

소셜 네트워크에서의 온톨로지를 이용한 지능형 뮤직 차트의 설계

김도형*, 손종수*, 정인정*
*고려대학교 컴퓨터정보학과
e-mail : {gouske, mis026, chung}@korea.ac.kr

Intelligent Music Chart using Ontology on Social Network

Do-Hyung Kim*, Jong-Soo Shon*, In-Jung Chung*
*Dept. of Computer and Information, Korea University

요 약

최근 전 세계적으로 소셜 네트워크 서비스의 사용자가 많이 증가하면서 많은 사람들이 소셜 네트워크 서비스를 이용하고 있다. 소셜 네트워크 서비스를 사용하는 사용자들은 이를 이용하여 많은 정보를 공유한다. 본 논문에서는 소셜 네트워크 서비스 사용자가 공유하는 정보 중 음악과 관련된 정보와 개방형 API 를 이용하여 MP3 파일의 메타데이터인 ID3 태그 정보를 검색한다. 검색된 결과와 소셜 네트워크 서비스 사용자 정보를 이용하여 ID3 태그 온톨로지를 생성하고 생성된 온톨로지와 온톨로지 추론기를 사용하여 음악과 관련된 다양한 순위 분석 결과와 음악 및 사용자 추천 서비스를 사용자들에게 제공하기 위한 시스템을 설계한다. 본 논문에서 제안한 시스템을 사용하여 제공된 정보는 음악 관련 산업에서 마케팅과 사업 전략 자료 등 다양한 형태로 사용될 것으로 기대한다.

1. 서론

최근 트위터[1]나 페이스북[2] 등과 같은 소셜 네트워크 서비스의 사용자가 증가하면서 사용자들은 해당 소셜 네트워크 서비스를 사용하여 다양한 정보를 공유한다. 특히 트위터를 사용하는 사용자들은 #nowplaying 등의 해쉬태그를 자신의 글에 포함하여 자주 듣는 곡의 아티스트와 제목을 포함한 음악 관련 정보를 공유한다. 본 논문에서는 이러한 음악과 관련된 해쉬태그를 포함한 트위터의 글을 수집하여 글에 포함된 아티스트와 곡 제목 그리고 개방형 API 를 사용하여 곡의 메타데이터를 검색한다. 검색된 정보와 소셜 네트워크 서비스의 사용자 정를 사용하여 ID3 태그 온톨로지를 생성하고 생성된 온톨로지와 온톨로지 추론기를 사용하여 사용자의 다양한 질의에 대한 결과를 보여준다.

본 논문에서 제안된 시스템은 다양한 음악 순위 분석 결과를 토대로 사용자의 취향에 맞는 음악이나 즐겨 듣는 곡이 같은 사용자의 추천과 같은 개인화된 서비스를 제공해줄 뿐만 아니라 음악 산업과 관련된 분야에서 마케팅 및 서비스 제공, 사업 전략을 위한 기반 자료로 활용될 것으로 기대한다.

본 논문의 구성은 다음과 같다. 2 장에서는 소셜 네트워크 서비스의 한 종류인 트위터와 온톨로지, 이전에 발표된 추천 시스템 그리고 온톨로지를 이용한 MP3 메타데이터 관리 시스템에 대해서 살펴보고 3 장에서는 본 논문에서 제안한 시스템의 구성에 대해 설명하고 실험한다. 그리고 4 장에서 결론을 맺고 향

후 과제에 대해 설명한다.

2. 배경 및 관련연구

2.1 온톨로지

2001 년 팀 버너스리(Tim Berners-Lee)가 웹 상의 많은 정보들 중에 사용자가 원하고 사용자에게 유용한 정보를 더욱 빠르고 정확하게 제공하기 위한 시맨틱 웹을 제안하였다[3].

온톨로지는 이러한 시맨틱 웹의 핵심 기술 중의 하나로 특정 분야를 개념화 하기 위해 명시적으로 정형화한 명세[4]이며, 현재의 웹과 같은 이기종 분산 처리 환경에서 자료를 공유하기 위해 자료들 사이의 관계-의미 정보를 명시적이고 정형화된 형태로 표현하여 기계가 이해하고 처리할 수 있도록 하기 위한 지식 표현 방법이다.

온톨로지를 생성하기 위한 마크업 언어들에는 여러 종류가 있지만 본 논문에서는 온톨로지를 생성하기 위해 RDF[5]와 OWL[6]을 사용하였다.

2.2 트위터

소셜 네트워크 서비스는 경계시스템 안에서 공개 혹은 반공개 프로필을 생성하고 그들의 연결을 공유하는 다른 사용자들의 목록을 분명히 표현할 수 있고, 시스템 내의 다른 사용자에게 의해 만들어진 목록과 그들의 목록이 연결되고 보여질 수 있는 개인적인 웹 기반 서비스들이라 정의한다[7].

특히 트위터[1]는 최근 이슈가 많이 되고 있고 사

용자의 수 또한 많이 늘어난 소셜 네트워크 서비스이다. 트위터는 “지저귀다”라는 뜻으로 사용자의 생각이나 현재 상황, 그리고 사용자가 알고 있는 유용한 정보들을 짧은 글로써 전 세계 사용자와 공유한다. 사용자가 트위터를 통해 정보를 공유할 때 트위터의 기능 중 해쉬태그를 사용하면 사용자가 올리는 글의 성격을 다른 사용자들이 쉽게 파악할 수 있다. 해쉬태그는 글을 올릴 때 사용자들이 특정 키워드 앞에 #을 붙여서 올리는 것으로 해쉬태그를 붙여서 글을 쓰면 다른 사용자가 해쉬태그를 클릭했을 때 같은 해쉬태그를 포함한 글들을 쉽게 찾을 수 있다.

본 논문에서는 이러한 트위터의 해쉬태그를 이용하여 음악과 관련된 태그 중 많은 사용자들이 사용하는 #nowplaying 과 이 외에 다른 음악 관련 해쉬태그들을 이용해 트위터에서 공유되고 있는 음악 관련 글들을 수집한다.

2.3 추천 시스템

기존의 추천 시스템은 사용자의 정보를 이용해 음악이나 다른 사용자들을 추천해 주는 기능에 머물러 있었다. FOAF(Friend of a Friend)를 이용하여 작성된 사용자 프로필의 정보를 분석하여 사용자가 관심 있는 아티스트나 음악 장르 정보를 토대로 음악을 추천하고 RSS(RDF Site Summary)를 이용해 아티스트와 관련된 뉴스 등의 정보를 사용자에게 제공해주었다[8].

기존의 추천 시스템들은 사용자가 즐겨 듣는 음악과 비슷한 장르의 음악이나 사용자가 즐겨 듣는 음악과 같은 음악을 즐겨듣는 사용자를 추천하는 것이 시스템 기능의 전부였다.

본 논문에서는 음악 메타데이터들과 소셜 네트워크 서비스에서 제공되는 사용자 정보를 이용해 온톨로지를 생성하고 온톨로지 추론기를 이용하여 국가별, 작곡가별, 장르별 인기곡 순위 등을 포함하는 다양한 순위 분석 결과를 사용자에게 제공한다. 또한 이를 바탕으로 사용자들에게 음악이나 사용자 추천과 같은 개인화된 서비스의 제공하는 시스템을 제안한다.

본 논문에서는 온톨로지 추론을 위해 Jena Framework[10]와 Sparql[11]을 이용한다.

2.4 온톨로지를 이용한 MP3 메타데이터 관리 시스템

이전 연구에서 우리는 온톨로지를 이용한 MP3 메타데이터 관리 시스템을 제안하였다[9]. 온톨로지를 이용한 MP3 메타데이터 관리 시스템은 MP3 파일에 기술되어 있는 MP3 메타데이터인 ID3 Tag 의 정보가 잘못 기술되었거나 불확실할 경우, 그리고 같은 아티스트나 노래에 대해 서로 다르게 기술되어 있을 때 온톨로지와 개방형 API 를 사용하여 MP3 파일의 메타데이터를 일관성 있게 관리해주는 시스템이다.

온톨로지를 이용한 MP3 메타데이터 관리 시스템인 ID3 태그 관리 시스템은 사용자가 가지고 있는 MP3 파일에 대해서 ID3 태그 온톨로지를 생성한다. 온톨로지를 생성할 때 부족한 정보는 개방형 API 를 이용한 웹 검색을 통해 얻은 후 온톨로지에 기술한다.

본 논문에서 제안하는 시스템은 ID3 태그 관리 시스템의 확장된 형태로 사용자가 가지고 있는 MP3 파일뿐만 아니라 트위터의 음악 관련 글에 포함된 아티스트와 곡 제목, 트위터 사용자 정보를 사용하여 온톨로지를 생성한다. 그리고 생성된 온톨로지를 이용하여 MP3 파일의 메타데이터 관리뿐만 아니라 온톨로지 추론기를 사용하여 음악과 관련된 다양한 형태의 분석 자료를 사용자에게 제공한다.

3. 시스템 구성 및 실험

본 논문에서 제안한 시스템의 구성도는 <그림 1>과 같다. 제안한 시스템 구성도를 기반으로 프로토타입 프로그램을 만들어 실험을 수행한다. 시스템을 구성하는 각 모듈의 역할은 다음과 같다.

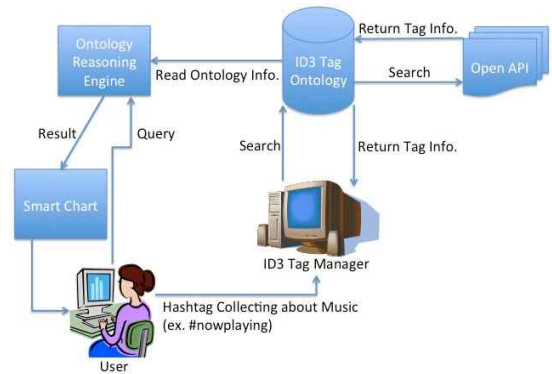


그림 1 시스템 구성도

- ID3 태그 매니저(ID3 Tag Manager) : ① 트위터에 올라오는 음악 관련 글들을 수집하여 아티스트와 곡 제목, 그리고 사용자 정보를 추출한다. 또한 ② MP3 파일의 ID3 태그 정보를 읽어오거나 기술한다. 그리고 ③ 사용자의 질의를 작성하고 그 결과를 확인한다.
- ID3 태그 온톨로지(ID3 Tag Ontology) : MP3 파일의 메타데이터인 ID3 태그 정보를 일관성 있게 관리하며 공유하는 기능을 수행한다. 또한 저장되어 있는 각 곡에 대해서 해당 곡을 트위터에 공유한 사용자 정보와 공유된 횟수를 기술하여 Sparql 을 이용한 사용자 질의에 대한 결과를 추론하는 기반 자료로 사용된다.
- 온톨로지 추론 엔진(Ontology Reasoning Engine) : 시스템을 통해 생성된 온톨로지의 정보를 이용하여 사용자 질의를 처리한 후 결과를 사용자에게 보여준다.

트위터의 음악 관련 글을 이용한 온톨로지 생성과 사용자 질의 처리 과정은 다음과 같다.

첫째, 트위터 사용자들이 음악과 관련된 해쉬태그를 포함한 글을 올리면 ID3 태그 매니저가 주기적으로 해당 글들을 수집하여 글을 올린 사용자의 정보와 글에 포함되어 있는 아티스트, 곡 제목을 추출한다. <그림 2>는 사용자들이 올린 트위터 글들을 수집한

화면이다. 이어서 수집된 글 중 그룹 2PM 의 Only You 라는 곡에 대한 정보를 수집하고 수집된 정보를 이용해 사용자 질의에 대한 결과를 사용자에게 제공하는 과정을 보여준다.



그림 2 수집된 트위터의 음악 관련 글

둘째, ID3 태그 매니저는 추출된 아티스트, 곡 제목 정보를 이용해 ID3 태그 온톨로지를 검색하고 온톨로지에 정보가 있으면 곡에 대한 정보에 글을 올린 사용자 정보를 추가하고 트위터에 공유된 횟수 값을 1 증가시켜 저장한다. 반면에 ID3 태그 온톨로지에 곡에 대한 정보가 없으면 아티스트와 곡 제목을 매개변수로 사용하여 개방형 API 에 질의를 보내고 질의에 대한 결과로 <그림 3>과 같은 XML 형태의 문서를 얻게 된다.

```
<!-- 음악 -->
<item id="203111" seq="0">
  <title>
    <![CDATA[2PM : digital single - Only You (Winter Special) (2008)]]>
  </title>
  <maniad:shorttitle>
    <![CDATA[Only You (Winter Special)]]>
  </maniad:shorttitle>
  <maniad:artist>
    <id>![CDATA[153003]]</id>
    <name>![CDATA[2PM]]</name>
  </maniad:artist>
  <maniad:tracks>
    <disc no="1">
      <title>![CDATA[]]</title>
      <song track="1" id="2203324">
        <title>![CDATA[Only You (Winter Special)]]</title>
        <runningtime>![CDATA[]]</runningtime>
        <performer id="153003">![CDATA[2PM]]</performer>
      </song>
    </maniad:tracks>
  <maniad:products>
    <product>
      <seqno>0</seqno>
      <releasedate>2008-12-10</releasedate>
      <release>![CDATA[트렌 엔터테인먼트]]</release>
    </product>
  </maniad:products>
</!-- 음악 -->
```

그림 3 개방형 API 통해 검색한 음악 정보

이를 이용해 글을 올린 사용자 정보, 공유된 횟수 값과 함께 <그림 4>와 같이 ID3 태그 온톨로지를 생성한다.

```
<!-- 음악 -->
<SN rdf:about="#Only You (Winter Special)">
  <rdfs:label>Only You (Winter Special)</rdfs:label>
</SN>
<AR rdf:about="#2PM">
  <rdfs:label>2PM</rdfs:label>
</AR>
<AA rdf:about="#2PM">
  <rdfs:label>2PM</rdfs:label>
</AA>
<AN rdf:about="#Only You (Winter Special)">
  <rdfs:label>Only You (Winter Special)</rdfs:label>
</AN>
<EN rdf:about="#1">
  <rdfs:label>1</rdfs:label>
</EN>
<RY rdf:about="#2008-12-10">
  <rdfs:label>2008-12-10</rdfs:label>
</RY>
<TN rdf:about="#1">
  <rdfs:label>1</rdfs:label>
</TN>
<!-- 음악 -->
```

그림 4 생성된 음악 정보 온톨로지

셋째, 생성된 온톨로지와 Jena Framework, 그리고 Sparql 을 이용하여 사용자의 질의에 대한 추론을 수행한다. 본 논문에서는 “ 그룹 2PM 의 노래 중 인기 곡 순위와 그 곡들을 만든 작곡가, 그리고 해당 작곡가가 만든 곡들 중 가장 인기있는 곡들 순위” 를 질의로 정하여 실험을 수행한다. <그림 5>는 사용자의 질의를 Sparql 로 작성하여 수행한 결과이다.

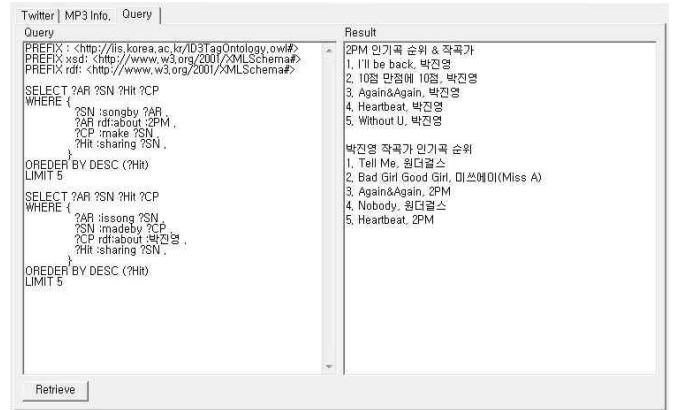


그림 5 사용자 질의를 Sparql 로 작성하여 검색한 결과

질의에 대한 결과로 그룹 2PM 의 인기곡 랭킹 5 위 안의 곡들은 모두 작곡가 박진영이 작곡했다는 것을 알 수 있다. 그리고 작곡가 박진영이 만든 곡 중 가장 인기있는 곡 순위에 대해서도 알 수 있다.

4. 결론 및 향후 연구 과제

최 전 세계적으로 소셜 네트워크 서비스가 이슈가 되고 있고 이를 이용하는 사용자들이 증가함에 따라 소셜 네트워크 서비스에서 공유되는 정보의 종류도 다양해지고 많아졌다. 그러나 기존의 추천 시스템들은 FOAF 나 RSS 등을 이용해 사용자 정보를 기반으로 음악을 추천해주거나 음악의 순위를 평가하는 것이 기능의 대부분이었다.

본 논문에서는 소셜 네트워크 서비스를 이용해 공유되는 다양한 종류의 정보들 중에서 음악 관련 정보들을 추출하여 이를 이용해 곡에 대한 ID3 태그 정보를 검색하고 소셜 네트워크 서비스의 사용자 정보와 함께 온톨로지를 생성한다. 그리고 생성된 온톨로지와 추론기를 이용하여 다양한 형태의 음악 순위 정보와 이를 바탕으로 하는 음악 추천, 사용자 추천 등과 같은 여러 형태의 서비스를 사용자에게 제공해 준다.

온톨로지를 이용한 추론을 통해 국가별 인기 음악 순위, 아티스트별 인기 국가 순위 등과 같은 다양한 음악 순위 분석 결과를 사용자에게 제공하고 이러한 순위 정보를 기반으로 음악 및 사용자 추천 서비스도 효과적으로 제공할 것이라 기대된다. 또한 이러한 자료를 바탕으로 음악 관련 산업 분야에서 마케팅 및 사업 전략 자료 등으로 활용될 것이라 기대된다.

향후 연구 과제로는 질의어 처리 과정에서 일반 사용자도 쉽게 질의를 할 수 있도록 사용자가 Sparql

을 직접 입력하지 않고 자연어로 된 질의를 Sparql 로 변환해 주도록 기능을 확장한다. 그리고 시각화 도구를 사용하여 질의어에 대한 결과를 사용자가 쉽게 알아볼 수 있도록 할 것이다. 또한 다양한 언어로 되어 있는 트위터의 글들을 모두 수집하여 음악 관련 정보를 추출할 수 있도록 기능을 강화한다. 프로토타입이 아닌 전체 시스템을 구축하고 사용자 편의성을 높여 줄 수 있도록 시스템을 확장하는 연구가 필요하다.

참고문헌

- [1] <http://twitter.com> - “트위터”
- [2] <http://www.facebook.com> - “ 페이스북”
- [3] T. Berners-Lee, J. Hendler, and O. Lassila. “The semantic web”. Scientific American, 284(5):34–43, May 2001.
- [4] Gruber, T. R., “A Translation Approach to Portable Ontology Specifications”. Knowledge Acquisition, 5(2):199-220, 1993. See also What is an Ontology? <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>.
- [5] Graham Klyne and Jeremy J. Carroll., “Resource Description Framework (RDF):Concepts and abstract syntax”. W3C Working Draft, 2003. Available at <http://www.w3.org/TR/2003/WD-rdf-concepts-20030123>.
- [6] Mike Dean, Dan Connolly, Frank van Harmelen, James Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein. “OWL web ontology language reference”. W3C Working Draft, 31 March 2003. Available at <http://www.w3.org/TR/2003/WD-owl-ref-20030331>.
- [7] boyd, D. M. & Ellison, N. B.. “Social Network Sites: Definition, History and Scholarship”. Journal of Computer-Mediated Communication, 13(1), article 1. (2007)
- [8] O. Celma, M. Ramirez, and P. Herrera, “Foafing the music: A music recommendation system based on rss feeds and user preferences,” in International Conference on Music Information Retrieval, 2005
- [9] 온톨로지를 이용한 MP3 메타데이터 관리 시스템. 2009 경영정보학 춘계통합학술대회 논문집 pp. 117. 2009. 06. 12.
- [10] <http://jena.sourceforge.net/index.html> - “Jena - A Semantic Web Framework for Java”
- [11] <http://sparql.orf> - “Sparql”