}$ 와  $N_{ac}$ 로 생성한 그래프는 모두  $Q_{act}$ 의 서브 그래프가 된다. 따라서 우리는 먼저 두 그래프 AT와 AC에 대하여 설명한다.

그래프 AT는  $Node_{AT}$ 와  $Edge_{AT}$ 로 구성된 그래프이다.  $Node_{AT}$ 는 사용자 집합 A와 태그 집합 T로 구성되며  $Edge_{AT}$ 는 집합 A의 어느 원소 a와 집합 T의 어느 원소 t가 서로 연결되어있을 때 그 연결들의 집합으로 구성된다.  $Node_{AT}$ ,  $Edge_{AT}$ , 그래프 AT의 표현은 다음과 같다.

$$\begin{aligned} Node_{AT} &= A \cup T \\ Edge_{AT} &= \{(a, t) \in A \times T \mid (a \in A, t \in T)\} \\ Graph_{AT} &= (Node_{AT}, Edge_{AT}) \end{aligned}$$

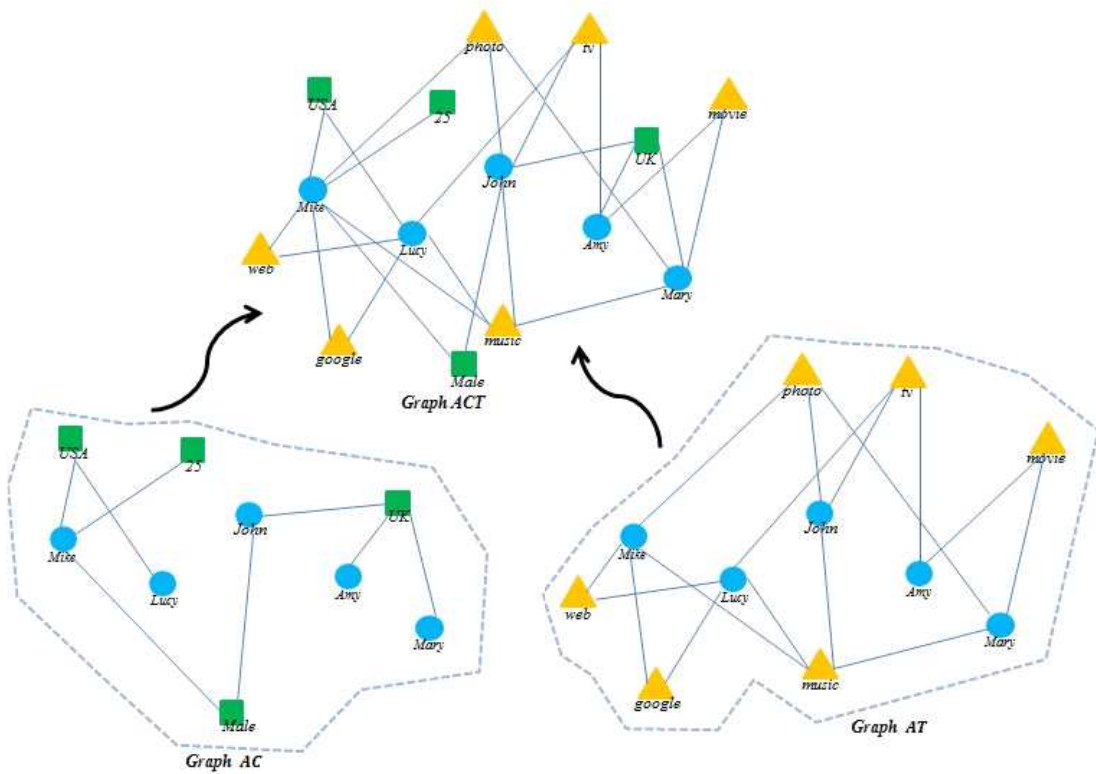
그리고 그래프 AC는  $Node_{AC}$ 와  $Edge_{AC}$ 로 구성된 그래프이다.  $Node_{AC}$ 는 사용자 집합 A와 특징 집합 C로 구성되며  $Edge_{AC}$ 는 집합 A의 어느 원소 a와 집합 C의 어느 원소 c가 서로 연결이 되어있다면 그 연결들의 집합으로 구성된다.  $Node_{AC}$ ,  $Edge_{AC}$  그리고 그래프 AC의 표현은 다음과 같다.

$$\begin{aligned} Node_{AC} &= A \cup C \\ Edge_{AC} &= \{(a, c) \in A \times C \mid (a \in A, c \in C)\} \\ Graph_{AC} &= (Node_{AC}, Edge_{AC}) \end{aligned}$$

본 논문에서 사용자, 특징, 태그를 모두 포함한 그래프 ACT는 그래프 AT와 그래프 AC의 꼭지점들과 간선들의 합 집합으로 정의한다.

$$\begin{aligned} Node_{ACT} &= Node_{AT} \cup Node_{AC} \\ Edge_{ACT} &= Edge_{AT} \cup Edge_{AC} \\ Graph_{ACT} &= (Node_{ACT}, Edge_{ACT}) \end{aligned}$$

다음 (그림 3)은 그래프 AT와 그래프 AC가 소셜 네트워크 그래프 ACT로 표현된 예제이다. (그림 3)에서 좌측은 그래프 AC의 예이며 우측은 그래프 AT의 예제이다. 그리고



(그림 3) 그래프 AC, 그래프 AT, 그래프 ACT

(그림 3)의 중앙 부분은 두 그래프  $AC$ 와  $AT$ 의 합집합인 그래프  $ACT$ 의 예제이다. (그림 3)에서 보이는 바와 같이 그래프  $ACT$ 는 모든 사용자, 태그, 특징들을 꼭지점으로 갖는 그래프이며 사용자, 태그, 특징을 구별하기 위하여 그래프에서는 사용자를  $\circ$  모양으로, 태그를  $\triangle$  모양으로, 특징을  $\square$  모양으로 표현하였다.

3.5 소셜네트워크 분석 및 인터넷 자원의 추천

본 절에서는 위 3.4절에서 보인 그래프  $ACT$ 를 활용한 소셜 네트워크 분석 과정을 설명한다. 소셜 네트워크 분석에는 중심성 측정(Measure of degree centrality), 매개 중심성 측정(Betweenness centrality), 근접 중심성 측정(Closeness centrality), 아이겐 벨류 중심성 측정(Eigen-value centrality) 등이 있다. 이 중에서 중심성 측정은 그래프에서 어떤 꼭지점이 얼마나 중심적인 역할을 하는지 측정하는 지표로 활용된다[32]. 매개 중심성 측정은 그래프 안의 여러 꼭지점들 중에서 정보의 유통이나 자료의 흐름에 있어 매개자적 역할을 하는 정도를 측정하는 지표이다. 그리고 근접 중심성은 그래프 안에 위치한 두 꼭지점들이 서로 가까운 정도를 측정하는데 사용된다. 아이겐 벨류 중심성 측정은 그래프가 점차 확대되고 안정화 될 때 꼭지점이 가질 수 있는 잠재적 영향력을 측정하는데 사용한다. 매개 중심성 측정이나 근접 중심성 측정은 그래프 안에서 임의의 꼭지점이 가지는 위상을 표현하는데 한계가 있으며 아이겐 벨류 중심성은 현재 상태보다는 잠재력을 측정하는데 유용하므로 현재 상황에서 가장 유용한 콘텐츠를 추천하는데 부족함이 있다. 중심성 측정은 연결된 꼭지점이 얼마나 많은지를 표현하므로 그래프에서 꼭지점의 위상을 표현하는데 적절하다 [33]. 따라서 본 논문에서는 여러 소셜 네트워크 분석 기법들 중에서 중심성 측정[31]을 활용한다.

(1) 중심성 측정

우리는 소셜 네트워크 분석 기법 중에 한가지인 중심성(Degree of centrality)을 활용하여 태그와 특징의 중심성을 계산한다. 중심성은 0에서 1사이의 실수로 그 값이 계산되는데 중심성의 크기가 크면 클수록 전체 그래프에서 크다는 것을 의미하며 중심성이 큰 꼭지점은 다른 꼭지점들과의 연결에 있어 허브(Hub)역할을 담당한다고 볼 수 있다. 아래 (식 1)은 중심성을 분석하기 위한 기초 단계로서 전체 꼭지점의 집합이  $I$ 라고 할 때 주어진 꼭지점  $i$ 와 연결된 꼭지점들의 개수를 구한다. 그리고 (식 2)에서 보이는 바와 같이 전체 꼭지점의 수가  $n$ 이라고 할 때 어떤 꼭지점의 중심성 값( $C_D(i)$ )를 계산한다.

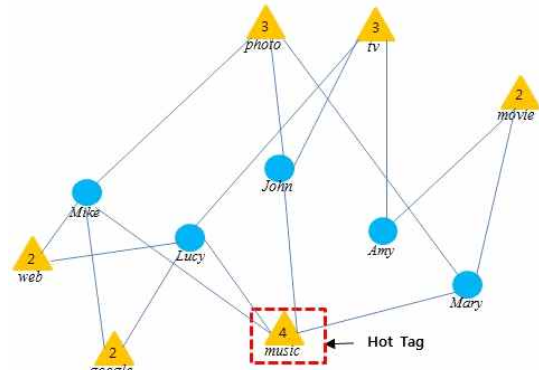
$$degree(i) = \sum_{j \in V} a_{ij} \tag{식 1}$$

$$C_D(i) = \frac{degree(Node)}{n-1} \tag{식 2}$$

우리는 (식 1)과 (식 2)를 이용하여 그래프  $ACT$ 의 모든 꼭지점들에 대하여 중심성을 계산한다.

(2) 핫태그의 선정

핫태그는 전체 태그들 중에서 많은 사용자들에 의해 사용되고 있는 태그를 의미한다. 핫태그는 소셜 네트워크의 크기에 따라 상대적인 중심성을 가질 수 있으므로 본 논문에서는 태그  $t$ 의 반영률 변수  $\beta_t$ 를 정의하여 사용한다.  $\beta_t$ 는 0에서 1사이의 값을 설정할 수 있는데 0에 가까울수록 더 많은 수의 핫태그가 선정되게 되며 이 때 선정된 핫태그의 중심성은 상대적으로 더 작다. 그리고  $\beta_t$ 가 1에 가까울수록 더 적은 수의 핫태그가 선정되게 되며 이 때 선정되는 핫태그는 더 큰 중심성을 갖는다.  $\beta_t$ 를 계산한 다음에는 태그  $t$ 의 임계값  $\theta_t$ 을 계산하는데  $\theta_t$ 는 전체 사용자의 숫자에  $\beta_t$ 를 곱한 값으로 계산하며  $\theta_t$ 보다 큰 중심성을 갖는 태그를 핫태그로 선정한다.  $\theta_t$ 의 계산은 다음 (식 3)과 같다.



(그림 4) 핫태그의 선정

$$\theta_t = \frac{Number\ of\ tags + Number\ of\ actors}{Number\ of\ nodes} \times \beta_t \quad (0 \leq \beta_t \leq 1) \tag{식 3}$$

$$Hot\_Tag = \{t \mid C_D(t) \geq \theta_t\} \tag{식 4}$$

예를 들어,  $\beta_t$ 가 0.3이라고 가정할 때 (그림 3)에서 보이는 그래프  $ACT$ 의 핫태그를 계산하면 다음과 같다.

$$C_D(web) = \frac{2}{15-1} = 0.14$$

$$C_D(google) = 0.14, \quad C_D(movie) = 0.14$$

$$C_D(photo) = 0.21, \quad C_D(tv) = 0.21$$

$$C_D(music) = 0.28$$

$$\theta_t = \frac{|tag|+|actor|}{|node|} \times \beta_t = \frac{6+5}{15} \times 0.3 = 0.220$$

$$C_D(music) \geq \theta_t$$

$$Hot\_Tag = \{ music \}$$

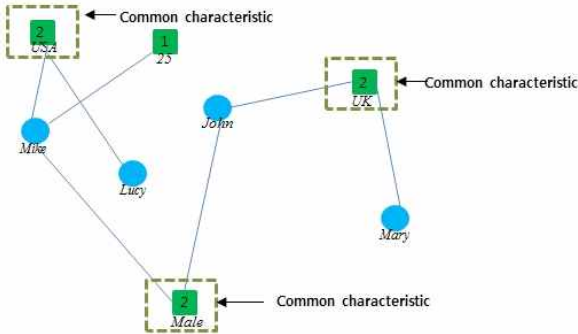
우리는 태그들 중에서  $C_D(t)$ 이 임계값  $\theta_t$ 보다 큰 값을 가지는 태그를 핫태그로 정의하였으므로 (식 4)에 의해 'music' 태그를 핫태그로 선정한다. (그림 4)는 (그림 3)에서 보인 그래프  $ACT$ 의 일부분이며 선정된 핫태그의 결과를 보인다.

(3) 일반 특성의 선정

일반 특성은 *foaf:interest*에 포함된 항목들 중에서 많은 사용자들에 의해 사용되고 있는 항목을 의미한다. 핫태그의 선정과 마찬가지로 소셜 네트워크의 크기에 따라 일반 특징 역시 상대적인 중심성을 가질 수 있으므로 본 논문에서는 특징  $c$ 의 반영률 변수  $\beta_c$ 를 정의하여 사용한다.  $\beta_c$ 는 0에서 1사이의 값을 설정할 수 있는데 0에 가까울수록 더 많은 수의 일반 특징이 선정되게 되며 이 때 선정된 일반 특징의 중심성은 상대적으로 더 작다. 그리고  $\beta_c$ 가 1에 가까울수록 그 반대의 경우가 된다.  $\beta_c$ 를 계산한 후에는 특징  $c$ 의 임계값  $\theta_c$ 를 계산하는데 이  $\theta_c$ 는 전체 사용자의 숫자에  $\beta_c$ 를 곱한 값으로써  $\theta_c$ 보다 큰 중심성 값을 갖는 특징을 일반 특징으로 선정한다.  $\theta_c$ 의 계산은 다음 (식 5)와 같다.

$$\theta_c = \frac{\text{Number of characteristic} + \text{Number of actors}}{\text{Number of nodes}} \times \beta_c \quad (0 \leq \beta_c \leq 1) \quad (\text{식 } 5)$$

$$\text{Common\_Characteristic} = \{c \mid C_D(c) \geq \theta_c\} \quad (\text{식 } 6)$$



(그림 5) 일반특징의 선정

일반 특징 또한 핫태그와 마찬가지로 (그림 3)의 그래프 ACT에서 (식 5)와 (식 6)을 이용해 계산한다. 다음은  $\beta_c$ 가 0.2라고 가정했을 때의 계산 결과이다.

$$C_D(\text{USA}) = \frac{2}{15-1} = 0.14$$

$$C_D(\text{UK}) = 0.14, \quad C_D(\text{Male}) = 0.14$$

$$C_D(25) = 0.07$$

$$\theta_c = \frac{|\text{characteristic}| + |\text{actor}|}{|\text{node}|} \times \beta_c = \frac{4+4}{15} \times 0.2 = 0.107$$

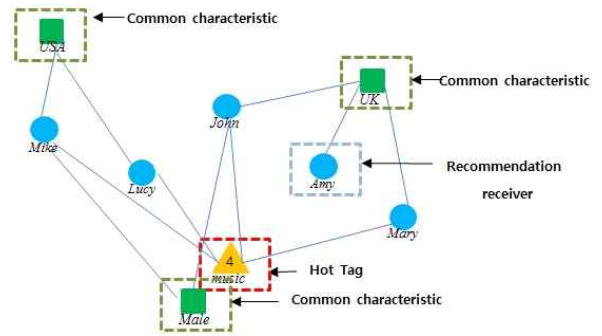
$$C_D(\text{USA}), C_D(\text{UK}), C_D(\text{Male}) \geq \theta_c$$

$$\text{Common\_Characteristic} = \{ \text{USA}, \text{UK}, \text{Male} \}$$

위와 같이 (식 5)와 (식 6)을 이용해 계산하면 (그림 5)에서 보이는 바와 같이 3개의 일반특징을 선정할 수 있다.

(4) 추천 받을 대상자의 선별과 인터넷 자원의 추천

마지막 단계로서, 우리는 추천을 받을 대상자와 추천 항목을 찾는다. 추천을 받을 대상자는 같은 일반 특징을 가지고 있는 사용자 중에서 핫태그를 사용하고 있지 않은 사용자를 의미한다. 그리고 추천 항목은 시스템이 사용자에게 추천할 자원을 의미한다. (식 4)에서 선정된 핫태그는 여러 사용자가 사용하고 있는 태그이므로 많은 자원과 연결이 되어있다. 그리고 위 (식 6)에서 일반 특징을 계산하는데 이를 통해 같은 일반 특징을 가지는 사용자 그룹을 나눌 수 있다.



(그림 6) 자원 추천

우리는 같은 일반 특징을 가지는 사용자들이 사용하는 핫태그를 선택하여 그 핫태그에 연결된 자원을 추천 항목으로 사용하며 사용자에게 이들을 추천한다.

(그림 6)은 (그림 3)에서 보인 그래프 ACT의 일부분이며 핫태그와 일반 특징 그리고 사용자를 보인다. 우리는 같은 일반 특징을 가진 사용자들에게 연결된 핫태그를 연결하여 자원을 추천한다. 예를 들어 (그림 6)에서 일반 특징 UK에 연결된 사용자는 {John, Amy, Mary}인데 이 중에서 'John'과 'Mary'는 'Music'이라는 핫태그와 연결이 되어있으나 'Amy'는 'Music'과 연결이 있지 않으므로 'Amy'에게 핫태그 'Music'를 추천한다.

4. 실험 및 평가

4.1 실험

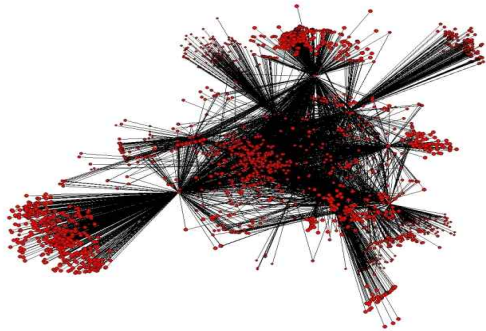
우리는 본 논문에서 제안한 방법의 실험을 위하여 마이블 로그로그, 라이브저널, 딜리셔스에서 데이터를 수집하였다. 실험을 위해 수집한 데이터는 2,676개의 꼭지점으로 구성된 소셜 네트워크이며 이 소셜 네트워크는 18명의 사용자와 98개의 특징 그리고 2,560개의 태그로 구성되어있다. 우리는 무작위로 라이브저널과 딜리셔스의 아이디에서 18개의 사용자 아이디를 선택하였다. <표 1>에서 보이는 바와 같이, 이 18개의 사용자 아이디를 이용하여 라이브저널에서 98개의 특징을 추출하였으며 딜리셔스에서 2560개의 태그 꼭지점을 추출하였다.

<표 1> 라이브저널과 딜리셔스에서 추출한 데이터

	라이브 저널	딜리셔스
사용자	특징	태그
18	98	2560



데이터를 수집한 후 우리는 그래프 및 소셜 네트워크 분석 도구인 패이젝(Pajek)과 UCINET6을 사용하여 실험을 진행하였다. (그림 7)은 본 논문에서 실험을 위해 사용한 그래프 ACT의 전체 모습을 보인다.



(그림 7) 실험에서 사용된 그래프 ACT

우리는 수집한 데이터로 그래프 ACT를 생성한 후  $\beta_t = 0.2$ 를 기준으로 태그들의 중심성을 측정하였다. 이를 통해 30개의 핫태그를 추출되었으며 <표 2>는 그 결과를 보인다.

<표 2> 추출된 핫태그

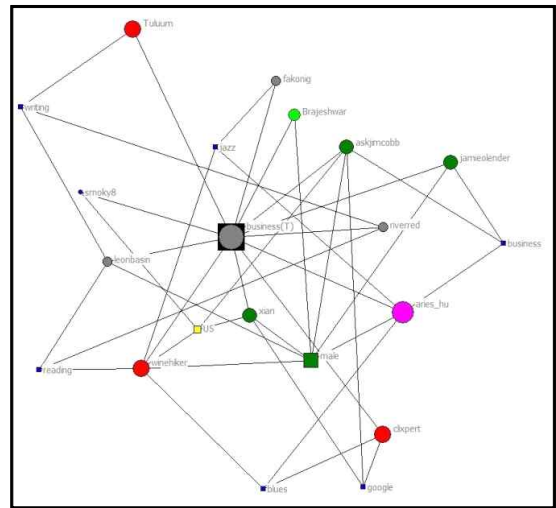
추출된 핫태그			
business	Technology	blog	photo
Work	Politics	Download	Tool
Online	News	Twitter	Cards
Music	Health	Ebook	Job
Learning	Free	Google	Marketing
Education	Money	Web	Networking
Software	Video	art	books
Advertising	Funny		

다음으로 우리는  $\beta_c = 0.2$ 를 기준으로 일반 특징을 추출하였으며 <표 3>은 그 결과를 보인다.

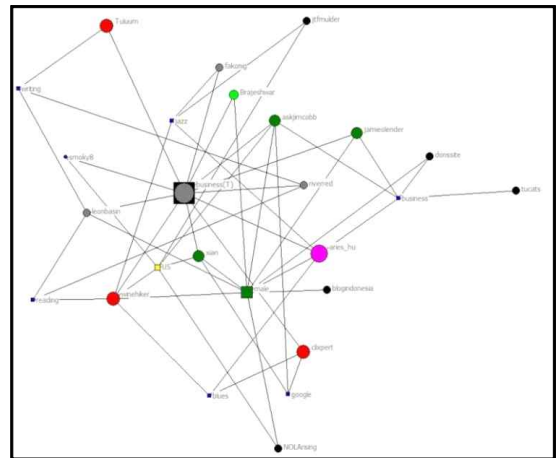
<표 3> 일반 특징의 추출 결과

추출된 일반 특징			
Writing	80's	Female	UK
Apple	US	Reading	70's
Jazz	Discovery	novel	50's
Art	Travel	Nature	Google
literature	Love	Food	advertising
Photography	movies	music	prayers
religion	CA	painting	singing
cooking	education	charities	internet
Male	Business	Religion	Education

(그림 8)은 그래프 ACT에서 핫태그 'business'를 통해 추출된 일반 특징에 대한 그래프의 표현이다. (그림 8)은 핫태그와 일반 특징의 관계를 보여주고 있는데 이 그림을 통해 우리는 business, google, blues 등이 일반 특징임을 알 수 있다. 그리고 이들은 (그림 8)에서 작은 사각형 모양으로 표현되었다.



(그림 8) 핫태그 'business'(가운데 사각형)와 일반 특징들(작은 사각형)



(그림 9) 일반 특징을 통한 추천 항목 선정(검은 원)

(그림 9)은 (그림 8)에서 보인 일반 특징을 이용하여 선정된 추천 대상자를 보인다. 선정된 추천 대상자는 (그림 9)에서 검은 원으로 표시되었다. 그리고 다음 <표 4>는 핫태그 'business'를 추천 받을 대상자의 리스트이다.

<표 4> 핫태그 'business'를 추천 받을 대상자

핫태그	추천 항목
business	blogindonesia
	jtfulder
	donssite
	tucats
	NOLArising

4.2 평가

우리는 본 논문에서 제안한 방법의 성능을 검증하기 위하여 아이템 기반 추천 방법과 본 논문에서 제안한 방법을 비교 평가한다. 아이템 기반 추천 방법은 사용자들의 추천 및 투표에 의해 자원이 평가되고 높은 점수를 받은 자원이 추천되는 방법이다[7]. 우리는 본 논문에서 제안한 방법과 아

이템 기반 추천 방법의 비교 평가를 위하여 [25]와 [26]에서 활용한 적중률(Hit-ratio)과 재현율(Recall)로 평가하는 방법을 따른다.

다음 (식 7)은 본 논문에서 제안한 방법을 통해 만들어진 핫태그의 적중률을 계산하는 식이며 (식 8)은 아이템 기반 추천 방법에서 사용하는 아이템의 적중률을 계산하는 식이다. (식 7)과 (식 8)에서  $u$ 는 사용자를 의미하며  $|Test_u|$ 는 소셜 네트워크에 있는 사용자의 태그 수이다. 그리고  $|Hottag_u|$ 는 사용자  $u$ 의 핫태그 개수이며  $|Item_u|$ 는 사용자  $u$ 의 아이템의 총 개수이다.

$$Hit\_Ratio(u) = \frac{|Test_u \cap Hottag_u|}{|Test_u|} \quad (식 7)$$

$$Hit\_Ratio(u) = \frac{|Test_u \cap Item_u|}{|Test_u|} \quad (식 8)$$

(식 7)과 (식 8)은 모두 적중률을 계산하는 수식으로서 전체 태그의 수에서 핫태그나 아이템이 몇 개나 일치하는지 계산한 결과를 보인다. 계산 결과의 범위는 최대 1부터 최소 0까지 가능하지만 본 실험에서는 자원 전체에서 핵심적인 역할을 하는 몇 개 단어를 사용하므로 1에 근접한 결과가 나올 확률은 매우 드물다.

$$Recall = \frac{\sum_{u=1}^k Hit\_Ratio(u)}{k} \times 100\% \quad (식 9)$$

(식 9)는 시스템에 전체  $k$ 명의 사용자가 있다고 할 때 사용자들의 적중률 합을  $k$ 로 나눈 값의 백분율이다. 다시 말해, (식 9)는 문서 집합에서 추천된 핫태그나 아이템이 전체적으로 얼마나 높은 적중률을 보였는지를 보인다.

(식 7), (식 8), (식 9)에서 보인 대로 우리는 같은 데이터를 가지고 각각 본 논문에서 제안한 방법의 적중률과 재현율을 계산하였으며 아이템 기반 추천 방법의 적중률과 재현율을 계산하였다. 먼저 <표 5>는 아이템 기반 추천 방법의 적중률 계산 결과를 보인다.

<표 5> 아이템 기반 방법의 적중률

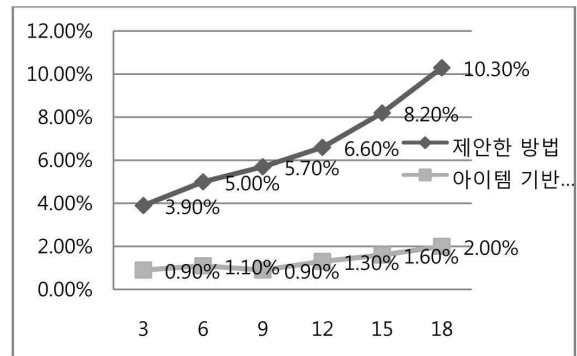
번호	사용자	태그	아이템	적중률
1	Brajeshwar	864	4	0.005
2	Tuluum	246	9	0.037
3	Riverred	118	1	0.009
4	aries_hu	270	4	0.015
5	jamieolender	118	4	0.034
6	askjimcobb	182	3	0.016
7	valeriovallari	70	1	0.014
8	smoky8	280	1	0.004
9	NOLArising	26	0	0.000
10	Tucats	83	3	0.036
11	Fakonig	268	2	0.007
12	Winehiker	677	10	0.015
13	Jtfmulder	73	2	0.027
14	Blogindonesia	78	2	0.026
15	Leonbasin	485	5	0.010
16	Xian	563	4	0.007
17	Donssite	119	1	0.008
18	Clixpert	10	1	0.100
평균				0.020

그리고 다음 <표 6>은 본 논문에서 제안한 방법으로 추출한 핫태그의 적중률 계산 결과를 보인다.

<표 6> 본 논문에서 제안한 방법의 적중률

번호	사용자	태그	핫태그	적중률
1	Brajeshwar	864	26	0.030
2	Tuluum	246	24	0.098
3	Riverred	118	21	0.178
4	aries_hu	270	16	0.059
5	jamieolender	118	20	0.169
6	Askjimcobb	182	12	0.065
7	valeriovallari	70	18	0.257
8	smoky8	280	19	0.068
9	NOLArising	26	2	0.077
10	Tucats	83	7	0.084
11	Fakonig	268	25	0.093
12	winehiker	677	25	0.037
13	Jtfmulder	73	8	0.110
14	Bloginesia	78	12	0.154
15	Leonbasin	485	29	0.060
16	Xian	563	29	0.051
17	Donssite	119	8	0.067
18	Clixpert	10	2	0.200
평균		251	16	0.103

<표 5>는 [26]에서 제안한 아이템 기반 추천 방법으로 계산한 적중률이다. 아이템 기반 추천 방법은 사용자의 작성한 글에 포함된 태그를 기반으로 추천하므로 관련 태그를 추가적으로 반영하지 못한다. 그러나 <표 6>에서 보이는 바와 같이 본 논문에서 제안한 방법은 사용자가 작성한 글에 포함된 태그뿐만 아니라 핫태그로 선정된 태그를 반영함으로써 더 높은 적중률 결과를 보인다.



(그림 10) 재현율 결과의 비교

<표 5>와 <표 6>의 결과를 비교하였을 때 아이템 기반 추천 방법에 비해 본 논문에서 제안한 방법의 적중률이 더 우수한 것을 확인할 수 있다.

(그림 10)은 <표 5>와 <표 6>의 결과를 비교한 그래프이다. (그림 10)에서 세로축은 재현율이며 가로축은 사용자의 수이다. 재현율은 검색한 질의어에 대하여 목적인 결과가 얼마나 많이 나타났는지 보이는 지표이며 높은 재현율을 보일수록 입력한 질의어와 일치하는 결과가 많다는 것을 의미한다. (그림 10)을 통해 우리는 본 논문에서 제안한 방법의 재현율이 아이템 기반 추천 방법에 비해 더 높음을 확인할 수 있다. 그리고 본 논문에서 제안한 방법이 아이템 기

만 추천 방법에 비해 사용자의 수가 많아 질수록 재현율의 상승 폭이 더 크다는 것을 알 수 있다. 따라서 사용자의 참여가 많아질수록 본 논문에서 제안한 방법이 기존의 아이템 기반 추천 방법에 비하여 점진적으로 높은 재현율을 보인다는 것을 알 수 있다. 그리고 본 논문에서 제안한 방법을 소셜 네트워크 서비스에 실제로 적용하게 되면 본 논문에서 실험한 규모보다 많은 사용자와 태그를 사용하게 되므로 더 높은 재현율을 기대할 수 있다.

## 5. 결 론

최근에 들어, 커뮤니티 기반의 웹사이트는 빠르게 발전하고 있다. 그러나 적절한 자원을 추천하기 위한 방법에 대한 연구는 여전히 과제로 남아있다. 이는 대부분의 방법들이 사용자 계정에 가입된 관심사만을 사용할 뿐, 사용자의 나이, 국가, 성별 등의 사용자 특징은 사용하지 않는 것에 기인한다. 이에 따라 본 논문에서는 온라인 소셜 네트워크 웹 사이트에서 FOAF와 SNA를 활용한 추천 방법을 제안하였다.

본 논문에서 제안한 방법은 커뮤니티 기반 웹 사이트들의 FOAF에서 개인 프로파일 정보와 태그 데이터를 추출하고 분석하여 인터넷 자원의 추천을 수행한다. 이 방법은 기존의 추천 방법과는 달리 사용자의 특징, 소셜 네트워크에서의 사용자 위상 그리고 태그를 동시에 콘텐츠 추천 과정에 반영하여 자원-공유-네트워크(Resource-Sharing-Network)를 향상시킬 수 있다.

본 논문에서 제안한 방법을 검증하기 위하여 우리는 2000꼭지점이 넘는 소셜 네트워크 데이터를 3종류의 웹 사이트에서 다운로드 받아 실험하였으며 그 결과를 평가하였다. 실험 결과를 통해 본 논문에서 제안한 방법이 아이템 기반 추천 방법보다 보다 훌륭한 콘텐츠 추천 품질을 보이는 것을 확인하였다. 그리고 본 논문에서 제안한 방법은 서비스 사용자의 수가 더 많아질수록 더 훌륭한 콘텐츠의 추천 품질이 좋아지는 특징이 있는 것을 실험을 통해 알아내었다. 따라서 본 논문에서 제안하는 방법을 활용하면 서비스 제공자는 사용자들에게 보다 적합한 자원을 추천해 줄 수 있으며 사용자의 수가 많아 질수록 서비스의 신뢰도를 높일 수 있다.

본 논문에서는 커뮤니티 기반 웹 사이트의 사용자를 대상으로 콘텐츠 추천 방법을 적용하고 실험하였다. 따라서 블로그, 소셜 네트워크 서비스 등에서는 본 논문에서 제안한 방법을 직접적으로 도입하여 활용할 수 있으며 이에 따른 개선된 콘텐츠 추천 결과를 기대할 수 있다. 그리고 본 논문에서 제안한 방법을 커뮤니티 기반 웹 사이트 및 각종 서비스에 적용하면 개인화된 추천과 보다 파급력이 있는 자원 공유가 가능하다. 뿐만 아니라, 본 논문에서 제안한 방법은 e-비즈니스, 정치, 사회분야에서 전문가 추천, 정보검색, 이슈 분석, 상품추천 등에 적용이 가능하다.

## 참 고 문 헌

- [1] L. Rainie, "The state of blogging.," Pew Internet & American Life Project, [http://www.pewinternet.org/PPF/r/144/report\\_display.asp](http://www.pewinternet.org/PPF/r/144/report_display.asp). 2005.
- [2] Xin Li., Lei Guo., Yihong Zhao, "Tag-based Social Interest Discovery," International World Wide Web Conference Committee (IW3C2), ACM 978-1-60558-085-2/08/04, 2008.
- [3] Anon Plangprasopchok., Kristina Lerman, "Exploiting social annotation for automatic resource discovery," AAAI workshop on Information Integration from the Web, 2007.
- [4] [http://en.wikipedia.org/wiki/FOAF\\_\(software\)#cite\\_note-0](http://en.wikipedia.org/wiki/FOAF_(software)#cite_note-0)
- [5] Maitrayee Mukherjee, Lawrence B. Holder, "Graph-based Data Mining on Social Networks," KDD'04, August 22 - 25, Seattle, WA, USA, 2004.
- [6] Katarzyna Musiał, Przemysław Kazienko, and Tomasz Kajdanowicz, "Social Recommendations within the Multimedia Sharing Systems," WSKS 2008, LNAI 5288, pp.364 - 372, 2008.
- [7] Amanda Lenhart, Kristen Purcell, Aaron Smith and Kathryn Zickuhr, "Social Media & Mobile Internet Use Among Teens and Young Adults," Pew Internet & American Life Project, 2010.
- [8] Mukund Deshpande and George Karypis, "Item-Based Top-N Recommendation Algorithms," ACM Transactions on Information Systems, Vol.22, pp.143 - 177, 2004.
- [9] Jun Wang, Arjen P. de Vries and Marcel J.T. Reinders, "Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion," SIGIR'06, August 6 - 11, 2006, Seattle, Washington, USA, 2006.
- [10] Iván Cantador, Alejandro Bellogín and David Vallet, "Content-based Recommendation in Social Tagging Systems," Fourth ACM conference on Recommender systems Barcelona, Spain, 2010.
- [11] Christian Desrosiers and George Karypis, "A Comprehensive Survey of Neighborhood-based Recommendation Methods," Springer Science & Business Media, LLC, 2011.
- [12] Noor Ali-Hasan, Lada A Adamic, "Expressing social relationships on the blog through links and comments," Proc. of International Conference on Weblogs and Social Media, 2007.
- [13] Michael F. Schwartz., David C. M. Wood, "Discovering shared interests using graph analysis," Communications of the ACM, 36(8):78 - 89, 1993.
- [14] Pu Wang and HongWu Ye, "A Personalized Recommendation Algorithm Combining Slope One Scheme and User Based Collaborative Filtering," International Conference on Industrial and Information Systems, 2009.
- [15] Ahmed Hassan, Dragomir Radev, Junghoo Cho and Amruta Joshi, "Content Based Recommendation and Summarization in the Blogosphere," Third International ICWSM Conference, 2009.

[16] Gábor Takács, István Pilászy, Bottyán Németh and Domonkos Tikk, "Scalable Collaborative Filtering Approaches for Large Recommender Systems," Journal of Machine Learning Research 10, 623-656, 2009.

[17] L. M. de Campos, J. M. Fernández-Luna, J. F. Huete and M. A. Rueda-Morales, "Using second-hand information in collaborative recommender systems," Soft Compute, 2010.

[18] Buhwan Jeong, Jaewook Lee and Hyunbo Cho, "An iterative semi-explicit rating method for building collaborative recommender systems," Expert Systems with Applications, Vol.36, Issue 3, Part 2, pp.6181-6186, 2009.

[19] Benjamin Heitmann and Conor Hayes, "Using Linked Data to Build Open, Collaborative Recommender Systems," Association for the Advancement of Artificial Intelligence, 2010.

[20] Luis M. de Campos, Juan M. Fernández-Luna, Juan F. Huete and Miguel A. Rueda-Morales, "Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks," International Journal of Approximate Reasoning Vol.51, Issue 7, pp.785-799, 2010.

[21] Toine Bogers, Antal Van Den Bosch, "Collaborative and Content-based Filtering for Item Recommendation on Social Bookmarking Websites," ACM RecSys '09 Workshop on Recommender Systems and the Social Web, New York, NY, USA, 2009

[22] <http://www.mybloglog.com/buzz/help/#a2005022821522712>

[23] <http://en.wikipedia.org/wiki/Delicious>

[24] <http://www.livejournal.com/site/about.bml>

[25] Ae-Ttie Ji, Cheol Yeon, Heung-Nam Kim, and Geun-Sik Jo, "Collaborative Tagging in Recommender Systems," AI 2007, LNAI 4830, pp.377 - 386, 2007.

[26] George Karypis, "Evaluation of Item-Based Top-N Recommendation Algorithms," Proceedings of the tenth international conference on Information and knowledge management, 2001.

[27] Mohsen Jamali and Martin Ester, "A random walk model for combining trust-based and item-based recommendation," 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009.

[28] Hoon-Ki Lee, Jung-Tae Kim, Jong-Hoon Lee and Eui-Hyun Paik, "Personalized Recommendation System for the Social Network Services Based on Psychographics," Fifth International Conference on Internet and Web Applications and Services, 2010.

[29] San-Yih Hwang, Chih-Ping Wei, Yu-chin Huang and Yun Tang, "Combining Coauthorship Network and Content for Literature Recommendation," Pacific Asia Conference on Information Systems, 2010.

[30] Jia Zhou, Tiejian Luo, "A novel approach to solve the sparsity problem in collaborative filtering," Networking, Sensing and Control (ICNSC), 2010 International Conference, pp.165-170, 10-12 April, 2010.

[31] Irwin, M. D., and Hughes, H. L., "Centrality and the structure of urban interaction: measures, concepts, and applications," Social Forces, 17-51, 1992.

[32] Casciaro, T., "Seeing things clearly: social structure, personality, and accuracy in social network perception," Social Networks, 20, 331-351, 1998.

[33] Scott, J., "Social Network Analysis," Sage publications, Thousand Oaks, CA, 1991.



### Qing Wang

e-mail : wangqing@korea.ac.kr

2008년 중국 North University of China, Department of Computer Science and Technology(학사)

2012년 고려대학교 컴퓨터정보학과 (공학석사)

2012년~현 재 중국 운남성 전력실험연구원 재직  
관심분야: 소셜 네트워크 분석, 시맨틱 웹, 인공지능 등



### 손종수

e-mail : mis026@korea.ac.kr

2003년 고려대학교 경영정보학과(학사)

2007년 고려대학교 전산학과(이학석사)

2007년~현 재 고려대학교 전산학과 박사과정

관심분야: 인공지능, 시맨틱 웹, 소셜 네트워크, 기계학습 등



### 정인정

e-mail : chung@korea.ac.kr

1978년 서울대학교 계산통계학과(학사)

1980년 한국과학원 전산학과(이학석사)

1989년 미국 University of IOWA, Computer Science(이학박사)

1990년~현 재 고려대학교 컴퓨터정보학과 교수

관심분야: 인공지능, 시맨틱 웹, 지능형 시스템 등