

소셜미디어 서비스를 위한 콘텐츠 서비스 기술

Content service technology for social media services

□ 이종설, 김경원, 장달원, 신사임, 이명준, 윤경로* / 전자부품연구원, 건국대학교*

요약

소셜 미디어는 대중들의 관심있는 생각, 정보, 이슈들을 공유하는 공개 미디어 플랫폼이다. 이를 통해 개인의 관심있는 텍스트, 이미지, 동영상, 음악 등 다양하고 거대한 콘텐츠가 재생산되고 공유되었다. 콘텐츠의 중심이 채널기반 단항 방송 중심에서 사회 다수를 위한 양방향 소셜 환경으로 이동한 것이다. 소셜 미디어의 장점은 관심사나 관계의 내용을 공유함으로써 양방향 의사 소통이 가능하고, 상호 연관성에 따라 새롭게 정의된 의미있는 정보를 기반으로 콘텐츠를 제공하는 소셜 서비스가 가능하다는 점이다. 이에 본 논문에서는 지난 수년간 인터넷과 방송 환경에서 방송과 웹 콘텐츠를 접목시키기 위해 개발된 다양한 기술 및 서비스에 대해 살펴본다. 또한, 기존의 다양한 웹 콘텐츠를 분석하여 소셜 브로드캐스팅 서비스의 운영을 지원하는 자동 지식 기반 구축시스템에 대해 소개한다.

1. 서론

인터넷과 디지털 콘텐츠 기술의 등장과 함께 촉발된 콘텐츠의 혁명은 콘텐츠를 사용자에게 전달하기 위한 채널의 고민에서 벗어나, 어떤 콘텐츠를 어떻게 사용자에게 제공하는가에 대한 콘텐츠 친화성으로 발전하였다. 특히, 소셜미디어 서비스의 등장으로 기존 한정된 콘텐츠 제공자들은 더 이상 생산자로서의 역할이 희미해지고, 일반 사용자들이 다양하고 거대한 콘텐츠의 생산과 배포의 주역이 되는 시대가 되었다. 이와 같은 서비스 대중이 콘텐츠의 생산과 소비를 주도하는 거대 콘텐츠 시장은 광고와 같은 다양한 수익창출 모델을 성공적으로 도출하였고 적재된 콘텐츠만으로도 경쟁력을 보유하게 되었

※ 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2017-0-00318, 건전한 미디어 소비환경 제공을 위한 소셜 IoM 기반 트러스트 미디어 생성·제어 프레임워크 기술 개발)

다. 그러나, 콘텐츠의 양이 늘어감에 따라, 사용자들은 원하는 콘텐츠에 접근하기 위해서 다양한 검색 환경을 거쳐야 했고 이를 위해 사용자 정보 및 콘텐츠의 기술정보를 이용하는 다양한 검색 및 추천 방법들이 필요해졌다. 또한, 미디어 서비스 플랫폼은 소셜 미디어에 대량으로 유입되는 데이터를 분석하고 새로 정의된 의미있는 정보를 기반으로 콘텐츠 서비스를 제공하는 새로운 방법에 대한 필요성이 발생하였다. 효과적인 소셜 미디어 서비스의 구축과 운영을 위해 소셜 미디어 상에 공유되고 있는 집단 지성을 추출하거나, 연관 콘텐츠와 매핑된 양질의 지식베이스의 구축과 운영이 요구되어진 것이다 [1]. 지식베이스 구축을 위해서는 급변하는 대중의 관심을 분석하여 대중적 이슈 및 새로운 이슈에 대한 정보가 반영되어야 한다. 따라서 멀티미디어 서비스에 대한 지식베이스는 대중적인 데이터를 기반으로 주기적이고 빠른 업데이트가 보장되어야 한다. 이에 본 고에서는 기존 디지털방송의 시작부터 진행되어 온 다양한 방송 및 멀티미디어 콘텐츠에 대한 서비스와 기술을 살펴보고, 최근 연구되고 있는 지식기반 콘텐츠 제공 기술에 대한 동향을 살펴본다.

II. 관련 연구

〈그림 1〉에서와 같이 Brian Solis와 JESS3에 의해 2008년 시작된 소셜 미디어 환경의 시각 맵인 Conversatoin Prism을 보면 소셜 미디어를 위한 다양한 환경과 각 산업별 노력들을 살펴볼 수 있다 [2]. 이와 같은 소셜미디어 서비스를 위해 기존 방송 콘텐츠와 웹 환경을 접목시키려는 접근 방법은 크게 2가지 경우로 분류가 된다. 첫째로는 방송 콘



〈그림 1〉 Conversation Prism

텐츠를 중심으로 웹 환경에 접근시키는 방법, 두번째로는 웹 환경에서 방송 콘텐츠를 접목시키는 방법이다. 이에 본 장에서는 소셜미디어 서비스를 위한 기존 연구들을 알아본다.

1. 방송기반 소셜 서비스

아날로그 방송 시절에는 TV 시청자들은 신문 지상을 통해 각 채널별 방송가이드를 접할 수 있었다. 이와 같은 방송 정보 제공 방식은 디지털 방송에 있어서도 전자 프로그램 안내(Electronic Program Guide)를 통해 그대로 계승되었다. 디지털방송이 보편화 되면서 TV와 셋톱박스는 CPU와 메모리를 보유하게 되었고, 이것을 활용하여 사용자의 시청 정보 및 방송가이드를 활용하여 새로운 서비스를 도출하였다. 이를 위한 대표적인 표준이 TV-Anytime Forum이다.

TV-Anytime포럼은 저장 공간을 갖는 디지털 방

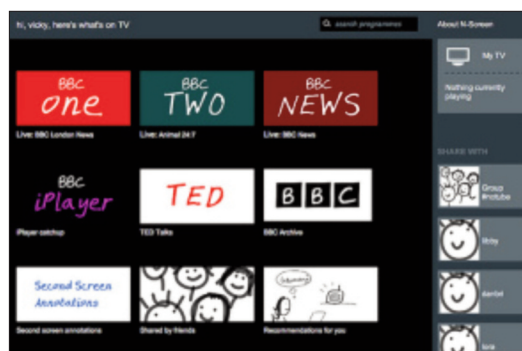
송환경에서 콘텐츠와 부가정보의 저장 및 활용을 통해 사용자가 원하는 방법으로 원하는 시간에 다양한 형태(기존의 방송 서비스 및 on-line 대화형 서비스 등)의 프로그램을 제공받을 수 있게 하는 것을 목표로 한다 [3]. 궁극적으로는 방송과 인터넷이 결합된 통합 서비스 환경에서의 관련 표준을 완성하였고, 유럽전기통신표준기구(ETSI)를 통해 2005년 릴리즈되었다. TV-Anytime 표준은 시스템 표현, 메타데이터, 콘텐츠 참조 기술, 권리관리, 양방향 전송 방법 등에 대한 표준을 확정하였고, 이를 통한 다양한 서비스 모델을 도출하였다. TV-Anytime Forum의 2단계 표준을 살펴보면 사용자 정보와 대용량 메타데이터에 기반한 웹기반 인터페이스를 정의하고 있다. 이를 통해 사용자들은 TV 환경에서 TV 콘텐츠와 관련된 소셜 서비스가 가능해졌다. TV-Anytime 표준은 ATSC, ARIB, DVB 등 다양한 지역 표준에 포함되어 표준화 되었다. TV-Anytime 표준의 가장 큰 성과는 방송 콘텐츠에 대한 메타데이터 표준을 공고히 하였고, 이를 통한 부가서비스 모델을 도출하였다는 점이다. TV-Anytime 이후 시맨틱 웹의 개념과 TV 콘텐츠와의 접목을 시도하는 기술 개발이 등장하였다. 시맨틱 TV는 TV 프로그램과 EPG 정보에 대한 시맨틱 기

술자를 추가하여 웹을 통해 방송 콘텐츠를 위한 리치한 데이터를 제공하는 것을 목표로 한다. 이를 통해 사용자 중심의 개인화 된 콘텐츠와 추천 서비스 제공이 가능하고 나아가 개인별 TV 사용 경험에 대한 공유가 가능하다. 2009년 시작된 NoTube 프로젝트는 시맨틱 기술을 통한 개인화 된 뉴스, 개인화 된 프로그램 가이드와 광고, 소셜 TV와 N-스크린 콘텐츠 추천 등 TV 중심의 소셜미디어 서비스 제공을 목표로 하였다[4]. 이와 같은 방송 콘텐츠 중심의 메타데이터와 사용자 정보, 웹 정보 등이 혼합된 환경에서의 미디어 처리 기술은 FbbTV 등으로 그 대로 연결이 되어졌고, 대용량의 멀티미디어 정보와 사용자 소셜 정보간의 처리 방법에 대한 연구들이 진행되어졌다.

2. 웹 기반 소셜 서비스

방송 콘텐츠와는 달리 웹기반 콘텐츠는 동영상 뿐만 아니라 텍스트, 이미지, 음악, 동영상 등 다양한 콘텐츠가 콘텐츠로써 의미를 갖는다. 대용량의 콘텐츠를 위해 서비스 플랫폼은 사용자들의 편리한 접근을 위해 다양한 방법들이 개발되어졌다.

대부분의 서비스는 기본적으로 인터넷상의 수많

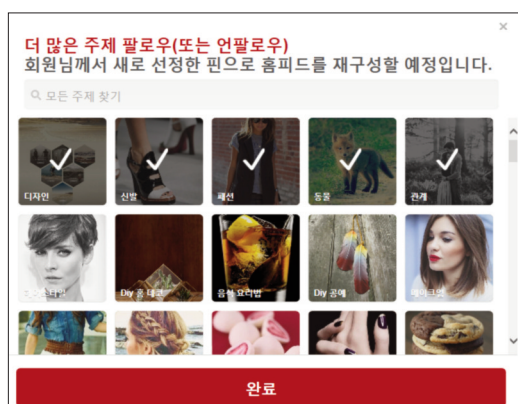


〈그림 2〉 NoTube의 N-스크린

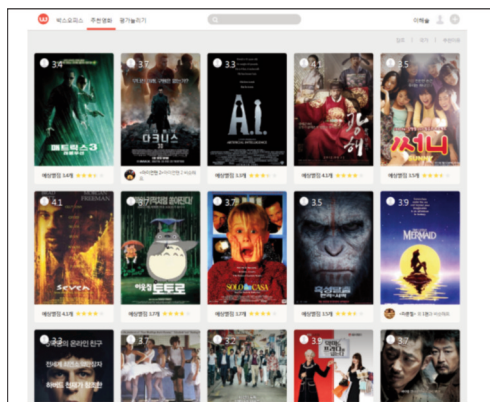


〈그림 3〉 NoTube 포털

은 정보들 중 사용자 개인이 필요로 하고 검증된 콘텐츠들을 골라주는 서비스를 지향하고 있으며 이는 사용자 맞춤형 서비스, 추천 서비스, 소셜 큐레이션 서비스 등으로 불리운다. 이와 같은 기술은 최근 SNS 시장에서 큰 인기를 끌고 있는데, 이는 정보 과잉 시대에 개인별 유용한 정보를 제공할 수 있도록 도와주기 때문이다. 최근 대표적인 웹 기반 소셜 서비스로는 핀터레스트, 왓차 등이 있다. 핀터레스트는 흥미(Interest)있는 것을 핀(Pin)으로 꽂아둔다는 의미를 내포한 서비스명이다 [5]. 페이스북의 ‘타임라인’과 유사하게 핀터레스트에는 ‘핀보드’라



<그림 4> 핀터레스트 서비스



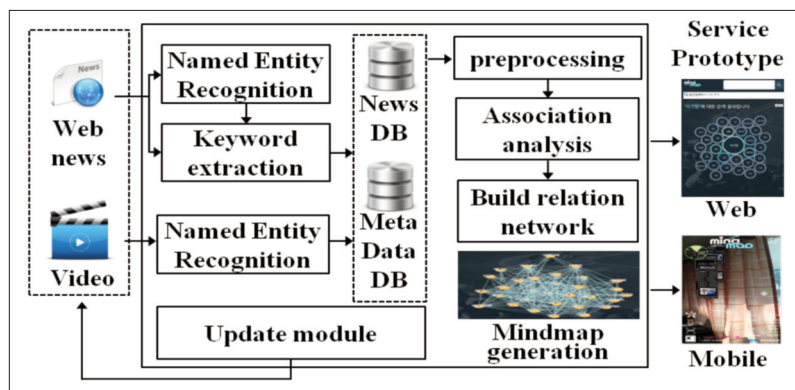
<그림 5> 왓차서비스

는 공간이 있어서 사용자가 자신의 핀보드에 취향에 맞는 이미지를 골라 Pin it(게시)할 수 있다. 핀터레스트는 이미지 중심으로 서비스되기 때문에 기타 SNS보다 직관적이며 트래픽 발생량이 높아 페이스북과 트위터를 잇는 새로운 서비스로 자리매김하고 있다. 왓츠 서비스는 개인의 취향을 분석해 영화에 대한 예상 별점을 제공하고 사용자가 좋아하는 영화를 추천해 준다 [6]. 2013년 1월 이용자들이 별점을 매긴 영화 평가 건수가 500만건으로 480만여건인 네이버 영화를 넘어섰다. 5월 말까지 1,000만건을 넘었으며, 모바일 어플리케이션은 출시한 지 2주만에 1,760만개의 데이터를 축적하였다. 구글코리아와 제휴해 지식그래프 검색 결과에도 영화 평가를 제공하고 있다.

III. 지식베이스 기반 소셜데이터 처리 기술

본 장에서는 전자부품연구원에서 개발중인 콘텐츠 관련 빅데이터를 지식기반 소셜미디어 서비스를 제공하는 마인드맵 기반 기술과 마인드맵을 통해 도출된 미디어 객체와 사용자 정보에 기반한 콘텐츠 생성기술에 대하여 설명한다.

KiS-M 시스템은 소셜 미디어 콘텐츠의 관계정보 생성을 위한 규칙 및 지식베이스 구축을 위해 대용량 데이터베이스를 구축한다. 시스템은 웹 데이터 수집 / 업데이트 모듈; 키워드 / 콘텐츠 마인드 맵 모듈의 세 모듈로 구성된다[7]. “키워드 마인드 맵”은 소셜 데이터에서 추출한 키워드를 바탕으로 대상 도메인에서 의미가 있다고 여겨지는 키워드 관계를 추출하며, “콘텐츠 마인드 맵”은 대상 도메인의 콘텐츠와 키워드 마인드 맵 간의 매핑 정보를 포함하



(그림 6) KiS-M (KETI Intelligent Social-Mindmap) 시스템 구조

다. 시스템은 웹 데이터의 수집을 위해 사회 흐름이 반영된 인터넷 뉴스 기사를 자동 수집한다. 인터넷 뉴스 기사 분석을 통해 특정 분야의 주요 키워드나 연관 키워드를 얻을 수 있다. 따라서 인터넷 뉴스 기사를 수집하기 위한 크롤러를 구현하였다. 본 시스템에서는 다양한 도메인에서의 서비스 제공을 위해 뉴스 기사들 중 국내야구/국내축구/연예/부동산/여행 분야에 관련된 기사를 수집하여 분석에 사용하였으며, 현재 다양한 도메인으로 확장 중이다. 키워드 마인드맵에서는 전처리된 기사를 데이터베이스에 등록하고 특정 기간의 기사를 대상으로 키워드가 언급된 횟수를 측정한다. 이 때, 빈도수가 많은 키워드를 특정 기간의 주요 키워드로 보았다. 또한 주요 키워드가 포함된 기사를 대상으로 기사에 나타난 모든 단어들의 빈도수를 측정했을 때, 빈도수가 높은 단어를 주요 키워드와 관련이 높은 연관 키워드로 보았다. 아래 수식은 앞서 설명한 분석 과정을 수식화하여 나타낸 결과이다. R 은 마인드맵에 포함되어 있고 collocation 기반으로 말뭉치에서 추출한 모든 관계 정보 집합을 의미한다. r_i 은 R 에 속한 각각의 관계 정보를 표현하는데, r_i 은 $\text{exists}(key_m, D)$ 가 성립하는 M 차원의 관계 벡터로 표현된다.

$\text{exists}(key_m, D)$ 는 key_m 과 대상 키워드 D 가 같은 문서에 함께 등장하는 것이 관찰되는 경우는 1, 그렇지 않으면 0을 도출하는 이진 함수로 정의된다. 또한, N 은 마인드맵에서 관찰되는 모든 관계정보의 개수

$$R = \{r_1, r_2, \dots, r_i, \dots, r_N\}$$

$r_i : M - \text{dimensional vector containing exists}(key_m, D)$

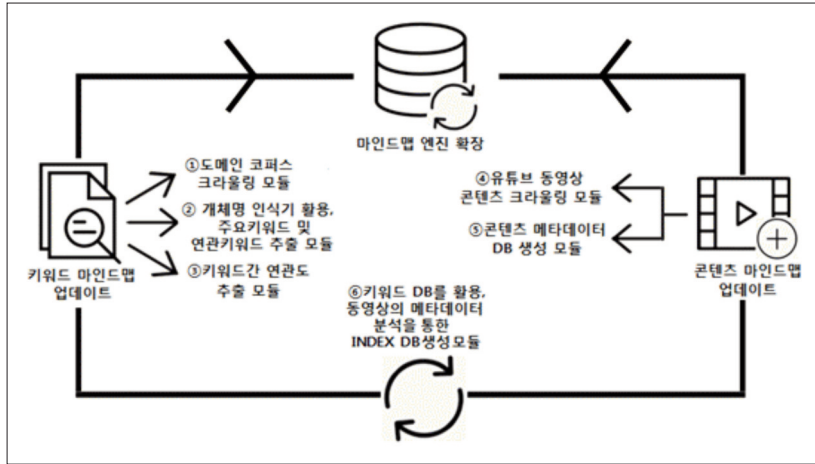
$$\text{exists}(key_m, D) = \begin{cases} 1 & \text{if } key_m \text{ is appeared in } D \\ 0 & \text{else} \end{cases}$$

$N = \text{number of relations}$

로 정의된다.

마인드맵을 통해 도출된 연관관계망을 통하여 추출된 연관 콘텐츠들은 소셜 순위화 엔진을 통하여 연관도의 강도를 반영하여 재순위화된다. 이 과정에서 콘텐츠 마인드맵의 연결 가중치를 반영하고, 마인드맵 상의 연관 키워드가 사용자가 직접 입력한 키워드인지 마인드맵에 의하여 확장된 키워드인지 등을 가중치로 반영하여 순위화한다. 이 과정에서 확장된 질의는 마인드맵 콘텐츠의 메타정보를 벡터화하여 표현한 후, 벡터 유사도 비교 방식을 통하여 수행된다.

마인드맵 엔진의 정확성 및 가용성을 높이기 위



〈그림 7〉 마인드맵 자동 확장 프로세스

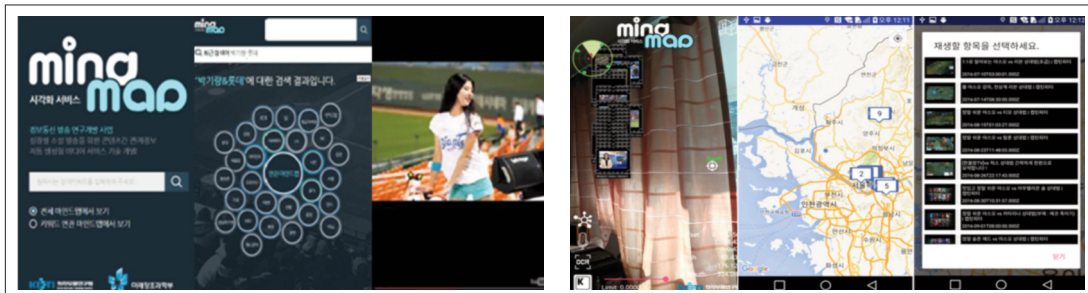
해서는 마인드맵 엔진이 가지고 있는 데이터의 표본의 크기가 클수록 더 정확한 연관도를 바탕으로 콘텐츠를 제공할 수 있다. 따라서 데이터의 자동 추가 구축을 위하여 마인드맵 자동 확장 프로세스를 갖는다. 자동 확장 프로세스에서는 각 도메인별 새롭게 추가되는 데이터를 자동 축적을 수행하고 데이터를 분석하여 추가한다.

자동 확장 과정을 통해 마인드맵 엔진의 데이터는 점차 ‘빅데이터’의 특징이라고 할 수 있는 데이터의 양, 데이터의 생성 속도, 데이터의 다양성을 가지는 마인드맵 엔진으로 발전할 것이다. 이에 따

라 대량의 데이터를 빠르게 처리하고 분석할 수 있는 분산처리 플랫폼으로 확장하였다. 사용자의 질의에 의해 제공되는 콘텐츠는 사용자의 선호 정보 및 사용 기록을 바탕으로 가중치를 적용하며, 이를 통해 각 미디어 객체별 중복성 제거 및 선택을 통해 사용자에게 제공되어진다.

IV. 결 론

본 논문에서는 기존 디지털 방송과 웹환경에서의



〈그림 8〉 KISM 서비스 UI

소셜서비스에 대하여 살펴보았다. 또한 개발중인 콘텐츠의 지식정보에 기반한 콘텐츠 마인드맵 구축 기술 및 시스템에 대해 설명하였다. 시스템은 자동화 프로세스를 통해 새로운 트렌드 키워드에 기반한 콘텐츠 연관정보의 자동 확장이 가능하다. 향후

시스템에서는 콘텐츠의 의미기반 시공간정보 분석 기술에 대한 추가 연구를 진행할 예정이다. 또한, 고정된 도메인 기반이 아닌 자동으로 도메인을 생성 및 관리할 수 있는 시스템으로 확장할 예정이다. [

참고 문헌

- [1] Z. Shi, H. Rui, and A. B. Whinston, "Content Sharing in a Social Broadcasting Environment: Evidence from Twitter," *Information Systems*, Vol 38, No. 1, pp. 123-142, 2014.
- [2] <https://conversationprism.com/>
- [3] <http://www.tv-anytime.org/>
- [4] <https://notube.tv/>
- [5] <https://www.pinterest.co.kr/>
- [6] <https://watcha.net/>
- [7] 이해술, 신사임, 장달원, 이종설, 장세진, "소셜 방송 서비스 지원을 위한 콘텐츠 마인드맵 지식베이스 자동 구축 기술 개발", 한국컴퓨터종합학술대회, 2016.

필자 소개



이종설

- 2001년 : 충북대학교 정보통신공학 석사
- 2013년 : 건국대학교 컴퓨터공학과 박사 수료
- 2001년 ~ 현재 : 전자부품연구원 스마트미디어연구센터
- 주관심분야 : 멀티미디어방송, 스마트미디어 서비스



윤경로

- 1987년 : 연세대학교 컴퓨터전자공학
- 1989년 : 미시건대 전자공학 석사
- 1999년 : 시라큐스대학 컴퓨터 정보과학 박사
- 현재 : 건국대학교 교수
- 주관심분야 : 스마트미디어, 영상처리, 멀티미디어 정보 및 메타데이터 처리

필자 소개



장달원

- 2002년 : KAIST 전자공학과 학사
- 2003년 : KAIST 전자공학과 석사
- 2010년 : KAIST 전자공학과 박사
- 주관심분야 : 멀티미디어 신호처리, 머신러닝



김경원

- 2003년 : 한국외국어대학교 컴퓨터 및 정보통신공학과 컴퓨터공학 전공 석사
- 2004년 ~ 현재 : 전자부품연구원 스마트미디어연구센터
- 주관심분야 : 멀티미디어방송, 클라우드 컴퓨팅, 빅데이터 응용서비스



신사임

- 2002년 : KAIST 전산학과 석사
- 2004년 : KAIST 전산학과 박사 수료
- 2006년 ~ 2016년 : 전자부품연구원 스마트미디어연구센터
- 2017년 ~ 현재 : 전자부품연구원 인공지능연구센터
- 주관심분야 : 인공지능, 자연언어처리, 기계학습, 데이터마이닝, 대화모델



이명춘

- 2014년 : 연세대학교 컴퓨터과학과 석사
- 2014년 ~ 현재 : 전자부품연구원 스마트미디어연구센터
- 주관심분야 : 인공지능, 기계학습, 딥러닝