

소셜 네트워크 기반 팟캐스트 검색시스템[☆]

Social Network based Podcast Search System

정 옥 란^{1*}
Ok-Ran Jeong

요 약

SNS와 스마트 기기가 보편화 되면서, 뉴미디어로 떠오르고 있는 팟캐스트 이용자가 계속적으로 증가함에 따라 전문 검색서비스의 필요성이 대두되고 있다. 본 연구에서는 사용자가 팟캐스트를 검색할 때, 사용자들의 소셜 네트워크 정보를 이용하여 관심도가 높은만한 팟캐스트를 추천하는 시스템을 설계 및 구현하였다. 제안한 소셜 네트워크 기반 팟캐스트 검색시스템(PODSSO)은 필요한 팟캐스트 정보만을 웹에서 수집하고, 사용자들의 소셜 네트워크를 분석하여, 사용자들에게 보다 신뢰성 있고 관심있는 팟캐스트 검색 결과를 얻을 수 있게 해준다.

주제어 : 소셜 네트워크, 팟캐스트, 추천, 검색

ABSTRACT

As the number of podcast users consistently increases which is rising as a new media along with the generalization of SNS and smart devices, the necessity for advanced search service is on the rise. This study designed and implemented a system which recommends a podcast to the users who search podcast by using their social network information. Suggested social network-based podcast search system (PODSSO) collects necessary podcast information only, analyzes social network of the users and makes the users have reliable and interested podcast search results.

☞ keyword : Social Network, Podcast, Recommendation, Search

1. 서 론

웹 기반 디지털 산업이 국내의 경제의 중요한 역할을 담당하며 차세대 기술 환경에서 소셜 네트워크 기반 기술이 급속도로 발전하고 있다. 웹의 발달과 이에 따른 다양한 네트워크 및 인터넷 기술 발전으로 인터넷 기반 디지털 매체와 그에 따른 콘텐츠들이 폭발적으로 증가하고 있다[1]. 대량으로 증가하는 인터넷 정보자원에서 사용자들이 찾고자하는 정보들을 찾기에는 점차 더욱 어려워지는 있는 현실이다. 이에 따라 사용자들이 필요한 정보 자원을 효율적으로 검색하거나 추천하는 정보기술이 뒷받침되어야 할 것이다. 구글, 야후, 네이버 등 일반적인 검색은 많이 발전되어 가고 있지만, 반면에 사용자의 특

성이나 사용자에 초점을 맞춘 검색은 다양한 분야에서 어려운 현실이다.

특히 팟캐스트는 동영상 파일 형태이므로, 사용자가 원하는 정보 여부를 시청하기 전에는 판별하기 어렵다. 이러한 불편이 있는 상황이지만, 사용자들에게 필요한 팟캐스트 파일을 전문적으로 검색이나 추천해주는 연구는 아직까지 미흡한 상황이다.

본 연구에서는 SNS의 활성화에 따른 소셜 네트워크 사이트에서 사용자들의 소셜 정보를 이용하여, 새로운

소셜 커뮤니티로 떠오르는 팟캐스트 검색시스템을 구현하여 사용자들에게 유용한 검색 결과를 얻을 수 있게 하였다. 이를 수행하기 위해 팟캐스트 정보만을 웹에서 구별하여 팟캐스트 크롤링을 이용하여 수집하고, 소셜 네트워크 정보를 추출하기 위해 대표적인 소셜 네트워크 사이트에서 제공하는 Open API를 이용하여 사용자 소셜 정보를 수집하였다. SNS사이트에서 추출된 인물관계, 작성한 글, 관심사를 고려하여, 사용자 성향에 따라 팟캐스트를 추천해준다. 제안한 시스템은 PODSSO(팟쇼, Social Network based PodCast Serarch System)이며, 다음과 같이 세부분으로 구성되었다.

¹ Dept. of Software Design & Management, Gachon University, Gyeonggi-Do, 461-701, Korea

* Corresponding author (orjeong@gachon.ac.kr)

☆ 이 논문은 2013년도 가천대학교 교내연구비 지원에 의한 결과이며, 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업지원금을 받아 수행된 것임. (2012-0004177)

[Received 12 March 2013, Reviewed 15 March 2013, Accepted 16 April 2013]

첫째, 팟캐스트 크롤러(podcast Crawler): 웹사이트로부터 팟캐스트 정보만을 효율적으로 수집할 수 있는 팟캐스트 크롤러를 설계 및 구현하였다. 수집 시 고려해야 하는 팟캐스트 여부 판별 기능, 최신 정보를 위한 재방문 주기 최적화와 팟캐스트 변경 감지 기능 등을 갖추도록 하였다.

둘째, 소셜 정보 추출(SNS Data Extractor): 사용자들에게 소셜 정보를 검색시스템에 적용하기 위해 소셜 네트워크상에 관계를 맺고 있는 커뮤니티 관련 소셜정보를 추출하였다. SNS 사이트에서 사용자들의 소셜 네트워크 정보를 확보하기 위해, SNS사이트에서 제공하는 Open API를 사용하였고, 실제 연구에서 적용한 사이트는 Facebook, Twitter, Me2day이다.

셋째, 팟캐스트 추천(Podcast Recommender): 사용자 성향에 따라 가장 적합한 팟캐스트를 추천해주는 추천알고리즘을 구현하여, 제안한 시스템에 적용하였다. 추천알고리즘은 SNS 사이트에서 인물관계, 작성한 글, 관심사를 고려하였다.

본 논문의 구성은 다음과 같다. 2장에서는 소셜네트워크, 팟캐스트, 웹 콘텐츠 추천방법에 대한 관련연구를 설명하고, 3장에서는 본 연구에서 제안하는 팟캐스트 검색 시스템에 대해 구성 및 단계별 팟캐스트 검색 방법에 대해 설명한다. 4장에서는 단계별로 실제 구현된 기술내용에 대해 설명하고, 5장에서는 실험 및 평가를 하였으며, 6장에서 결론을 서술한다.

2. 관련연구

2.1 소셜 네트워크

소셜네트워크는 웹상에서 공통된 관심사를 가지고 있는 사용자들 간의 관계를 쌓으며, 폭넓은 네트워크를 형성할 수 있도록 해주며, 최근 웹 2.0기반으로 다양한 SNS가 제공되고 있다. 빠르게 보급되는 커뮤니티 기반의 웹 사이트들을 통해 사용자들은 서로 관계를 맺고 즐겨찾기, 음악, 블로그, 사진 등을 서로 공유한다. 최근 모바일 스마트 기기의 빠른 대량 확산으로 인해 소셜 네트워크를 실시간으로 이용하는 사용자들이 급증하고 있다. [2]에서는 소셜 네트워크를 기반으로 하는 소셜 커뮤니티 사이트를 대상으로 블로그 검색 프레임워크를 제안하였다. 소셜 커뮤니티에서 집단지성의 특성을 추출하고, 이를 이용하여 질의 확장을 통하여 주어진 주제를 보다 깊이 있게 소셜 검색을 고려함으로써 사용자에게 보다

신뢰성 있고 유용한 검색결과를 보여주었다.

2.2 팟캐스트

팟캐스트는 애플의 아이팟(Ipod)과 방송(Broadcasting)을 결합해 만든 신조어로, 오디오 또는 비디오 파일 형태로 뉴스나 각종 콘텐츠를 제공하는 것을 말한다. 기존 라디오와 달리, 방송 시간에 맞춰 들을 필요없이 모든 프로그램이 자동으로 업데이트되므로 그 가운데 관심 있는 프로그램을 다운로드하여 시청할 수 있다. 국내에서도 공개 미디어 서비스 중에서 최근 팟캐스트를 통해 사회적/정치적 이슈를 일반적인 언론에서보다 좀 더 자유롭게 풍자하는 콘텐츠가 사용자들의 관심을 갖게 되면서, 팟캐스트 이용률이 증가하게 되었다. 하지만 팟캐스트만을 대상으로 하는 검색이나 추천서비스는 이뤄지지 않고 있어, 사용자 접근이 좋지 않은 것이 현실이다[3]. 해외에서는 팟캐스트를 주로 교육 보조재로서의 기능성에 주목하고 있다. 고드윈-존스(Godwin-Jones)는 언어교육도구로서 팟캐스트 기능성에 대해 다루었고[4], 리차드슨(Richardson)은 팟캐스트를 비롯하여 위키나 블로그 등 웹 2.0을 특징짓는 미디어들의 교육적 활용 방안에 대해서 언급하였고[5], 볼로스(Boulos)등은 의료교육현장에서 웹 2.0 미디어들의 활용방안에 대해 연구하였다.[6] 그 외에 상업적 수익모델로서의 RSS의 가능성과 기술적 문제점에 대해 연구되어왔다[7].

2.3 웹 콘텐츠 추천방법

SNS 커뮤니티 웹 사이트 등에서 대량의 웹 콘텐츠를 이용하기 위해서 검색이나 추천은 중요한 역할을 하고 있다. 사용자들이 웹 사이트들의 콘텐츠 중에서 본인이 찾는 아이템, 뉴스, 음악 등을 좀 더 쉽게 찾기 위해 추천 시스템의 역할이 갈수록 중요해지고 있다[8].

효율적인 검색을 위한 추천시스템은 도서, 영화, 미디어, 뉴스, 기사, 팟캐스트 등 다양한 분야에서 광범위하게 사용될 수 있다. 이러한 검색을 하기위해서 사용자의 프로파일 및 사용 성향을 분석하여, 선호도의 예측한 후, 사용자에게 맞는 내용을 추천하는 방식이다. 사용자들이 만족하는 검색시스템은 추천방식을 통해 가능하다. 대표적인 기존연구 방법은 내용기반 콘텐츠 추천방법, 협업적 콘텐츠 추천방법과 하이브리드 추천 방법이 있다.

내용기반 콘텐츠 추천방법은 사용자의 선호도를 분석하여 추천 아이템과의 유사도를 비교하여, 선호도가 높은 아이템을 추천해주는 방식이다[9]. 이러한 방법은 사

용자가 속한 소셜 커뮤니티에서 비슷한 유형의 아이템 중 선호도가 높은 아이템을 추천해주는 방법도 제안되었다[10].

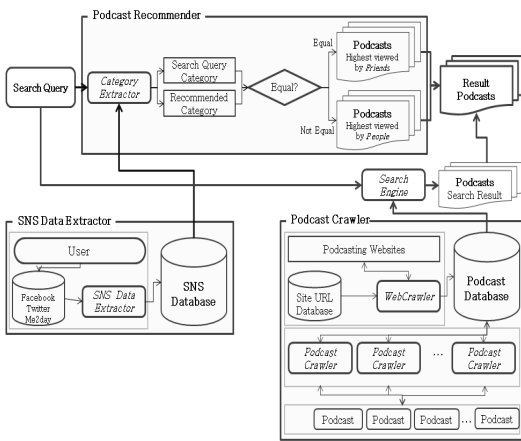
협업적 추천 방법은 취향이 비슷한 기존 사용자들에 의해 점수가 매겨진 내용을 기반으로 현재 사용자에게 아이템을 추천하는 방법이다[11]. 인터넷 콘텐츠를 제공하는 사이트에서 콘텐츠를 다운받아 사용하는 사용자들에게 일정기준의 점수를 제공받아, 그 점수에 대한 랭킹을 정한 후, 랭킹이 높은 순으로 다른 사용자들에게 추천하는 방법이다. 대표적인 기존 방법으로는 GroupLens [12], K-Nearest Neighborhood[13], Tapestry[14]등이 있다.

하이브리드 콘텐츠 추천방법은 위의 두 방법을 상호 보완한 것으로 현재 많은 연구가 진행되고 있다[15]. 내용기반 추천 방법과 협업적 추천 방법을 결합하여, 추출된 사용자의 취향과 아이템의 속성 및 아이템이 분류되어진 카테고리를 통해 추천이 이뤄진다.

3. 팟캐스트 검색시스템(PODSSO) 설계

3.1 PODSSO 설계 및 동작

제안된 시스템 PODSSO는 크게 3가지 모듈로 설계되었다. 팟캐스트 정보를 수집하기 위한 팟캐스트 크롤링 모듈, 사용자의 소셜 정보를 추출하기 위한 SNS 데이터 추출모듈, 팟캐스트 검색 결과를 추천해주는 팟캐스트 추천모듈로 구성되어 있다. 각 부분 구성에 대해서는 (그림 1)에서 보여주며, 동작 방법은 다음과 같다.



(그림 1) 팟캐스트 검색 시스템 구조도
(Figure 1) Podcast Search System

사용자가 검색어를 입력했을 때, 제안된 시스템을 통해 결과값을 얻는 과정에 대해 알아보겠다. 팟캐스트 크롤러를 통해 웹에서 해당 데이터를 수집하여 팟캐스트 정보를 저장해놓는다. 사용자가 검색어를 입력하면, 해당 사용자의 SNS정보를 추출하고, SNS 정보를 기반으로 추천 카테고리를 추출한다. 입력된 검색어의 카테고리를 분류한 후, 검색어의 카테고리 and 추천카테고리를 비교하여 동일한 카테고리에 있는지 비교한다. 동일한 카테고리에 있을 경우, 해당 카테고리 내에서 사용자의 친구들이 가장 많이 시청한 팟캐스트를 추천해준다. 동일한 카테고리에 없을 경우는 해당 카테고리 내에서 일반 사용자들이 가장 많이 본 팟캐스트를 추천해준다. 결과적으로 검색하는 팟캐스트 Top 5를 추천해주며, 사용자의 검색어가 포함된 모든 팟캐스트 목록을 보여준다. 위에서 설명한 전체적인 단계를 포함하고 있는 팟캐스트 검색시스템 구성도를 그림1에서 보여준다. 여기에서 주요 알고리즘은 다음 3.2, 3.3 에서 소셜 네트워크 특징 추출방법과 추천알고리즘이 사용되었다.

3.2 소셜정보 추출

사용자들에게 소셜 정보를 검색시스템에 적용하기 위해 본 연구에서는 사용자와 소셜 네트워크상에 관계를 맺고 있는 커뮤니티 관련 소셜 정보(글, 댓글, 친구, 사용자 정보, 글쓰기 링크)를 추출하였다. 추출하는 방식은 SNS사이트에서 제공하는 Open API를 사용하였고, 실제 연구에서 적용한 사이트는 Facebook, Twitter, Me2day이다. 각 사이트에서 소셜 정보를 추출하는 방법은 다음과 같다.

- (1) Facebook: Facebook은 자바스크립트, PHP, 모바일 플랫폼과 언어에 따라 Open API를 제공해주고, 외부 제 3자의 웹페이지까지 연결하기 위해서 Graph API를 플랫폼 컴포넌트에 포함시켰다. 이 Graph API를 통해 Facebook에 공개된 데이터를 마음껏 읽을 수 있고, 이를 활용해 Facebook 플랫폼 상에서 데이터를 쓸수 있다. 이를 활용해 유저들의 정보와 친구들의 목록을 가져왔다.
- (2) Twitter: Twitter는 친구, 타임라인 등의 정보를 제공하는 라이브러리 twitter4j를 통해 유저들의 정보를 얻을 수 있었다. 가져온 정보는 XML 형식으로 보이며, Me2day Parsing할 때 이용했던 XML SAX Parser를 이용해서 HashTable를 이용해서 넣었다. parsing을 할

때, 다른 사이트에서 id를 유저의 정보를 찾아서 그 정보를 저장했다. 하지만, Twitter에서는 screen_name으로 다른 유저들과 구별 할 수 있기 때문에 그 정보를 가져와야 한다는 특징이 있다.

- (3) Me2day: Me2day의 API요청은 me2day에서 제공하는 데이터 및 기능을 접근할 수 있다. me2API의 요청은 HTTP 요청 방식을 사용하고, XML 형식과 JSON 형식으로 받을 수 있고, 반환 받은 결과를 XML SAX Parser를 이용해서 우리가 필요한 유저의 정보와 친구들 목록을 가져와서, 가져온 정보들은 추천 알고리즘에 적용하기 위해 HashTable에 저장한다. 미투데이의 요청 방식은 자유로워서 글, 댓글, 미투, 친구, 사용자 정보, 글쓰기 링크 등과 같은 정보를 가져올 수 있다.

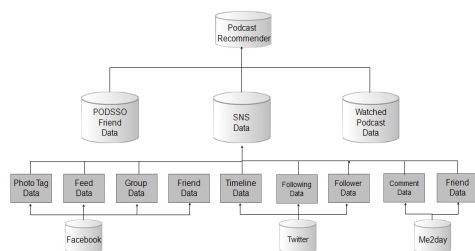
3.3 추천 알고리즘

추천알고리즘은 사용자가 검색을 할 때 사용자의 성향에 따라 팟캐스트를 추천해주는 알고리즘으로, 사용자에게 가장 적합한 팟캐스트를 검색 결과로 추천하는 방법이다. 본 연구에서는 SNS 사이트에서 인물관계, 작성한 글, 관심사를 고려하여, 추천 알고리즘에 적용하였다.

인물관계에 따른 추천 알고리즘은 사용자가 입력한 SNS 아이디를 이용하여 API를 통해 사용자의 친구들 목록을 모두 가져와 이를 이용한 알고리즘이다. 친구들 목록에 기반하여 친구들이 본 적이 있는 팟캐스트를 기록하고 그 횟수를 기록하여 친구들이 가장 많이 본 팟캐스트를 찾아 추천해주게 된다.

SNS는 내가 실제로 알고 지내는 사람, 가족, 연인, 친구들뿐만 아니라 연예인, 유명인사, 심지어 특정회사나 이벤트 등을 통해 수많은 사람들과 관계를 맺을 수 있다. 특히, Twitter의 경우 수락의 개념 없이 팔로우(follow)가 가능하다. 이러한 이유 때문에 단순하게 친구 목록을 가져와서 추천 알고리즘에 적용하게 된다면 그 신뢰성이 다소 떨어질 수밖에 없다. 추천 알고리즘의 신뢰성을 높이기 위해 SNS 상에서 교류가 많은 친구들을 찾고자 하였다.

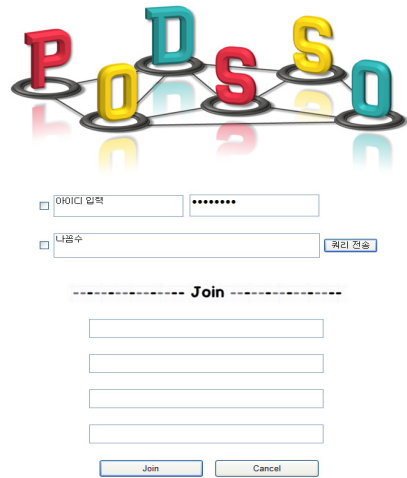
위에서 언급한 바와 같이 우리가 연구에서 사용한 SNS는 Facebook, Twitter, Me2day이다. 각각의 사이트에서 개발자들에게 제공해주는 API들을 분석하여 인간관계를 보다 의미있게 만들어줄 수 있는 정보들을 추천알고리즘에 적용하였다.



(그림 2) 팟캐스트 추천 방법
(Figure 2) Podcast Recommendation method

- ① Facebook: Facebook은 소셜 정보 중 group, photo tags, feed에 대한 정보를 이용하여, 추천알고리즘에 적용하였다. 먼저, group은 사용자가 어떠한 group에 속해 있는지에 대한 정보를 갖고 있다. 한 사용자가 어떠한 그룹에 속해 있다면 그 그룹에 속한 사람들은 보다 더 친밀한 관계가 있다 볼 수 있다. feed는 사용자가 직접 작성한 글 혹은 사용자의 친구들이 작성한 글, 사용자가 태그 되었거나 ‘좋아요’를 누른 글 등 사용자의 담벼락에 있는 모든 글의 목록을 제공한다. 이 정보에서 태그된 사람들의 목록, 댓글을 쓴 사람의 목록, ‘좋아요’를 누른 사람들의 목록을 이용한다. Photo tags는 담벼락에 쓰여진 글들 중에서 어떠한 사진에 Tag된 사람들의 목록에 관한 정보를 제공한다. 사진에 Tag가 되었다는 것은 친한 사이일 확률이 매우 높다고 판단할 수 있다.
- ② Twitter: Twitter는 다른 SNS들 보다 자유롭고, 친한 관계가 아니더라도 자유롭게 140자의 글을 올릴 수 있어 확산성은 매우 높지만 친구관계에 대한 연관 정도를 구하는 것은 상대적으로 어렵다. 따라서 가장 기본적으로 내가 일방적으로 친구를 맺는 Following, 상대방이 나를 일방적으로 친구 맺는 Follower의 관계를 이용하여 서로 Follow한 경우는 상대적으로 친밀도가 높다고 판단하여 가중치(weight)를 적용하게 된다. 또한 사용자의 타임라인(timeline)을 모두 가져올 수 있기 때문에 이를 이용하여 사용자가 작성한 Mention 정보를 통해 계산이 가능하다. Twitter에서 Mention이란 개인 메시지라는 개념으로 사용되기 때문에 Mention을 많이 보낼수록 관계가 높다고 볼 수 있으나 친구 관계가 아니더라도 보낼 수 있기 때문에 가중치를 적용한다. Twitter는 twitter4j와 같은 Java 기반의 오픈 라이브러리가 존재하여 이를 이용하였다.
- ③ Me2day: Me2day는 기본적으로 친구 목록을 가져올

때 어느 정도의 Data mining을 거쳐 상대적으로 친한 친구 순으로 목록을 제공해 준다. 또한 Twitter처럼 한 쪽이 일방적으로 친구를 맺을 수 있다는 특성이 있기 때문에 일방적으로 한쪽만 친구를 맺은 경우 보다는 서로 친구가 맺어져 있는 경우가 보다 친밀도가 높다고 볼 수 있기 때문에 서로 친구일 경우 가중치를 적용하게 된다. Me2day는 사용자가 작성한 댓글을 가져올 수 있어, 댓글 또한 가중치를 이용하여 계산에 이용하기에 적합한 요소가 된다. 하지만 댓글을 작성하는 것은 친하지 않더라도 단순하게 해당 글에 대한 공감을 표하기 위해서도 이용되기 때문에 적어도 5개 이상의 댓글을 작성하였을 때에만 서로 친구일 경우처럼 가중치를 적용한다.



(그림 3) PODSSO에서 쿼리전송
(Figure 3) Query transmission on PODSSO

4. 제안시스템(PODSSO) 구현

제안시스템(PODSSO)은 3가지 모듈로 구성되었으며, 각 모듈은 팟캐스트를 크롤링하는 모듈(팟캐스트 크롤러), 소셜 정보를 추출하는 모듈(SNS 데이터 추출), 팟캐스트 검색결과를 추천해주는 팟캐스트 추천 모듈(팟캐스트 추천)로 구현되었다.

4.1 팟캐스트 크롤러(Podcast Crawler)

팟캐스트 크롤러(Podcast Crawler)는 기존의 웹 크롤러 기반기술로 구현되었다[16]. 팟캐스트는 기본적으로 xml의 형태로 제공되기 때문에 xml 포맷을 따르고 있다. 또한 iTunes에서 제공하는 dtd를 만족시켜야 하기 때문에 모든 팟캐스트는 iTunes의 dtd 주소를 포함하고 있다. 이를 이용하여 사이트를 크롤링하면서 iTunes의 dtd가 포함된 문서를 팟캐스트로 인식하고 해당 팟캐스트의 주소를 데이터베이스에 저장하였다.

웹이라는 환경 자체가 불필요한 데이터가 굉장히 많기 때문에 모든 웹페이지를 다운받아 팟캐스트인지 구별하는 것은 매우 비효율적이다. 따라서 팟캐스트가 될 수 없고, 용량이 큰 미디어파일은 크롤링 대상에서 제외하였다. 또한 웹페이지간에 링크로 연결된 경우 사이클(cycle)이 형성될 수 있기 때문에 이미 크롤링 한 리스트를 관리해야 할 필요가 있다. 이러한 리스트 뿐만 아니라 앞으로 크롤(crawl)해야 할 리스트도 규모가 크기 때문에 메모리차원에서 관리하기에는 한계가 있다. 따라서 sqlite와 같은 가벼운 자바 데이터베이스(Java database)를 이용하여 크롤링 된 페이지에 대한 리스트를 관리하여 메모

리문제를 해결하였다.

4.2 소셜 정보 추출(SNS Data Extraction)

SNS(Social Network Service) 중에서 규모가 가장 큰 Facebook의 정보를 추출했다. 한 사람에 대해서 group, photo tags, feed에 대한 정보를 가져와서 가중치를 구하는 것이 목적이다. 이 정보들은 Facebook의 open API인 Graph를 이용해서 추출했다.

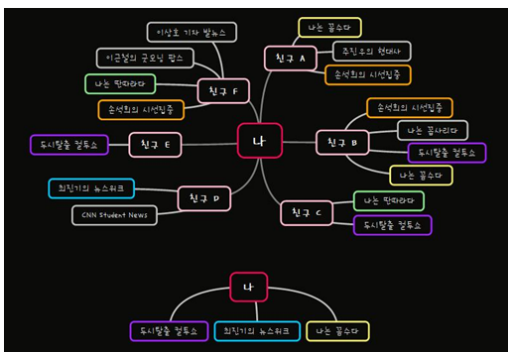
Facebook은 사용자에 대한 정보를 쉽게 가져갈 수 없도록 보안 장치가 되어 있는데 그 중 하나가 액세스 토큰(access token)이다. 한 사용자마다 고유의 액세스 토큰을 발급 받아야만 사용자의 정보들을 가져올 수 있다. 이 권한을 얻기 위해서는 Facebook에 앱(app)을 등록하고, 앱(app)에 대한 고유의 아이디(id)와 비밀번호(secret code)를 발급받아야 한다. 발급 받은 정보를 이용하여 HTTP 요청방식 중 GET방식을 이용하여 정보를 요청하면 각종 권한에 대한 허가를 할 수 있다. 사용자가 허가 버튼을 누르면 해당 사용자에 대한 액세스 토큰이 발급되게 되고, 그 토큰을 이용하여 허가된 정보들을 가져올 수 있다. 이렇게 얻은 토큰을 이용하여 사용자의 정보를 JSON 형태로 가져올 수 있었다. Java 기반으로 정보를 파싱(parsing)하였으며, SNS는 데이터가 굉장히 많기 때문에 순수하게 파싱의 속도가 가장 빠른 방식을 택하였다. 수많은 오픈 라이브러리(open library) 중에서 읽기(read)가 빠른 JACKSON을 이용하여 파싱하였다.

4.3 팟캐스트 추천(Podcast Recommendation)

추출된 SNS 데이터를 이용하여 사용자가 검색을 할 때, 관심도가 높은 팟캐스트를 추천해 주는 것이 주요한 목적이다.

Facebook은 다른 SNS와 다르게 개인의 정보가 많고, 좋아요, Page 구독 등과 같은 기능을 통해 사용자가 흥미를 가지고 있는 카테고리(Category)를 쉽게 알 수 있다. 따라서 이를 이용하여 사용자가 평소 친하게 지내는 사람일수록 흥미가 같을 것이라는 가정 하에 추천 알고리즘을 구현하였다. Facebook에서는 단순히 글을 남기는 것뿐만 아니라 다른 사람에게도 글을 남길 수 있고, 다른 사람의 글에 ‘좋아요’를 누르거나 댓글을 작성함으로써 상대방에게 공감을 표시하기도 한다. 또한 같이 찍은 사진을 올리면서 상대방을 태그하는 기능으로 친분을 나타내기도 한다. 사용자가 친구를 맺고 있는 모든 사용자들을 대상으로 데이터를 처리하면 시간적으로도 비효율적이고 신뢰도 높은 데이터가 나오지 않기 때문에 이를 이용하여 각각 다른 가중치를 적용해서 사용자와 다른 사용자 간에 친분도를 계산하였다.

계산된 친분도를 기준으로 친분도가 높은 사람들이 구독하고 있는 페이지의 정보를 가져온다. 이 계산을 위해서는 10명 또한 우리에게 Facebook 액세스 토큰이 주어져 있어야 한다. 10명이 구독하고 있는 페이지중에서 공통적으로 구독하고 있는 페이지 10개를 선별한다. 이렇게 선정된 페이지의 카테고리를 이용하여 해당 사용자가 관심 있을 것이라고 생각되는 카테고리를 얻어올 수 있고, 해당 카테고리에서 친밀도가 높은 10명이 가장 많이 본 팟캐스트 순으로 추천해 주게 된다.



(그림 4) 소셜 네트워크 친밀도 기반 팟캐스트 추천 (Figure 4) Podcast Recommendation using intimacy information based on social network

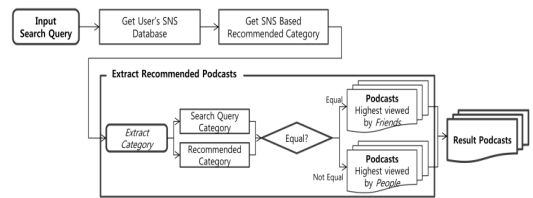
사용자의 친구 목록을 추출하여, 친분에 따른 인간관계도를 PODSSO시스템에서는 다음 (그림 5)와 같이 보여준다.



(그림 5) PODSSO 에서 표현된 인간관계도 (Figure 5) Social Network on PODSSO

본 논문에서 제안한 시스템에서 실제 쿼리를 전송하여 결과를 얻기까지 과정을 다음 (그림 6)에서 보여주었다.

사용자가 검색을 위해 쿼리를 입력하면, 그 사용자의 소셜정보를 기반으로 하여, 해당 카테고리를 찾아낸 후 사용자와 친분이 있는 사람들의 정보와 일반사용자 정보를 비교하여, 가중치를 적용한 후 결과를 보여준다.



(그림 6) 제안된 팟캐스트 추천 과정 (Figure 6) Proposed Podcast Recommendation Flow

5. 실험

우리는 본 논문에서 제안한 방법의 실험을 위하여, 구현된 팟캐스트 회원을 대상으로 데이터를 수집하여, 국내 최초 팟캐스트 포털서비스 사이트인 팟빵(Podbbang) [17]과 비교 분석하였다. 각 사용자의 추천 만족도를 조사하여야 하지만, 정량적 평가를 위해 다음과 같은 방법

으로 실험을 하였다.

(그림 5)의 인간관계를 기반으로 친구목록을 추출한 뒤, 사용자 A의 등록된 글의 ‘좋아요’, 댓글, 태그등을 분석하여 가장 친구 친구들인 top 10 친구를 추출한다. 이렇게 추출한 친구들이 ‘좋아요’를 누른 페이지목록을 가져와서 가장 많이 ‘좋아요’한 페이지의 랭킹 목록을 가져온다. 이렇게 추출된 top 10 페이지의 카테고리과 사용자 A가 ‘좋아요’를 누른 페이지의 카테고리를 대상으로 적중률(Hit ratio)를 계산하였다. 인물 A가 관심있는 카테고리 이름(Ex. [대학교, 가천대], [Actor/Director, 이민호]...)을 추출할 수 있다. 위와 같은 방법으로 적중률을 구하고, 팟빵에서 추천해 주는 팟캐스트의 카테고리를 이용하여 적중률을 구하여 비교하였다. 다음 식1에서 u는 사용자를 의미하여 [Test_u]는 소셜 네트워크의 친구목록 top 10에 있는 사용자이다. [HotCategory]는 사용자u의 카테고리 수이다.

$$Hit_Ratio(u) = \frac{|Test_u \cap HotCategory_u|}{|Test_u|} \quad (식1)$$

적중률은 기준 카테고리는 Facebook에서 대표적으로 사용되는 카테고리를 기준으로 하여 실험하였고, 기준 카테고리는 다음 (표 1)과 같다.

(표 1) 팟캐스트 기준 카테고리
(Table 1) Podcast based Category

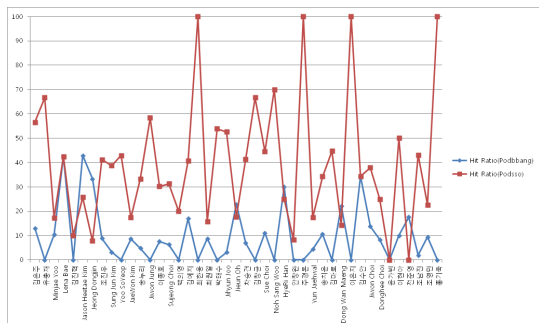
Category (Facebook)	
Media/News/Publishing	Travel/Leisure
TV	School/University
Movie	Sports/Recreation/Activities
Political Organization	Professional Services
Music	Church/Religious Organization
Comedian	Arts/Entertainment/Nightlife
Entertainer	Food/Grocery/ Beverages
Consulting/Business Services	Spas/Beauty/Personal Care
Bank/Financial Institution	Health/Medical/Pharmacy
Shopping/Retail	Biotechnology
Internet/Software	Community/Government/ Government Organization
Games/Toys	Non-Governmental Organization (NGO)
Cars	Baby Goods/Kids Goods

PODSSO사이트에 가입되어 있고, 3개의 소셜 웹사이트(Facebook, Twitter, Me2day)에 가입되어 있는 사용자를 대상으로 실험을 하였다. 결과는 (표 2)와 같다.

(표 2) 적중률 실험 결과
(Table 2) Experiment Result (Accuray)

번호	이름	적중률 (Podbbang)	적중률 (Podssso)
1	김은주	13.04	56.52
2	유총재	0	66.67
3	Minjae Yoo	10.34	17.24
4	Lena Bae	42.5	42.5
5	김진혁	0	10
6	Jason Kim	42.85714	25.71
7	Jeong Dongjin	33.33333	7.84
8	조진우	8.82353	41.17
9	Sung Jun Kim	3.22581	38.71
10	Yoo SoYeop	0	42.85
11	JaeWon Kim	8.69565	17.39
12	송누리	4.76190	33.34
13	Jiwon Jung	0	58.33
14	이종호	7.54717	30.18
15	Sujeong Choi	6.25	31.25
16	백의영	0	20
17	김예지	16.94915	40.67
18	최한웅	0	100
19	최원일	8.77193	15.78
20	박태수	0	53.84
21	Jihyun Joo	3.09278	52.57
22	Jieun Oh	22.78481	17.72
23	차승건	6.89655	41.37
24	김창균	0	66.66
25	Sue Choi	11.11111	44.45
26	Noh Sang Woo	0	70
27	HyeRi Han	30	25
28	안창완	0	8.33
29	주경돈	0	100
30	Yun Jaehwal	4.34783	17.39
31	송치윤	10.52632	34.21
32	김마로	0	44.73
33	Dong Maeng	22.22222	14.28
34	이은지	0	100
35	김수안	34.375	34.375
36	Jiwon Choi	13.79310	37.93
37	Donghee Choi	8.33333	25
38	윤가빈	0	0
39	이현아	10	50
40	전준영	17.64706	0
41	조남진	1.96078	43.13
42	조영민	9.43396	22.64
43	홍기욱	0	100
	평균	9.61908	39.52965

각 사용자에 대해 결과를 그래프로 (그림 6)과 같다. 실험 결과는 팟빵에서의 적중률인 9.6%보다 결과가 좋은 PODSSO의 적중률은 39.256%를 보여주었다. 실험을 통해서 기대한 수준의 검색시스템의 결과를 확인할 수 있었으며, 앞으로 검색시스템을 이용하는 사용자가 확대될수록 적중률은 더욱 향상될 것으로 기대된다.



(그림 6) 사용자별 팟빵과 PODSSO의 적중률 비교
(Figure 6) Comparison Podbbang with PODSSO on Accuracy

6. 결 론

본 연구에서는 소셜 네트워크의 인간관계 요소를 추출하여, 팟캐스트 검색시스템을 구현하여 적용하였다. 현재 일반 검색은 많은 기술 발전이 이뤄졌으나, 소셜 네트워크 요소를 적용하거나 특정 팟캐스트 검색 및 추천 시스템에 대한 연구는 미흡한 상황이다. 본 연구에서 제안한 소셜네트워크 기반 팟캐스트 검색 및 추천 방법은 소셜 네트워크 요소를 다양하게 활용할 수 있는 연구 제안이며, 또한 새로운 소셜 커뮤니티 사이트로 각광을 받고 있는 팟캐스트 사이트에 대한 검색에 활용하면 매우 유용하게 사용될 것이다.

참고문헌(Reference)

[1] G. Lawton, "Knowledge Management: Ready for Prime Time?, IEEE Computer, Vol. 34, No.2, pp 216-244, 2001.

[2] Ok-Ran Jeong and Jehwan Oh, "Social Community Based Blog Search Framework," Springer-Verlag, LNCS 7240, pp. 130-141, 2012.

[3] Incorporating User Search Goal Analysis in Podcast

Retrieval Optimization, 2008

[4] Godwin-Jones, "Emerging Technologies, Skype and Podcasting: Disruptive Technologies for Language," Language Learning and Technology, 2005

[5] Will Ricardson, "Blogs, Wikis, Podcasts, and Other Powerful Web Tools for Classrooms," Thousand Oaks, Corwin Press, 2006

[6] Maged N Kamel Boulos, Inocencio Maramba, and Steve Wheeler, "Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education," BMC Medical Education 2006, 2006

[7] Juan Jose Rodriguez Ponce, RSS and Mobile Multimedia Service, <http://www.tkk.fi/Opinnot/T-10.9.7510/2007/reports/RSS.doc>, 2007

[8] Katarzyna Musial, Przemyslaw Kazienko, 1 and Tomasz Kajdanowicz, "Social Recommendations within the Multimedia Sharing Systems," WSKS 2008, LNAI 5288, pp.364-372, 2008.

[9] Ivan Cantador, Alejandro Bellogin and David Vallet, "Content-based Recommendation in Social Tagging Systems," Fourth ACM conference on Recommender systems, Barcelona, Spain, 2010

[10] L. Rainie, "The state of blogging," Pew Internet & American Life Project, http://www.pewinternet.org/PPF/r/144/report_display.asp. 2005

[11] 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

[12] 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

[13] J. Suchal and P. Navrat, "Full Text Search Engine as Scalable K-Nearest Neighbor Recommendation System," IFIP WCC Series, Vol. 331, pp.165-173, 2010.

[14] D. Goldberg, D. Nichols, B.M. Oki and D. Terry, "Using Collaborative Filtering to Weave a Information Tapestry," Communications of the ACM, Vol.35, pp.61-70, 1992.

- [15] R. Burke, "Hybrid Web Recommender Systems," The Adaptive Web, Lecture Notes In Computer Science, Vol. 4321. Springer-Verlag, Berlin, Heidelberg, pp. 377-408, 2007
- [16] Beomjun Kim, Jongsung Park, Dongcheol Rha, Jeongwoo Lee, Chongjae Yoo, Soyeop Yoo and

Soojeong Choi, "Design and Implementation of Crawler for Podcast Search System," Proceedings of the 39th KISS(Korea Information Science Society) Fall Conference, Vol. 39, No.2(A), pp. 120-122, 2011.

[17] <http://www.podbbang.com/>

● 저 자 소 개 ●



정 옥 란

2005년 이화여자대학교 컴퓨터공학과 (공학박사)
2005년~2006년 서울대학교 컴퓨터공학부 (박사 후 연구원)
2007년 Univ. of Illinois of Urbana Champaign (visiting scholar)
2008년~2009년 성균관대학교 정보통신공학부 (연구교수)
2009년~현재 가천대학교 소프트웨어 설계·경영학과 (조교수)
관심분야 : 웹 마이닝, 정보검색, 추천 시스템, 소셜 컴퓨팅
E-mail : orjeong@gachon.ac.kr