

# 소셜네트워크 기반의 콘텐츠 추천 방법

Yun-feng Pei<sup>†</sup> · 손 종 수<sup>\*\*</sup> · 정 인 정<sup>\*\*\*</sup>

## 요 약

최근 웹 및 웹 콘텐츠의 양이 폭발적으로 증가함에 따라서 콘텐츠 추천 시스템(CRS, Contents Recommendation System)은 최근 중요한 이슈로 대두되었다. 이에 따라, 콘텐츠 추천 시스템에 대한 콘텐츠 추천 방법(CRM, Contents Recommendation Method)이 꾸준히 연구 및 소개되어 왔다. 그러나 전통적인 CRM들은 콘텐츠 생성자의 위상이 중요하게 여겨지는 웹 2.0 환경에서 활용하는데 부족함이 있다. 본 논문에서는 연결 정도 중심성 분석(Degree of centrality) 및 TF-IDF를 활용하여 양질의 콘텐츠를 추천하는 방법을 제안한다. 이를 위하여 본 논문에서는 RSS와 FOAF를 수집하여 TF-IDF와 연결 정도 중심성을 각각 분석한다. 그리고 분석된 두 값을 이용하여 콘텐츠를 추천한다. 본 논문에서 제안한 방법을 검증하기 위하여 우리는 시스템을 구현하였으며 콘텐츠 추천 결과를 보인다. 본 논문에서 제안한 방법을 사용하면 입력된 질의어에 대해 사용자와 콘텐츠의 관계를 분석하고 이를 통해 적절한 콘텐츠를 추출할 수 있다. 그리고 본 논문에서 제안한 방법을 통해 구축한 시스템은 전통적인 콘텐츠 추천 시스템과 달리 소셜네트워크에서 콘텐츠 생산자에 대한 중요도가 반영됨으로 보다 신뢰성이 있는 결과를 얻을 수 있다.

키워드 : 소셜네트워크, 소셜네트워크 분석, 콘텐츠 추천 방법, TF-IDF, FOAF, RSS

## Contents Recommendation Method Based on Social Network

Yun-feng Pei<sup>†</sup> · Jong-soo Sohn<sup>\*\*</sup> · In-jeong Chung<sup>\*\*\*</sup>

## ABSTRACT

As the volume of internet and web contents have shown an explosive growth in recent years, lately contents recommendation system (CRS) has emerged as an important issue. Consequently, researches on contents recommendation method (CRM) for CRS have been conducted consistently. However, traditional CRMs have the limitations in that they are incapable of utilizing in web 2.0 environments where positions of content creators are important. In this paper, we suggest a novel way to recommend web contents of high quality using both degree of centrality and TF-IDF. For this purpose, we analyze TF-IDF and degree of centrality after collecting RSS and FOAF. Then we recommend contents using these two analyzed values. For the verification of the suggested method, we have developed the CRS and showed the results of contents recommendation. With the suggested idea we can analyze relations between users and contents on the entered query, and can consequently provide the appropriate contents to the user. Moreover, the implemented system we suggested in this paper can provide more reliable contents than traditional CRS because the importance of the role of content creators is reflected in the new system.

Keywords : Social Network, Social Network Analysis, Contents Recommendation Method, TF-IDF, FOAF, RSS

## 1. 서 론

오늘날의 인터넷은 웹 2.0의 출현으로 인하여 콘텐츠의 생산주체가 서비스 제공자에서 서비스 수요자인 사용자들로 변화 되면서 다양한 소셜네트워크 서비스(Social network service)가 각광받게 되었다[1][2]. 소셜네트워크란 개인을

의미하는 수많은 노드들로 구성된 사회적 관계망으로써 오늘날의 소셜네트워크 서비스는 사용자들이 직접 참여할 수 있고 상호 작용이 가능하게끔 발전되고 있다. 소셜네트워크 서비스를 제공하고 있는 웹 사이트에서는 웹 2.0 기술을 활용하여 사용자의 경험 및 지식에 대한 콘텐츠를 작성하고 배포할 수 있는 환경을 제공하고 있다. 이에 따라 사용자들의 축적된 경험은 콘텐츠의 품질에 큰 영향을 미치고 있으며 사용자들은 소셜네트워크를 통해 추천 받은 콘텐츠를 보다 선호하고 있다[2].

한편, 인터넷이 발달하고 이에 수반되는 각종 정보기술 및 통신 네트워크가 발달이 눈부시게 진행됨에 따라서 이에

† 준 회원 : 고려대학교 컴퓨터정보학과 석사과정  
 \*\* 준 회원 : 고려대학교 컴퓨터정보학과 박사과정  
 \*\*\* 종신회원 : 고려대학교 컴퓨터정보학과 교수  
 논문접수 : 2011년 4월 1일  
 수정일 : 1차 2011년 5월 17일  
 심사완료 : 2011년 5월 17일

수반되는 각종 매체와 콘텐츠의 양이 정보의 홍수(Deluge of information)로 표현될 만큼 폭발적으로 증가하고 있다. 이에 따라, 사용자들이 접할 수 있는 아이템의 수와 종류가 대단히 많아지고 있으며, 사용자들이 원하고 필요로 하는 아이템을 얻기가 점점 어려워지고 있다. 따라서 사용자가 원하고 필요로 하는 아이템을 알맞게 추천하는 콘텐츠 추천 시스템(CRS, Contents Recommendation System)이 최근 중요한 이슈로 대두되었으며, 이에 대한 콘텐츠 추천 방법(CRM, Contents Recommendation Method)이 활발히 연구, 소개되어왔다.

전통적인 콘텐츠 추천 방법에는 내용 기반 추천(Content-based recommendation) 방법[3][4]과 협업적 추천(Collaborative recommendation) 방법[5][6]이 있다. 그리고 위 두 가지 방법을 혼합한 하이브리드 추천(Hybrid recommendation) 방법이 있다[7][8][9]. 내용기반 추천 방법은 사용자의 선호도와 콘텐츠의 텍스트, 태그 등의 정보를 활용하여 사용자에게 적절한 콘텐츠를 추천하는 방법이다. 내용기반 추천 방법을 활용하면 사용자의 선호도에 기초한 콘텐츠를 선별하고 추천할 수 있지만 사용자들 간의 상호작용이 반영되지 못한다는 단점이 있다[10]. 그리고 협업적 추천 방법은 어떤 콘텐츠에 대해 여러 사용자가 협업적으로 평가하고 그 평가된 값에 의하여 콘텐츠를 추천하는 방법이다. 협업적 추천 방법은 콘텐츠를 추천하는데 있어 여러 사용자의 주관적인 평가값을 반영하기 때문에 사용자의 경험과 지식의 축적하는데 효율적이다. 그러나 주제에 대한 콘텐츠의 적합성 등을 평가하기 위한 객관적 근거가 없는 한계를 가지고 있다[10]. 마지막으로 하이브리드 추천 방법은 인공지능, 데이터마이닝 등에서 활용되는 알고리즘들을 콘텐츠 추천에 활용함으로써 언급한 두 추천 방법의 한계점을 극복하고자 한 방법들이다. 그러나 대부분의 하이브리드 콘텐츠 추천 방법들은 특정 도메인에 국한된 콘텐츠를 추천하는 것에만 활용되며 콘텐츠를 생성한 사용자의 신뢰성 및 위상이 반영되고 있지 않다[11].

이에 따라 본 논문에서는 기존의 콘텐츠 추천 방법의 한계를 극복하기 위하여 FOAF(Friend-of-A-Friend)[12]와 RSS(RDF Site Summary)[13]를 활용한 소셜네트워크 기반의 콘텐츠 추천 방법을 제안한다. 본 논문에서 제안하는 방법은 소셜네트워크를 분석하기 위하여 FOAF를 사용하며 콘텐츠를 분석하기 위하여 RSS를 활용한다. FOAF와 RSS 모두 웹2.0 환경의 서비스들에서 대부분 제공하고 있는 XML기반의 어휘들이다.

본 논문에서 제안하는 방법은 크게 4개 단계로 이루어져 있다. 첫 번째 단계는 질의어의 입력 및 데이터의 수집이다. 이 단계에서는 웹2.0 기반 서비스에서 제공하고 있는 FOAF와 RSS를 수집하여 데이터베이스에 저장한다. 그리고 두 번째 단계로써 TF-IDF를 이용한 RSS 콘텐츠 분석이다. 유사도의 측정을 위하여 질의어와 여러 콘텐츠들과의 유사성을 계산하는데 여기서 TF-IDF(Term Frequency - Inverse Document Frequency)기법[14]을 활용한다. 세 번째 단계에

서는 FOAF를 이용하여 각 사용자의 연결 정도 중심성(Degree of centrality)을 이용하여 소셜네트워크 분석을 한다. 그리고 마지막 네 번째 단계에서는 3단계와 4단계에서 두 값을 반영하여 각 콘텐츠에 순위를 매기고 관련된 콘텐츠를 추천한다. 본 논문에서 제안하는 콘텐츠 추천 방법은 소셜네트워크 상에서 사용자들이 차지하는 위상과 평판을 반영한다. 본 논문에서 제안한 방법을 활용하면 사용자들이 소셜네트워크에서 획득한 경험과 지식이 활발하게 공유될 수 있다. 그리고 급속히 발전하고 있는 소셜네트워크 서비스 상에서 사용자들에게 신뢰성이 있는 콘텐츠를 효과적으로 추천할 수 있다.

본 논문의 구성은 다음과 같다. 2장과 3장에서 배경 기술과 관련 연구를 보이며 4장에서는 본 논문에서 제안한 방법의 시스템 아키텍처와 소셜네트워크 기반의 콘텐츠 추천 방법에 대해 설명하며 5장에서는 본 논문에서 제안한 방법을 예제를 통해 설명한다. 6장에서는 본 논문에서 제안한 방법을 이용한 구현 예제를 보인다. 마지막으로 7장에서 본 논문에서 제안한 방법의 공헌도 및 향후 연구과제에 대하여 토의한다.

## 2. 배경 기술

### 2.1 RSS와 FOAF

FOAF는 하나의 RDF/OWL 어휘집으로 XML 기반의 온톨로지이다[12]. 사용자의 정보와 행동을 기술한 FOAF는 기계가 내용을 읽고 이해할 수 있으며 사람과 사람간의 관계, 사람과 개체들 간의 관계를 나타낸다. 그리고 이를 이용하여 온라인 상에서 사용자에게 관련된 정보를 쉽게 공유할 수 있다[12]. 본 논문에서는 FOAF를 이용하여 사용자들 간의 소셜네트워크를 표현한다. 그리고 이를 이용하여 사용자 연결 정도 중심성을 분석한다.

RSS는 RDF/XML 기반의 표준 웹 콘텐츠 출판언어이며 [13] 웹 상에서 주로 사용하는 콘텐츠 표현 방식이다. 사용자가 원하는 정보를 얻기 위해 해당 사이트를 직접 방문하여야 하는 기존의 방식과 달리 RSS는 등록된 신규 콘텐츠를 사용자의 PC 및 다른 웹 사이트, 모바일 어플리케이션 등에서 자동적으로 수신할 수 있도록 도구를 제공한다[6]. 본 논문에서는 RSS를 통해 사용자가 생성한 콘텐츠를 수집한다. 그리고 TF-IDF를 이용하여 RSS 콘텐츠와 입력 받은 질의어와의 유사도를 측정한다.

### 2.2 TF-IDF

TF-IDF는 수 많은 문서 중에서 질의어로 입력 받은 단어가 각 문서 내에서 차지하는 중요도를 나타내는 지표이다 [14]. 일반적으로 TF-IDF는 검색 엔진에서 검색 결과의 순위를 결정하는 용도로 사용한다[15]. TF-IDF는 TF와 IDF로 구성된다. TF(Term Frequency)는 질의된 키워드가 문서에서 나타나는 빈도수이다. 질의된 키워드가 문서에서 많이 나타날수록 해당 문서에서 중요하다고 볼 수 있다. DF

(Document Frequency)는 전체 문서들에서 공통적으로 나타나는 단어의 빈도수이다. 만약 어떤 단어의 DF가 높다면 그 단어는 관사 등과 같이 문서의 특징을 나타내는 것이 아님을 의미한다. 따라서 중요하지 않은 단어를 필터링하기 위하여 DF의 역(Inverse)인 IDF (Inverse Document Frequency)를 사용한다. TF-IDF의 값은 계산된 TF의 값과 IDF의 값의 곱으로 계산된다.

### 2.3 소셜네트워크 분석

소셜네트워크 분석(Social network analysis)은 소셜네트워크에서 존재하는 서로 독립적인 노드(Node)들 간의 연결 관계를 이용하여, 노드들의 중심성(Centrality) 혹은 밀도(Density)등 사회적 관계를 분석하는 방법이다. 소셜네트워크 분석 방법 중에서 한가지 중요한 방법인 연결 정도 중심성은 소셜네트워크를 구성하는 어느 노드와 연결된 다른 노드들과의 연결 정도를 측정하는 지표이다. 이는 사회학, 경제학, 마케팅, 조직관리 등 다양한 연구분야에서 활용된다[16].

## 3. 관련 연구

### 3.1 콘텐츠 추천 시스템

인터넷의 발달과 이에 따른 다양한 정보기술의 발전으로 인하여 인터넷 기반 매체와 콘텐츠 그리고 정보의 양은 홍수로 불릴 만큼 크게 증가하고 있다[17]. 그러나 폭발적으로 증가한 정보자원의 양에 반비례하여 사용자가 원하는 콘텐츠를 찾아내기는 점차 어려워지고 있는 것이 현실이다. 이에 따라 사용자에게 필요한 정보자원을 추천하는 CRS에 대한 연구가 다각도로 이루어지고 있으며 이는 다양한 분야의 웹 사이트에서 CRS를 활용이 가능하다. 특히 CRS는 영화, 비디오, 도서, 뉴스 기사 등과 같은 비디오 콘텐츠의 추천은 물론 음악 등과 같은 오디오 콘텐츠의 추천에서도 광범위하게 사용될 수 있다.

일반적인 CRS는 사용자 및 아이템이 주어졌을 때 사용자 프로파일을 어떤 기준에 따라서 분석, 평가하며, 이 평가의 순위 매김(Ranking) 결과에 따라서 사용자의 예상 선호도가 높은 아이템을 예측하여 추천한다. 이를 수식화하면 (식 1)과 같이 표현된다.

$$z = f(u, I) \tag{식 1}$$

(식 1)에서  $z$ 는 사용자 선호도를 나타내며  $u$ 는 사용자,  $i$ 는 아이템을 나타낸다. 사용자 선호도  $z$ 를 결정짓는 함수  $f$ 는 여러 다양한 알고리즘에 의해 결정된다.

그리고 추천시스템은 주어진 아이템 리스트의 각 항목에 대한 사용자의 선호도를 계산하고, 경우에 따라서 어떤 제약조건을 고려하여 사용자에게 적합한 아이템을 추천할 수 있으며, 다음 (식 2)와 같이 표현된다.

$$w = g(u, i, c) \tag{식 2}$$

(식 2)에서  $w$ 는 추천 리스트,  $c$ 는 제약조건을 나타낸다.

### 3.2 콘텐츠 추천 방법

콘텐츠 추천 시스템(CRS)의 구현을 위하여 사용자 선호도의 예측 및 추천을 위한 여러 가지 다양한 방법들이 있으며 크게 다음 리스트와 같이 분류할 수 있다.

- 내용 기반 콘텐츠 추천 방법
- 협업적 콘텐츠 추천 방법
- 하이브리드 콘텐츠 추천 방법

#### 3.2.1 내용 기반 추천 방법

내용 기반 추천 방법은 정보검색에 기반을 두었으며, 아이템의 콘텐츠를 직접 분석하여 아이템과 아이템, 아이템과 사용자 선호도 사이의 유사성을 분석하여 아이템을 추천해주는 방법이다[3]. 일반적으로 사용자의 선호도는 개인마다 크게 다를 수 있다. 그러므로 내용 기반 추천 방법에는 특성화된 개인적인 추천 리스트의 제공을 가능하게 하기 위하여 사용자 선호도 프로파일을 사용한다. 사용자 선호도 프로파일은 특성화된 사용자의 선호도를 각 아이템의 내용과 비교하여 사용자의 선호도가 높을 것으로 예상되는 아이템을 추천할 때 활용된다. 즉, 사용자 선호도 프로파일은 사용자가 과거에 구매했던 도서나 영화, 음악, 읽었던 기사 또는 방문했던 URL의 내용 등과 같이 사용자가 과거에 선택했던 아이템들로부터 그와 유사한 콘텐츠를 추천 할 수 있도록 하는 기초 자료로 활용된다.

내용 기반 추천 방법은 사용자 선호도의 예측 및 추천 과정이 비교적 쉬운 장점이 있으나, 다음과 같은 여러 가지 단점들이 있다. 첫째 사용자의 선호도와 유사한 아이템을 찾는 과정에서 과거의 아이템과 다른 새로운 아이템을 접할 기회가 제한된 추천 아이템이 과도하게 특성화 될 수 있다. 둘째 내용기반 추천 방법을 한정된 범위의 텍스트 콘텐츠에 적용하기는 쉬우나, 영화, 음악, 음식점 등과 같은 아이템의 영상, 음향, 맛 등의 멀티미디어적인 특징을 파악하기에는 어려움이 있다. 마지막 단점은 사용자 선호도 프로파일의 생성을 위한 사용자의 평가 결과를 얻기 위한 사용자의 참여를 얻기가 쉽지 않은 점이다. 이 마지막 단점은 반드시 내용기반 추천 방법뿐만 아니라 사용자 평가 결과가 반영되는 다른 추천 시스템에도 공통으로 해당되는 문제이다.

내용기반 콘텐츠 추천 방법의 대표적인 연구사례로는 [3]와 [4]이 있다. [3]에서는 아이템의 상세한 정보와 사용자 프로필의 관심사를 이용하여 다양한 아이템을 사용자에게 추천하는 방법을 제안하였다. 그리고 [4]에서 제안한 시스템은 위치기반 모바일 환경에서 사용자의 프로파일에 포함된 관심사와 업체의 웹 사이트에 포함된 내용을 비교한 후, 해당 위치 범위 내에 유사도가 높은 업체의 웹 사이트를 사용자에게 추천한다. 그러나 [3]와 [4]에서 제안한 내용기반 추천 방법은 사용자의 성향이 반영된 아이템을 추천할 수 있는 반면에 새로운 사용자에게 대하여 설정된 프로파일이 없기 때

문에 정확한 아이템을 사용자에게 추천하기 어렵다. 뿐만 아니라 사용자들간의 상호 작용이 없기 때문에 양질의 아이템을 추천하는데 있어 한계점을 보인다[10].

### 3.2.2 협업적 콘텐츠 추천 방법

협업적 콘텐츠 추천 방법은 비슷한 성향에 따라 사용자들을 분류한 후, 기존 사용자들의 아이템에 대한 선호도 및 평가를 바탕으로 현재 사용자에게 아이템을 추천하는 방법이다[15]. 즉, 영화나 음악 등의 콘텐츠를 제공하고 있는 사이트의 경우, 콘텐츠를 다운 받은 사용자에게 평점을 입력 받은 후에 평점이 높은 순으로 다시 사용자들에게 추천하는 방식이 협업적 콘텐츠 추천 방법이다. 대부분의 협업적 콘텐츠 추천 방법은 K-Nearest Neighborhood[18]나 개미 식민지 최적화(ACO: Ant Colony Optimization) 알고리즘[19][20]등과 같은 클러스터링 알고리즘, 집단지성 기반 기법들을 활용한다. 협업적 콘텐츠 추천 방법은 데이터가 충분할 때 정확도가 높은 아이템을 추천할 수 있다는 점에서 내용기반 추천 방법이 가진 한계점을 부분적으로 해결할 수 있다. 그러나 협업적 추천 방법은 사용자의 사용자 선호도 프로파일을 사용하지 않으므로 사용자에게 특화된 콘텐츠를 추천하기 어려운 단점을 가지고 있으며 콘텐츠의 내용 자체 보다는 사용자의 주관적인 평가에 의존하는 단점이 있다.

협업적 콘텐츠 추천 방법을 사용한 대표적인 사례로는 Tapestry[5]와 GroupLens[6]가 있다. Tapestry 시스템은 이메일 또는 넷뉴스(NetNews)로부터 수집된 사용자의 태그이나 의견을 필터링한 후, 사용자에게 따라 분류하여 사용자에게 콘텐츠를 추천하는 시스템이다. GroupLens 시스템은 사용자들이 뉴스에 매긴 등급을 통해 사용자간의 관계를 계산하여 그룹화한 후, 유사한 등급을 가진 사용자에게 좋아할 만한 뉴스를 추천하는 시스템이다. [5]과 [6]은 모두 사용자의 의견이나 평점을 기초로 다른 사용자들에게 유사한 콘텐츠를 추천한다는 점에 있어서 내용기반 콘텐츠 추천 방법의 문제점을 어느 정도 해결하였다. 반면, 사용자 선호도 프로파일을 사용하지 않음으로 인하여 내용기반 추천 방법의 장점을 활용하지 못한다는 단점이 있다. 그리고 협업적 추천 방법은 데이터가 많을수록 복잡한 연산이 필요하고, 아이템에 대한 평가가 바뀌어지는 현상과 상대적으로 적은 사용자의 수로 인해 아이템 추천함에 있어서 낮은 정확도가 보인다[10].

### 3.2.3 하이브리드 추천 방법

하이브리드 추천 방법은 내용기반 추천 방법이나 협업적 추천 방법의 한계점을 보완하기 위하여 데이터마이닝, 유전자 알고리즘, 인공지능망 등과 같은 다른 방법이나 기술과 결합시킨 콘텐츠 추천 방법이다[21]. 하이브리드 콘텐츠 추천 방법은 내용기반 콘텐츠 추천 방법과 협업적 콘텐츠 추천 방법의 단점을 상호 보완한 것으로 현재 많은 연구가 진행되고 있다. [7]에서 제안된 하이브리드 방법은 보편적으로 알려진 내용기반 추천 방법과 협업적 추천 방법을 결합하여, 추출된 사용자의 취향과 아이템의 속성 및 카테고리들을 통해

사용자에게 아이템을 추천한다. [8]에서는 아이템의 등급, 평가, 특성 및 인구통계학 정보를 통해 사용자에게 아이템을 추천하는 방법을 제안하였다. [9]에서는 웹 마이닝 기반으로 여러 가지 기타 추천 방법들의 장점을 결합한 새로운 추천 시스템을 제안하였고, 이에 대한 좋은 추천 성능을 보인다. [22]에서는 행렬 인수분해 기법(Matrix factorization technique)을 이용한 콘텐츠 추천 방법을 소개하였다. 그러나 [7][8][9][22]등에서 제안한 하이브리드 콘텐츠 추천 방법들은 특정한 사이트나 도메인에만 적용되어 제한된 아이템이나 콘텐츠를 사용자에게 제공하는 단점이 있다[11]. 그리고 콘텐츠 생성자 및 사용자들이 소셜네트워크에서 지닌 위상 및 평판 또한 콘텐츠를 평가할 때 중요하게 작용될 수 있으나 이에 대한 연구도 부족한 현실이다.

## 4. 소셜네트워크 기반 콘텐츠 추천 방법

### 4.1 시스템 아키텍처

본 논문에서 제안한 소셜네트워크 기반 콘텐츠 추천 방법은 다음 (그림 1)과 같은 아키텍처를 통해 구현되었으며 그 결과는 6장에서 보이는 바와 같다.

시스템은 크게 데이터 수집 부분과 콘텐츠 추천 부분으로 구성된다. 데이터 수집 부분은 소셜네트워크 서비스에 존재하는 사용자들의 FOAF 및 RSS를 수집한다. 그리고 본 논문에서 필요로 하는 데이터만을 필터링한 후 데이터베이스에 저장한다.

콘텐츠 추천 부분은 콘텐츠 추출, TF-IDF 분석, 소셜네트워크 분석, 콘텐츠 추천으로 구성된다. 콘텐츠 추출 부분은 사용자가 시스템 인터페이스를 통해 입력한 키워드를 이용하여 데이터베이스에 저장된 질의어에 관련 콘텐츠를 추출한다. TF-IDF 분석 부분에서는 TF-IDF를 이용하여 입력된 질의어가 추출된 콘텐츠들 중에서 각각 차지하는 가중치를 분석하고 계산한다. 소셜네트워크 분석 부분은 연결 정도 중심성의 계산을 수행한다. 이 부분에서 각 사용자의 네트워크가 질의어에 관련된 모든 사람들 중에서 차지하는 연결 정도 중심성을 계산한다.

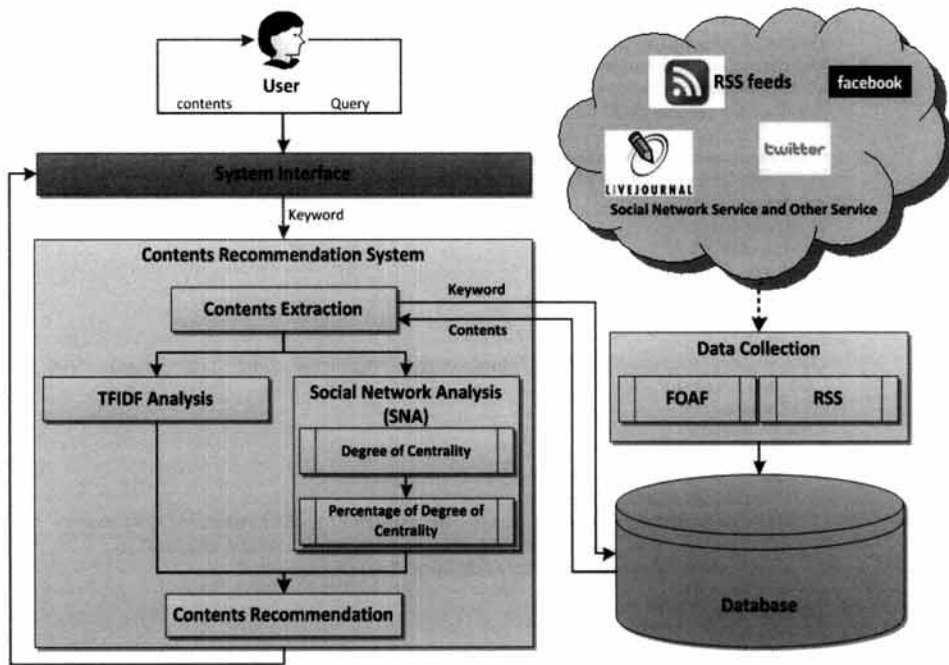
콘텐츠 추천 부분은 본 논문에서 제안한 소셜네트워크 기반 콘텐츠 추천 방법을 통해 계산된 콘텐츠 추천 값을 이용하여 상위 콘텐츠 순으로 콘텐츠를 정렬한다. 그리고 정렬된 콘텐츠를 사용자에게 반환한다.

### 4.2 소셜네트워크 기반 콘텐츠 추천 방법

본 논문에서 제안하는 소셜네트워크 기반 콘텐츠 추천 방법은 다음 보이는 바와 같이 총 4개 단계로 이루어진다.

#### • 단계 1: 질의어의 입력 및 데이터의 수집

1단계에서는 질의어를 입력 받으며 FOAF와 RSS 데이터를 수집한다. 질의어는 사용자가 추천 받고자 하는 키워드를 의미한다. 예를 들어 어떤 사용자가 'computer' 이라는 키워드에 대한 콘텐츠를 추천 받고자 할 때 시스템에



(그림 1) 시스템 구성

```

- <item>
  <guid isPermaLink="true">http://adafrog.livejournal.com/337514.html</guid>
  <pubDate>Tue, 14 Sep 2010 17:22:49 GMT</pubDate>
  <title>AO3 Invite</title>
  <link>http://adafrog.livejournal.com/337514.html</link>
  <description>Actually still have it. Leave a comment with your email, and I'll give
    screened.</description>
  <comments>http://adafrog.livejournal.com/337514.html</comments>
  <category>ao3</category>
  <category>flist</category>
  <dj:security>public</lj:security>
  <dj:reply-count>0</lj:reply-count>
</item>
    
```

(그림 2) RSS

입력한 'computer'이라는 단어가 질의어로 설정된다. 그 이후 시스템은 사용자들이 배포한 콘텐츠를 수집하기 위하여 소셜네트워크 서비스들에서 제공하고 있는 RSS를 수집한다. 4장 및 5장에서 보이는 예제와 구현된 시스템은 Facebook.com[23], LiveJournal.com[24], Twitter.com[25] 등에서 제공하고 있는 RSS 데이터를 활용하였다. 수집된 RSS는 (그림 2)에서 보이는 바와 같이 사용자가 작성한 콘텐츠의 제목, 작성 날짜, URI, 글 내용, 카테고리 등으로 구성되어 있다. 각각의 항목들은 모두 하나의 테이블로 데이터베이스에 저장된다.

그리고 FOAF 역시 대부분의 소셜네트워크 서비스에서 제공하고 있다. 사용자의 간단한 개인 정보와 선호 분야 등을 기술한 FOAF는 (그림 3)에서 보이는 바와 같다. FOAF에 포함된 사용자와 사용자간의 관계는 태그 <foaf:knows>로 나타난다. 시스템은 태그 <foaf:knows>를 통해 소셜네트워

크를 가상적으로 생성하며 이를 통해 연결 정도 중심성 분석을 수행한다. 그리고 <foaf:interest>는 해당 사용자의 관심사를 표현해주는 항목이다. <foaf:interest>는 입력된 질의어에 대하여 3단계에서 연결 정도 중심성 분석을 수행할 때 사용자 검색에 활용된다. <foaf:interest>와 <foaf:knows>는 데이터베이스에서 각각 다른 테이블에 저장되며 그 외의 나머지 FOAF 항목들 역시 다른 하나의 테이블에 저장된다.

• 단계 2: TF-IDF를 이용한 RSS 콘텐츠 분석

위 1단계에서는 RSS와 FOAF를 수집하고 소셜네트워크와 콘텐츠를 분석할 수 있도록 데이터베이스에 저장하는 과정을 보였다. 본 절에서 보이는 2단계는 사용자가 시스템에 검색하고자 하는 어떤 질의어를 입력하였다고 가정할 때 그 질의어와 가장 관련이 깊은 콘텐츠를 찾는 단계이다. 이를 위하여 본 논문에서는 정보검색 분야에서 널리 사용되고 있

```

- <foaf:Person>
  <foaf:nick>clannuisnigh</foaf:nick>
  <foaf:name>clannuisnigh</foaf:name>
  <foaf:openid rdf:resource="http://clannuisnigh.livejournal.com/" />
  <ya:country dc:title="UK" rdf:resource="http://www.livejournal.com/directory.bml?c
  <ya:city dc:title="London" rdf:resource="http://www.livejournal.com/directory.bml?
  opt_sort=ut&s_loc=1&loc_cn=UK&loc_st=&loc_ci=London" />
  <foaf:dateOfBirth>1983-02-19</foaf:dateOfBirth>
+ <foaf:page>
  <foaf:icqChatID />
  <foaf:aimChatID />
  <foaf:jabberID />
  <foaf:msnChatID />
  <foaf:yahooChatID />
+ <ya:blogActivity>
  <foaf:weblog rdf:resource="http://clannuisnigh.livejournal.com/" lj:dateCreated="200
  03-26T12:17:12" />
- <foaf:knows>
  - <foaf:Person>
    <foaf:nick>elfflame</foaf:nick>
    <foaf:member_name>elfflame</foaf:member_name>
    <foaf:tagLine>Little Fall of Rain</foaf:tagLine>
    <foaf:image>http://l-userpic.livejournal.com/33239411/3938525</foaf:image>
    <rdfs:seeAlso rdf:resource="http://elfflame.livejournal.com/data/foaf" />
    <foaf:weblog rdf:resource="http://elfflame.livejournal.com/" />
  </foaf:Person>
  </foaf:knows>
+ <foaf:knows>
+ <foaf:knows>
  
```

(그림 3) FOAF

는 TF-IDF 를 사용하여 입력된 질의어와 콘텐츠 사이의 유사도를 측정한다.

TF-IDF는 TF와 IDF로 구성된다. TF는 입력 받은 질의어가 특정 콘텐츠에서 차지하는 중요도를 평가하는 지표이다. TF는 다음 (식 3)과 같이 표현된다.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (식 3)$$

(식 3)에서  $i$ 는 사용자가 어떤 질의어를 입력했다고 했을 때 그 질의어를 의미한다. 그리고  $j$ 는 전체 콘텐츠 중에서 특정한 하나의 콘텐츠를 의미한다. 따라서  $tf_{i,j}$ 는 시스템에서  $i$ 라는 질의어를 입력 받았을 때 전체 단어 집합 중에서 입력 받은 질의어  $i$ 가  $j$ 문서에서 차지하는 중요도이다.  $n_{i,j}$ 는 입력 받은 질의어  $i$ 가 콘텐츠  $j$ 에서 나타난 횟수이며  $\sum_k n_{k,j}$ 는 콘텐츠  $j$ 가 포함하고 있는  $k$ 개의 전체 단어들  $i$ 가 나타난 횟수의 합이다. 즉,  $tf_{i,j}$ 는 특정 단어가 어떤 콘텐츠에서 차지하는 비중을 의미한다.

DF(Document Frequency)는 전체 콘텐츠 집합에서 특정 단어가 출현하는 빈도를 측정하는 값이다. 어떤 단어의 DF의 값이 높다면 많은 콘텐츠에서 그 단어를 사용하고 있음을 의미하며 이는 관사 등과 같이 흔히 쓰이는 단어이다. 따라서 TF-IDF에서는 TF값에 DF의 역수인 IDF (Inverse Document Frequency)를 곱하여 단어의 가중치를 적용한다. 다음 (식 4)는 IDF를 계산하는 식이다.

$$idf_i = \log \frac{|D|}{|\{d: t_i \in d\}|} \quad (식 4)$$

(식 4)에서  $idf_i$ 는 질의어  $i$ 를 포함한 콘텐츠가 전체 콘텐츠에서 차지하는 빈도수의 역수이다.  $|D|$ 는 전체 콘텐츠의 개수이며,  $|\{d: t_i \in d\}|$ 는 키워드  $i$ 를 포함한 콘텐츠의 개수이다.

(식 3)과 (식4)에서 각각 계산된 TF와 IDF의 값을 곱하면 (식5)와 같이 질의어  $i$ 가 콘텐츠  $j$ 에서 차지하는 가중치 TF-IDF값이 계산된다.

$$tfidf_{i,j} = tf_{i,j} \times idf_i \quad (식 5)$$

(식 5)에서 보인 계산을 거치면 사용자가 입력한 질의어에 대해 각 콘텐츠의 순위가 매겨진다.

• 단계 3: FOAF를 이용한 소셜네트워크 분석

본 단계에서는 소셜네트워크 분석 기법 중에서 한가지 중요한 기법인 연결 정도 중심성을 이용한다. 연결 정도 중심성이란 소셜네트워크를 구성하는 어느 사용자와 연결된 다른 사용자들과의 연결 정도를 측정하는 지표이다. 연결 정도 중심성으로 계산된 값은 각각의 사용자들이 소셜네트워크에서 다른 노드들과 얼마나 많은 연결이 있는지 표현한다. 본 절에서 보이는 연결 정도 중심성 분석은 질의어가 포함된 콘텐츠를 배포한 사람이 소셜네트워크에서 차지하는 중요도를 분석하는데 사용된다.

연결 정도 중심성은  $v$ 라는 사용자가 있을 때 (식 6)에서 보이는 바와 같이  $C_D(v)$ 로 표현된다.

$$C_D(v) = \frac{deg(v)}{n-1} \quad (식 6)$$

위 (식 6)에서  $deg(v)$ 는 사용자  $v$ 와 연결된 다른 사용자의 수이며  $n$ 은 전체 사용자의 수이다. 따라서  $C_D(v)$ 는  $v$ 와 연결된 사용자의 수를  $n$ 명의 사용자 중에서  $v$  자신을 제외한  $(n-1)$ 로 나눈 값으로 계산된다.

본 논문에서는 콘텐츠를 작성한 사람의 FOAF에서 <foaf:knows>를 이용하여  $C_D(v)$ 를 계산한다.  $C_D(v)$ 값은 다른 사용자와의 연결이 많으면 많을수록 1에 가까운 값이 나오게 되며 다른 사용자와의 연결이 적을수록 0에 가까운 값이 나오게 된다. 그리고 전체 사용자의 수가 많아지면 많아질수록  $C_D(v)$ 의 값은 전체적으로 0에 가까게 나오게 된다. 이런 경우 콘텐츠를 추천할 때 소셜네트워크에서의 중요도가 거의 반영되지 않는 문제가 있다. 따라서 본 논문에서는 질의어와 관련된 사람들 중에서 해당 사용자가 차지하는 비율을 계산하여 콘텐츠 추천에 그 값을 반영한다. 즉, 각 사용자의  $C_D(v)$ 를 질의어와 관련이 있는 전체 사용자  $C_D(v)$ 값의 합으로 나눈 후 그 값에 100을 곱하여 비율을 계산한다. 이는 다음 (식 7)를 통해 질의어와 관련 있는 사용자들 중에서 각 사용자가 차지하는 연결 정도 중심성 비율이 계산된다.

$$P_{i,m} = \frac{C_D(v_{i,m})}{\sum_{r=1}^n C_D(v_{i,r})} \times 100 \quad (\text{식 7})$$

(식 7)에서  $C_D(v_{i,m})$ 는 질의어  $i$ 와 관련된 콘텐츠를 배포한 사용자  $m$ 의 연결 정도 중심성 값이다.  $\sum_{r=1}^n C_D(v_{i,r})$ 는 질의어  $i$ 에 관련된 사용자가 총  $n$ 명 있을 때  $n$ 명의  $C_D(v)$ 의 합이다. 따라서  $P_{i,m}$ 은 질의어  $i$ 에 관련된 전체 사용자들의 연결 정도 중심성 합에서 특정 사용자  $m$ 의 연결 정도 중심성이 차지하는 비율이다.

$P_{i,m}$ 의 계산 결과를 통해 사용자가 입력한 질의어에 대한 콘텐츠 배포자의 순위를 결정한다.

• 단계 4: 콘텐츠 추천

단계 4에서는 위 단계 2와 3에서 계산된 TF-IDF와  $P_{i,m}$  값을 이용하여 사용자가 입력한 질의어에 대한 콘텐츠의 순위를 정한다. 즉, TF-IDF를 통해 계산된 콘텐츠의 중요도에  $P_{i,m}$ 을 통해 계산된 소셜네트워크에서의 연결 정도 중심성을 곱하여 콘텐츠 순위에 반영한다. 그리고 이 과정은 다음 (식 8)과 같다.

$$SR_{i,j,m} = tfidf_{i,j} \times P_{i,m} \quad (\text{식 8})$$

(식 8)에서  $SR$ (Social contents Recommendation)은 본 논문에서 제안하는 콘텐츠 추천 및 순위 결정 방법이다.  $SR_{i,j,m}$ 은 단계 2에서 보인  $tfidf_{i,j}$ 와 단계 3에서 보인  $P_{i,m}$ 의 곱으로 계산된다.  $tfidf_{i,j}$ 는 키워드  $i$ 에 대해 문서  $j$ 가 갖는 중요도를 표현한다. 그리고  $P_{i,m}$ 은 키워드  $i$ 에 대해 사용자  $m$ 가 차지하는 중요도를 표현한다. 따라서  $SR_{i,j,m}$ 은 키워드  $i$ 에 대하여 콘텐츠의 중요도와 소셜네트워크 안에서 해당 사

용자의 중요도가 모두 반영된 값으로 표현된다. 각 콘텐츠에 대하여  $SR$ 이 모두 계산되고 나면 콘텐츠의 순위를 정하여 사용자에게 그 결과를 출력한다.

4.3 콘텐츠 추천 알고리즘

본 논문에서 보인 소셜네트워크 기반 콘텐츠 추천 방법은 <표 1>에서 보이는 알고리즘을 통해 수행된다.

알고리즘은 사용자의 질의어  $i$ 를 입력 받으면 extract Contents() 함수를 이용하여  $i$ 가 포함된 콘텐츠를 추출한다. 그리고 TFcalculate() 함수를 이용하여 TF를 계산하며 degreeCentralityCalculate() 함수를 이용하여 연결 정도 중심성을 계산한다. 모든 콘텐츠에 대하여 이 계산들이 모두 수행되면 소셜네트워크 기반 콘텐츠 추천 값인  $SR$ 이 높은 순으로 콘텐츠 리스트를 반환한다.

<표 1> 소셜네트워크 기반 콘텐츠 추천 알고리즘

알고리즘1. Contents Recommendation	
Input: query $i$ Output: contents list	
1:	if( $i$ ) /*사용자에 의하여 입력된 질의어 $i$ */
2:	$d = \text{extractContents}(i)$ ; /*질의어에 관련된 콘텐츠를 추출하여, 변수 $d$ 에게 반환*/
3:	$idf = \text{IDFcalculate}(d.\text{count})$ ; /* IDF를 계산 */
4:	while( $j.\text{number} \leq d.\text{count}$ ) /*변수 $j$ 는 전체 콘텐츠 $d$ 에서 존재하는 특정 콘텐츠*/
5:	$tf = \text{TFcalculate}(i, j)$ ; /*키워드가 특정 콘텐츠 중에서 나타나는 빈도수를 계산한 후 변수 $TF$ 에게 반환*/
6:	$tfidf = tf * idf$ ; /*키워드가 특정 콘텐츠에서 차지하는 가중치를 계산*/
7:	$dc = \text{degreeCentralityCalculate}(j)$ ; /*특정 콘텐츠를 배포한 사용자의 연결 정도 중심성을 계산한 후, 변수 $DC$ 에게 반환*/
8:	$dcSum += dc$ ; /*키워드에 관련된 전체 콘텐츠들의 연결 정도 중심성의 합을 계산*/
9:	save " $j$ ", " $tfidf$ ", " $dc$ "; /*특정 콘텐츠에 관련된 TF-IDF값과 연결 정도 중심성을 데이터베이스에 저장*/
10:	$j.\text{number}++$ ;
11:	}
12:	while( $j.\text{number} \leq d.\text{count}$ ) 13: $p = j.dc / dcSum * 100$ ; /* 연결 정도 중심성 계산*/
14:	save $p$ about $j$ in the database; /*계산된 연결 정도 중심성 비율을 데이터베이스에 저장*/
15:	$SR = j.tfidf * j.p$ ;

```

/* 계산된 콘텐츠의 TF-IDF 및 사람의 연결 정도
중심성 비율을 통해 소셜네트워크가 반영된 새로운
콘텐츠의 값을 계산하여 변수 SR에 반환 */
16:   save SR about j;
/* 콘텐츠 j에 대한 소셜네트워크가 반영된 새로운
값 SR를 데이터베이스에 저장*/
17: )

18   sort by higher SR contents desc;
/*소셜네트워크가 반영된 새로운 콘텐츠의 값 SR를
높은 값을 갖는 순서대로 배열하여 사용자에게 추천*/
return contents;
19: )

Function extractContents()
Input: query i
Output: Contents list
1: if(tile and category include query "i"){
2:   extract contents from the database;
3:   return contents list;
4: }

Function TFcalculate();
Input: query i, content j
Output: tf
1: n = number of "i" in content "j";
2: totalWord = number of total words in content "j";
3: tf = n / totalWord;
4: return tf;

Function degreeCentralityCaculate()
Input: content j
Output: dc(degree of centrality)
1: deg = number of knows of "j";
2: n = number of all users;
3: dc = deg/(n-1);
4: return dc;

```

### 5. 예 제

본 장에서는 3장에서 설명한 첫 번째 단계부터 네 번째 단계까지의 소셜네트워크 기반의 콘텐츠 추천 방법을 예제를 들어 설명한다.

- 단계 1: 데이터의 수집 및 질의어의 입력

단계 1에서는 콘텐츠를 추천하기 위하여 데이터를 수집하고 사용자에게 질의어를 입력하는 과정이다. 본 예제에서는 다음과 같이 사용자와 질의어를 가정한다.

- 사용자 아이디: H0
- 질의어: computer
- 수집된 콘텐츠의 수: 1,000
- 수집된 FOAF의 수: 500

시스템은 Facebook, LiveJournal, Twitter 등과 같은 소셜 네트워크 서비스 사용자들의 FOAF 및 RSS를 수집한다. 만약 소셜네트워크 서비스에서 FOAF를 제공하지 않는다면 개방형 API(Open API)를 활용하여 사용자의 아이디와 연결 관계를 추출하며 이를 기반으로 기초적인 형태의 FOAF와 일을 작성한다. 1단계에서 보이는 데이터의 수집은 FOAF와 RSS를 모두 가지고 있지 않은 상태에서 시작하는 시스템을 기준으로 한다. 따라서 서비스 제공자와 같이 콘텐츠와 사용자의 정보를 모두 가지고 있는 경우에는 데이터 수집 단계를 생략할 수 있다.

콘텐츠 추천 시스템의 사용자는 키워드를 입력하며 시스템은 입력 받은 키워드를 이용하여 2 - 4 단계를 수행한다.

- 단계 2: TF-IDF를 이용한 RSS 콘텐츠 분석

위 단계 1을 통해 콘텐츠가 1,000개 수집되었다고 할 때 1,000개의 콘텐츠 중에서 다음 <표 2>과 같이 'computer'라는 단어를 포함하고 있는 5개의 콘텐츠가 있다고 가정한다.

<표 2> 키워드에 관련된 콘텐츠

콘텐츠	사용자	전체 단어의 개수	질의어가 나타난 횟수
C1	H1	100	5
C2	H2	80	3
C3	H3	200	8
C4	H4	50	3
C5	H5	60	1

<표 2>은 C1부터 C5까지 총 5개의 콘텐츠에 대해 각 사용자의 콘텐츠에서 추출된 전체 단어의 개수와 질의어가 나타난 횟수를 보인다. 여기서 (식3)에서 보인  $tf_{ij}$ 를 C1에 대해 계산하기 위하여  $i$ 에는 'computer'를 대입하며  $j$ 에는 C1을 대입한다. 따라서  $tf_{computer,C1}$ 는 다음과 같이 계산된다.

$$tf_{computer,C1} = (n_{computer,C1}) / (Total\ number\ of\ words)$$

$$tf_{computer,C1} = 5 / 100 = 0.05$$

그리고  $idf_{computer}$ 를 계산하기 위하여  $i$ 에 'computer'를 대입하고  $|D|$ 에 전체 콘텐츠의 개수를 그리고  $|d_i|$ 에 'computer'를 포함하고 있는 전체 콘텐츠의 개수를 대입한다. 이 때,  $idf_{computer}$ 는 다음과 같이 계산된다.

$$idf_{computer} = \log(1000 / 5) = 2.3$$

따라서  $tfidf_{computer,C1}$ 은 (식 4)에 의하여 다음과 같이 계산된다.

$$tfidf_{computer,C1} = tf_{computer,C1} * idf_{computer}$$

$$= 0.05 * 2.3$$

$$= 0.115$$



위에서  $tfidf_{computer,C1}$ 을 계산한 것과 마찬가지로 C2 부터 C5까지 모두 계산하면 <표 3>와 같다.

<표 3> TF-IDF 계산 결과

콘텐츠	사용자	TF	IDF	TF-IDF
C1	H1	0.0500	2.3	0.115
C2	H2	0.0375	2.3	0.086
C3	H3	0.0400	2.3	0.092
C4	H4	0.0600	2.3	0.138
C5	H5	0.0170	2.3	0.039

• 단계 3: FOAF를 이용한 소셜네트워크 분석

단계 3에서는 질의어 'computer'를 포함한 콘텐츠를 작성한 사용자가 소셜네트워크에서 다른 사람들에 비해 얼마나 중심에 속하는지를 측정한다. 이를 위하여 본 논문에서는 연결 정도 중심성을 계산하는데 이는 3장에서 보인 (식 6)과 같다.

위 예제 단계 1에서 보인 바와 같이 수집된 사용자의 FOAF는 50)개 이다. 여기서 (식 6)에서 보인  $C_D(v)$ 를  $H1$ 에 대해 계산하기 위하여  $v$ 에는 H1을 대입한다. 따라서  $C_D(H1)$ 는 다음과 같이 계산된다.

$$C_D(H1) = \text{deg}(H1) / ((\text{Total Number of Users}) - 1)$$

$$C_D(H1) = 30 / (500 - 1) \approx 0.06$$

$C_D(H1)$ 을 계산한 것과 마찬가지로 H2, H3, H4, H5를 각각 계산한 결과는 <표 4>와 같다.

<표 4> 연결 정도 중심성

사용자	<foaf:knows>의 개수	연결 정도 중심성
H1	30	0.06
H2	60	0.12
H3	25	0.05
H4	10	0.02
H5	80	0.16

그리고 계산된  $C_D(H1)$ 값을 이용하여 (식7)을 통해 질의어와 관련 있는 사용자들 중에서 각 사용자가 차지하는 연결 정도 중심성 비율이 계산된다. 여기서 (식 7)에서 보인  $P_{i,m}$ 에서  $i$ 에는 'computer'를 대입하며  $m$ 에는  $C_D(H1)$ 를 대입한다. 따라서  $P_{computer,H1}$ 은 다음과 같이 계산된다.

$$P_{computer,H1} = C_D(H1) / (C_D(H1) + C_D(H2) + C_D(H3) + C_D(H4) + C_D(H5)) * 100$$

$$P_{computer,H1} = 0.06 / (0.06 + 0.12 + 0.05 + 0.02 + 0.16) * 100$$

$$= 14.63$$

위에서  $P_{computer,H1}$ 을 계산한 것과 마찬가지로 H1부터 H5까지 모두 계산하면 <표 5>와 같다.

<표 5> 연결 정도 중심성 비율

사용자	연결 정도 중심성	$P_{i,m}$
H1	0.06	14.63
H2	0.12	29.26
H3	0.05	12.19
H4	0.02	4.87
H5	0.16	39.02

• 단계 4: 콘텐츠 추천

단계 4에서는 위 단계 2와 3에서 계산된 TF-IDF와  $P_{i,m}$  값을 이용하여 사용자가 입력한 질의어에 대한 콘텐츠의 순위를 정한다. 이를 위하여 본 논문에서는 콘텐츠의 중요도와 소셜네트워크 안에서 해당 사용자의 중요도가 모두 반영된 값을 계산하는데 이는 3장에서 보인 (식 8)로 계산된다.

<표 6>는 C1부터 C5까지의 콘텐츠에 대해 각 콘텐츠의 TF-IDF와 이를 배포한 사용자의 연결 정도 중심성 비율이 보인다. 여기서 (식 8)에서 보인  $SR_{i,j,m}$ 를 C1에 대해 계산하기 위하여  $i$ 에는 'computer'를 대입하고  $j$ 에는 C1를 대입하며  $m$ 에는 H1을 대입한다. 따라서  $SR_{computer,C1,H1}$ 는 다음과 같이 계산된다.

$$SR_{computer,C1,H1} = tfidf_{computer,C1} * P_{computer,H1}$$

$$SR_{computer,C1,H1} = 0.115 * 14.63$$

$$\approx 1.682$$

위에서  $SR_{computer,C1,H1}$ 을 계산한 것과 마찬가지로 C2부터 C5까지 모두 계산하면 <표 6>과 같다.

<표 6> 각 콘텐츠의  $SR_{i,j,m}$  계산 결과

콘텐츠	사용자	TF-IDF	$P_{i,m}$	$SR_{i,j,m}$
C1	H1	0.115	14.63	1.682
C2	H2	0.086	29.26	2.525
C3	H3	0.092	12.19	1.121
C4	H4	0.138	4.87	0.672
C5	H5	0.039	39.02	1.525

마지막으로 각 콘텐츠에 대하여  $SR_{i,j,m}$ 이 모두 계산되고 나면 콘텐츠의 순위를 정하여 사용자에게 그 결과를 출력한다.

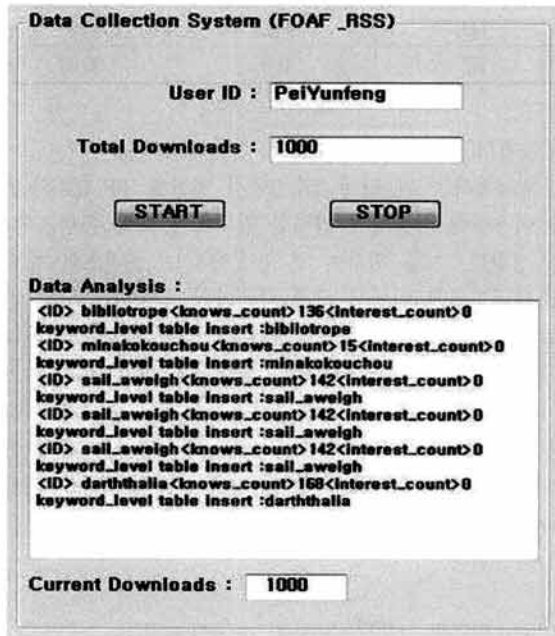
6. 시스템 구현 및 평가

제안한 소셜네트워크 기반 콘텐츠 추천 방법을 검증하기 위하여 본 논문에서는 시스템을 구현하였다. 제안한 시스템은 데이터 수집부분과 콘텐츠 추천 부분으로 구성된다. 데이터 수집부분의 개발환경은 마이크로소프트 비주얼 스튜디오

오 2008과 마이크로소프트 SQL서버를 이용하였으며 개발언어는 C#를 이용하였다. 콘텐츠 추천부분의 개발환경은 이클립스(Eclipse)과 PHP 5 언어 및 HTML언어를 이용하였다.

6.1 데이터의 수집

본 논문에서 구축한 시스템의 데이터 수집부분의 인터페이스는 (그림 4)에서 보인다.



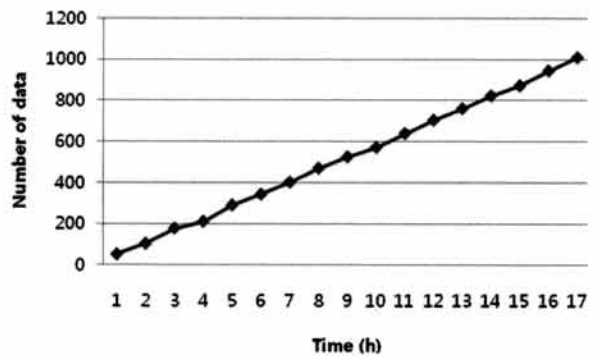
(그림 4) 데이터 수집 인터페이스

데이터 수집부분에서 시스템은 입력 받은 사용자의 아이디를 통해 사용자 자신 및 관계가 있는 사람들의 FOAF 및 RSS를 소셜네트워크 서비스에서 제공한 API를 통해 다운받는다. 다운 받은 FOAF 및 RSS들은 시스템을 통해 분석한 후 필요한 데이터를 추출하여 데이터베이스에 저장한다.

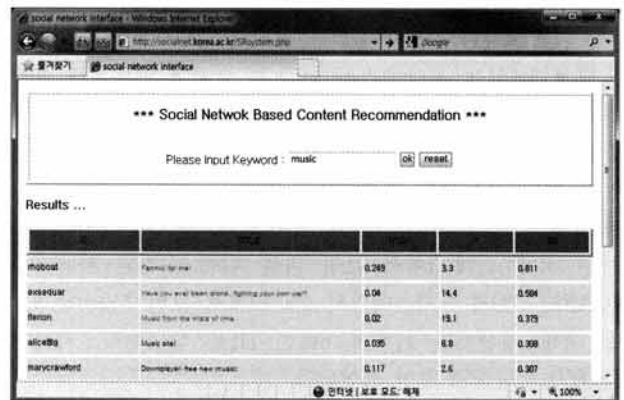
본 논문에서 보이는 방법의 구현을 위하여 우리는 1,012개의 FOAF와 12,373개의 RSS를 수집하였다. 그리고 수집한 FOAF를 활용하여 109,426개의 foaf:know를 수집하였다. 대부분의 소셜네트워크 서비스에서는 RSS를 제공하고 있으므로 RSS의 수집에는 문제가 없으나 만약 사용자의 FOAF를 제공하지 않는 경우에는 본 논문에서 제안한 방법을 활용하기 어렵다. 따라서 우리는 소셜네트워크 서비스에서 제공하고 있는 개방형 API를 활용하여 FOAF를 수집하지 못한 사용자의 아이디와 다른 사용자와의 연결관계를 수집하고 이를 통해 기초적인 형태의 FOAF를 작성하였다. 따라서 본 논문에서 보이는 모든 데이터는 각 사용자에 대하여 FOAF와 RSS가 모두 함께 저장되어있다.

그리고 FOAF와 RSS의 데이터 수집 시간은 인텔 펜티엄4, 메모리 1기가 시스템을 기준으로 (그림 5)에서 보이는 바와 같이 1000명의 데이터를 수집하는데 약 16시간이 소요된다.

본 논문에서 제안한 콘텐츠의 추천 방법은 FOAF와 RSS를 수집한 후 질의어를 입력 받으므로 시스템의 동작속도에



(그림 5) 데이터의 수집 시간



(그림 6) 콘텐츠 추천

는 큰 영향을 미치지 않는다. 그리고 한번 수집한 데이터는 삭제되지 않고 누적되므로 시스템의 가동시간이 늘어날수록 사용 가능한 데이터의 수가 계속적으로 증가한다. 또한 등록된 FOAF의 사용자가 콘텐츠를 작성하면 실시간으로 업데이트된 RSS가 수집되어 추천 결과에 반영된다.

콘텐츠 추천 시스템 인터페이스

본 논문에서 구축한 시스템의 콘텐츠 추천부분의 인터페이스는 (그림 6)와 같다. 사용자는 콘텐츠를 추천을 이용하여 찾고자 하는 콘텐츠에 대한 키워드 'music'를 시스템에 입력한다. 시스템은 입력된 키워드를 이용하여 구축된 소셜네트워크를 바탕으로 키워드에 관련된 콘텐츠를 추출하고, 본 논문에서 제안한 기법을 통해 사용자에게 키워드에 관련된 콘텐츠 중에서 내림차순으로 추천한다.

추천된 콘텐츠 중에는 콘텐츠를 작성한 사람의 아이디, 콘텐츠에 대한 링크가 포함된다. 사용자는 추천된 콘텐츠의 링크를 이용하여 보다 쉽게 콘텐츠를 접근할 수 있다.

7. 결론

최근의 인터넷에서 사용자들은 소셜네트워크 등을 이용하여 다른 사용자들과 경험과 지식을 공유하고 있다. 사용자들의 활동은 대부분 소셜네트워크 서비스를 이용하면서 이루어지며 소셜네트워크를 통해 협업적 콘텐츠를 생성하고 있다. 이에 따라 무한대에 가까운 정보들 가운데서 사용자에게 필요한 정보를 정확하게 검색하고 추천하기 위한 콘텐츠

츠 추천 방법이 활발한 연구가 진행되고 있다. 그러나 기존에 소개된 콘텐츠 추천 방법들은 사용자들 간의 상호작용이 반영되지 못하는 단점이 있으며 소셜네트워크에서 사용자가 가지는 위상 및 평판을 반영하지 못하는 단점이 있다.

이에 따라 본 논문에서는 기존 콘텐츠 추천 방법들의 문제점을 극복하고 소셜네트워크에서 콘텐츠를 생성한 사용자의 위상을 반영하기 위한 콘텐츠 추천 방법을 제안하였다. 본 논문에서 제안한 콘텐츠 추천 방법은 질의어의 입력 및 데이터의 수집 단계, TF-IDF를 이용한 RSS콘텐츠 분석 단계, FOAF를 이용한 소셜네트워크 분석 단계, 콘텐츠 추천 단계를 통하여 질의어에 관련된 콘텐츠들을 분석한 후 사용자의 위상 및 평판이 반영된 신뢰성이 있는 적절한 콘텐츠를 사용자에게 추천한다.

본 논문에서 제안한 콘텐츠 추천 방법은 기존 콘텐츠 추천 방법에서는 사용하지 않았던 연결 정도 중심성을 활용하였다. 전통적인 웹에서는 콘텐츠를 서비스 제공자가 생성하였으므로 콘텐츠 생성자의 연결 정도 중심성은 중요하게 여겨지지 않았다. 그러나 최근의 웹에서는 콘텐츠의 생성자가 사용자로 변화하게 되었으므로 콘텐츠의 추천에 있어서도 사용자의 위상과 평판이 중요하게 영향을 미친다. 따라서 본 논문에서 제안한 방법은 기존에 제안된 많은 콘텐츠 추천 방법이 반영하지 못했던 소셜네트워크를 반영하고 한계점을 극복한 콘텐츠 추천 방법이다.

본 논문에서 제안한 방법을 활용하면 사용자들의 경험과 지식이 활발하게 공유되고 협업적으로 콘텐츠를 생성하고 보완 할뿐만 아니라 최근의 급속히 발전하고 있는 소셜네트워크 서비스 상에서 신뢰성이 있는 콘텐츠를 사용자에게 효과적으로 추천할 수 있다. 뿐만 아니라 소셜네트워크 바탕으로 사용자들간의 콘텐츠 공유를 더욱 활발하게 촉진 시킬 수 있다.

## 참 고 문 헌

- [1] A. L. Barabási, H. Jeong, E. Ravasz, Z. Neda, A. Schuberts and T. Vicsek, "Evolution of the Social Network of Scientific Collaborations," *Physica A*, pp.590-614, 2002.
- [2] R. Kumar, P. Ragbavan, S. Rajagopalan and A. Tomkins, "The Web and Social Networks," *Computer*, pp.32-36, 2002.
- [3] M. J. Pazzani and D. Billsus, "Content-Based Recommendation Systems," *The Adaptive Web Lecture Notes in Computer Science*, Vol.4321/2007, pp.325-341, 2007.
- [4] W. S. Yang, H. C. Cheng and J. B. Dia, "A Location-Aware Recommender System for Mobile Shopping Environments," *Expert Systems with Applications*, Vol.34, No.1, pp.437-445, Jan. 2008.
- [5] D. Goldberg, D. Nichols, B. M. Oki and D. Terry, "Using Collaborative Filtering to Weave an Information Tapestry," *Communications of the ACM*, Vol.35, pp.61-70, 1992.
- [6] J. A. Konstan, B. N. Miller, D.Maltz, J. L. Herlocker, L. R. Gordon and J. Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News," *Communications of the ACM*, Vol.40, No.3, pp.77-87, Mar., 1997.
- [7] A. Albadvi and M. Shahbazi, "A Hybrid Recommendation Technique Based on Product Category Attributes," *Expert Systems with Applications*, Vol.36, No.9, pp.11480-11488, Nov., 2009.
- [8] M. A. Ghazanfar, A. Prugel-Bennett, "A Scalable, Accurate Hybrid Recommender System," *Knowledge Discovery and Data Mining*, pp.94-98, 2010.
- [9] M. Göksedef and Ş. Gündüz-Öğüdücü, "Combination of Web Page Recommender Systems," *Expert Systems with Applications*, Vol.37, No.4, pp.2911-2922, Apr., 2010.
- [10] M. Montaner, B. López and J. L. de la Rosa, "A Taxonomy of Recommender Agents on the Internet," *Artificial Intelligence Review*, Vol.19, No.4, pp.285-330, 2003.
- [11] M. Zanker and M. Jessenitschnig, "Collaborative Feature-Combination Recommender Exploiting Explicit and Implicit User Feedback," *IEEE Conference on Commerce and Enterprise Computing*, pp.49-56, 2009.
- [12] <http://www.foaf-project.org>
- [13] <http://www.rssboard.org/rss-specification>
- [14] G. Salton and C. Buckley, "Term Weighting Approaches in Automatic Text Retrieval," *Information Processing and Management*, Vol.24, No.5, pp.513-523, 1988.
- [15] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative Filtering Recommender Systems," *The Adaptive Web Lecture Notes in Computer Science*, Vol.4321, pp.291-324, 2007.
- [16] J. Scott, "Social Network Analysis," SAGE Publications, 2000.
- [17] G. Lawton, "Knowledge Management: Ready for Prime Time?," *IEEE Computer*, Vol.34, No.2, pp.216-244, 2001.
- [18] J. Suchal and P. Návrat, "Full Text Search Engine as Scalable K-Nearest Neighbor Recommendation System," *IFIP WCC Series*, Vol.331, pp.165-173, 2010.
- [19] T. I. Wang, K. T. Wang and Y. M. Huang, "Using a Style-Based Ant Colony System for Adaptive Learning," *Expert Systems with Applications*, Vol.34, No.4, pp.2449-2464, May, 2008.
- [20] Y. J. Yang and C. Wua, "An Attribute-Based Ant Colony System for Adaptive Learning Object Recommendation," *Expert Systems with Applications*, Vol.36, No.2, pp.3034-3047, Mar., 2009.
- [21] R. Burke, "Hybrid Web Recommender Systems," *The Adaptive Web, Lecture Notes In Computer Science*, Vol. 4321. Springer-Verlag, Berlin, Heidelberg, pp.377-408, 2007.
- [22] Y. Koren, R. Bell and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *IEEE Computer*, Vol.42, No.8, pp.30-37, Aug., 2009.
- [23] <http://www.facebook.com>
- [24] <http://www.livejournal.com>
- [25] <http://www.twiiter.com>



### Yun-feng Pei

e-mail : beayunfeng@korea.ac.kr  
 2009년 중국 Northeastern University,  
 Software Engineering(학사)  
 2010년~현 재 고려대학교 컴퓨터정보학과  
 석사과정  
 관심분야: 인공지능, 시맨틱 웹, 소셜네트  
 워크 등



### 정인정

e-mail : chung@korea.ac.kr  
 1978년 서울대학교 전자계산학(학사)  
 1980년 한국과학원 전산학과(석사)  
 1989년 미국 Univ. of Iowa, Computer  
 Science(박사)  
 1990년~현 재 고려대학교 컴퓨터정보학과  
 교수

관심분야: 인공지능, 시맨틱 웹, 소셜네트워크, 지능형 웹서비스 등



### 손종수

e-mail : mis026@korea.ac.kr  
 2003년 고려대학교 경영정보학과(학사)  
 2005년 고려대학교 전산학과(이학석사)  
 2005년~현 재 고려대학교 컴퓨터정보학과  
 박사과정  
 관심분야: 인공지능, 시맨틱 웹, 소셜네트  
 워크, 기계학습 등