

대규모 콘텐츠 추천을 지원하기 위한 UCI 메타데이터와 변환서비스의 기능 개선

나문성[†], 이재동^{**}

요 약

콘텐츠 추천 시스템은 콘텐츠에 대한 사용자의 선호도를 예측하고, 예측된 선호도가 높은 콘텐츠를 추천하는 시스템을 말한다. 디지털 식별자는 디지털 네트워크 환경에서 추상적인 작품(Work)이나 디지털 형태로 제작된 콘텐츠 등을 식별하는 역할을 한다. 디지털 식별자는 콘텐츠 추천 시스템에서 주로 이용되는 내용기반 여과 기법과 협업여과 기법에서 효과적으로 활용될 수 있다. 본 논문에서는 UCI 국가표준 디지털 식별자를 대규모 콘텐츠 추천 분야에 효과적으로 활용할 수 있도록 기존 UCI 메타데이터를 확장하고 변환서비스를 개선하는 방안을 제시한다. UCI 메타데이터의 개선은 콘텐츠 추천에 필요한 요약, 키워드, 장르, 연령구분, 평점, 리뷰 항목을 추가하는 것이며, 변환서비스의 개선은 결과페이지에 콘텐츠에 대한 선호도 정보를 입력하는 부분을 포함함으로써 콘텐츠에 대한 선호도 정보를 수집할 수 있도록 하는 것이다. 개선된 UCI를 운용하는 시스템을 설계하고 구현함으로써 본 논문에서 제안한 개선 방안이 콘텐츠 추천에 활용될 수 있음을 보인다.

Improvement of UCI Metadata and Resolution Service for Massive Contents Recommendation

Moon-Sung Na[†], Jae-Dong Lee^{**}

ABSTRACT

Contents Recommender System predicts user's preferences towards contents, and then recommends highly-predicted contents to user. Digital Identifier plays its part in identifying abstract works or digital contents in digital network environment. Digital Identifier could be effectively used in content-based filtering and collaborative filtering that are mainly used in Contents Recommender Systems. Therefore, this paper proposes an improvement of UCI metadata and resolution service for effective use of UCI in massive contents recommender systems. UCI metadata is expanded by adding elements such as abstract, keyword, genre, age, rate and review. Resolution service allows the operation systems to collect user preference for content by including input part of preference in a result page. This paper also designs and implements an improved UCI operation system and shows that the proposed improvement of UCI metadata and resolution service could be used for massive contents recommendation.

Key words: Digital Identifier(디지털 식별자), Recommender System(추천 시스템), UCI Metadata(UCI 메타데이터), Resolution Service(변환 서비스), Content-based Filtering(내용기반 여과), Collaborative Filtering(협업 여과)

※ 교신저자(Corresponding Author) : 나문성, 주소 : 서울시 마포구 상암동길 250-15(121-270), 전화 : 02)3153-1180, FAX : 02)3153-3001, E-mail : nms@kocca.kr
접수일 : 2009년 11월 27일, 수정일 : 2009년 12월 8일
완료일 : 2009년 12월 8일

[†] 정회원, 한국콘텐츠진흥원 제작지원 본부장
^{**} 정회원, 단국대학교 컴퓨터학부 컴퓨터공학전공 교수
(E-mail : letsdoit@dankook.ac.kr)
※ 본 연구는 2009학년도 단국대학교 대학연구비 지원으로 연구되었음.

1. 서 론

콘텐츠 추천 시스템(Recommender System)은 개별 사용자가 선호할 수 있는 영화, 음악, 뉴스 기사, 도서 등과 같은 콘텐츠를 추천해주는 시스템을 말한다. 최근과 같이 다양한 콘텐츠가 생산되고 소비되는 환경에서 콘텐츠 구매 의사를 결정할 때에 검색에 의한 방법보다는 선호도를 기반으로 추천에 의한 구매 의사 결정이 보다 효과적일 수 있다[1]. 추천 시스템에서 사용하는 기법은 사용자의 선호도를 예측하고, 예상 선호도가 높은 콘텐츠를 추천하는 방식에 따라 다양한 방식이 활용되고 있다[2,3].

디지털 식별자는 90년대 중반 디지털 네트워크 환경에서 지적 재산물에 대한 권리보호가 중요한 이유로 부각되면서 개발되기 시작하였다[4]. 해외에서는 DOI(Digital Object Identifier)가 출판 및 학술 분야를 중심으로 국제적으로 활용되고 있으며, 국내에서는 UCI(Universal Content Identifier)가 영화, 음악, 만화 등과 같은 다양한 장르에서 국가표준 디지털 식별자로 활용되는 초기 단계에 있다[5]. 현재의 UCI 식별자는 과거 문화부 COI(Content Object Identifier) 식별자가 제공하는 기능이 통합되어 디지털 형태로 제작된 콘텐츠 뿐만 아니라 작품(Work), 표현(Expression) 또는 실현(Performance) 등과 같은 추상적인 대상도 식별할 수 있도록 확장되었다.

디지털 식별자는 디지털 네트워크 환경에서 유통되는 모든 유형의 콘텐츠 또는 작품을 식별할 수 있기 때문에 저작권자, 제작자, 서비스 사업자, 소비자 모두에게 유익한 유통 기반을 제공해 준다. 디지털 식별자가 ISBN과 같은 일반적인 코드와 가장 크게 차별될 수 있는 것은 변환 서비스(Resolution Service)를 제공한다는 점이다. DOI의 경우 디지털 식별자를 클릭할 경우 대부분 해당 콘텐츠를 구매할 수 있는 장바구니 페이지 또는 구매 의사를 결정할 수 있도록 제목, 저자, 초록 등의 정보가 제공되는 웹 페이지로 이동한다. 이러한 특징으로 인하여 디지털 식별자는 Actionable Identifier[6]라고 불리기도 한다.

하지만, 현재의 UCI 메타데이터는 제목, 유형, 표현형태, 형식, 기여자 등으로 구성되어 있어서 콘텐츠 추천에 활용하기에는 부족함이 있다. UCI 식별체계가 콘텐츠 추천에 효과적으로 활용되기 위해서는 메타데이터의 확장이 필수적이다. 따라서 본 논문에서

는 이러한 UCI 디지털 식별체계가 콘텐츠 추천에 효과적으로 활용될 수 있도록 하기 위해 지원해야 하는 메타데이터 체계, 변환 서비스 기능을 제안한다. 또한, 이를 효과적으로 운용할 수 있는 시스템을 설계하고 구현함으로써 본 논문에서 제안한 개선 방안이 대규모 콘텐츠 추천에 활용될 수 있음을 보인다.

본 논문의 구성은 다음과 같다. 1장 서론에 이어 2장에서는 다양한 콘텐츠 추천 기법과 디지털 식별체계의 핵심 요소를 설명하고, 3장에서는 대규모 콘텐츠 추천을 위한 UCI 식별체계 개선방안을 제시한다. 4장에서는 개선된 UCI 메타데이터를 운용할 수 있는 시스템을 설계하고 구현한다. 마지막으로 5장에서 결론을 도출하고 향후 연구방향에 대하여 언급한다.

2. 관련연구

본 장에서는 콘텐츠 추천에서 주로 사용되는 내용기반 여과와 협업여과에 대하여 설명하고, 대표적인 디지털 식별체제로 DOI(Digital Object Identifier)와 UCI(Universal Content Identifier)에 대하여 알아본다.

2.1 콘텐츠 추천 기법

콘텐츠 추천 시스템은 특정 사용자에게 해당 콘텐츠에 대한 사용자의 선호도(Preference)를 예측하여 예상 선호도가 높은 콘텐츠를 추천한다[7]. 선호도 예측을 기반으로 하는 콘텐츠 추천 방법으로는 Rule-based Filtering, Content-based Filtering, Collaborative Filtering, Item-based Filtering 등과 같이 다양한 방법들이 사용되고 있다[2]. 이러한 방법들 중에서 Content-based Filtering, Collaborative Filtering 그리고 이 둘을 혼합한 Hybrid 방식이 주로 활용되고 있다[3,8-10].

2.1.1 Content-based Filtering 방식

Content-based Filtering 방법은 콘텐츠에 포함된 텍스트 정보 또는 콘텐츠와 관련된 메타데이터 정보를 분석하여 콘텐츠 간의 유사성 및 사용자 선호도를 바탕으로 콘텐츠를 추천하는 방식이다[11]. 전통적으로 정보 검색(Information Retrieval) 기법에 기반하여 활용되어 왔다. 사용자 선호도와 관계없이 콘텐츠 간의 유사성을 기초로 사용자에게 유사 콘텐츠를

제시할 수도 있는 장점을 가진다[12].

특정 사용자에게 최적의 콘텐츠를 제시하기 위해서는 사용자의 취향이나 선호 성향을 파악해야 하는데, 이를 위해 사용자 프로파일(User Profile)을 사용해 왔다. 콘텐츠 자체 또는 메타데이터로부터 분석된 정보와 사용자 프로파일에 등록된 정보와의 관련도 기반으로 사용자 선호도가 높을 것으로 추정되는 콘텐츠를 제시한다.

Content-based Filtering 방법은 음악, 영화, 사진 등과 같이 콘텐츠 자체에 텍스트 정보가 포함되어 있지 못하므로 전적으로 메타데이터에 의존할 수 밖에 없는 한계를 가진다[13]. Content-based Filtering 방법은 콘텐츠 자체에 포함된 정보(대부분 텍스트 정보를 활용) 또는 콘텐츠에 대한 메타데이터에서 사용자 선호도와 관련된 정보를 추출하여 분석하는 것이 핵심이라 할 수 있다. 따라서 디지털 식별체계 메타데이터에 사용자 선호도와 관련된 정보를 추출하고 분석하기 위하여 기존 콘텐츠 추천에서 많이 활용한 항목인 장르, 키워드, 연령구분, 요약 등 표준적인 메타데이터 항목을 추가적으로 보완한다면 Content-based Filtering 방식의 활용성이 크게 달라질 수 있다.

2.1.2 Collaborative Filtering 방식

Collaborative Filtering 방법은 사용자와 유사한 취향을 가지는 다른 사용자들이 선호하는 콘텐츠에 대한 정보를 활용하여 특정 콘텐츠에 대한 선호도를

예측하여 추천하는 방식이다[14]. 사용자들은 특정 콘텐츠에 대하여 평점을 매기고, 특정 콘텐츠에 대하여 유사한 평점을 부여한 사용자 그룹을 분석한다. 사용자에게 대한 콘텐츠 추천은 유사 성향을 가지는 사용자 그룹이 내린 평점에 근거하여 해당 콘텐츠에 대한 선호도를 예측하여 추천이 이루어진다[15]. 따라서 Collaborative Filtering 방법은 다수의 사용자가 평가한 내용을 수집하는 것이 핵심이라 할 수 있다. 디지털 식별자는 평가하는 대상 콘텐츠를 명확하게 식별하며, 변환 서비스 결과 페이지에서 해당 콘텐츠에 대한 정보를 수집하고 분석한 평점 결과 및 리뷰를 제공할 수 있다.

2.2 디지털 식별체계

디지털 식별체계 시스템은 디지털 콘텐츠 식별을 위한 시스템으로 해외에서는 DOI(Digital Object Identifier)가 활용되고 있으며, 국내에서는 UCI(Universal Content Identifier)가 활용되고 있다[5].

2.2.1 DOI 식별체계

DOI (Digital Object Identifier) 식별체계는 디지털 콘텐츠의 저작권정보를 포함하고 있는 고유 식별 체계로 하나의 콘텐츠에 하나의 DOI 부여한다[4]. DOI 식별체계는 콘텐츠내 요소에 대하여 DOI를 발급할 수 있는 특징을 가지며, 디지털 콘텐츠에 고유한 식별번호를 부여하고 이를 URL로 변환하여 인터넷상의 디지털 콘텐츠 위치에 접근할 수 있게 함으로

표 1. DOI RA기관 현황

DOI RA 기관명	소속국가	비 고
Copyright Agency Ltd.	호주	호주의 저작권 관리기관
CrossRef	미국	출판사들이 참여하는 비영리 기관인 PILA(Publishers International Liking Association, Inc.)가 운영하는 국제적인 학술 정보 Reference Linking 서비스 기관
mEDRA	이탈리아	Italy 출판협회가 주축이 된 multilingual European DOI RA로 인터넷 문서간의 영구적인 인용 서비스를 제공하고 있는 기관
Nielsen BookData	영국	도서관 관련 서지 정보를 서비스하는 기관
OPOCE (Office des Publications EU)	EU	EU내의 출판분야를 공식적으로 관장하는 기구
R.R. Bowker	미국	미국 출판물의 ISBN 관리기관
TIB (Technische Informations Bibliothek)	독일	독일의 국립과학기술도서관
Wanfang Data Co., Ltd	중국	중국 내 과학기술 관련 정보 서비스 기관

써 디지털 콘텐츠의 저작권을 보호할 수 있게 한다. 현재 DOI 식별체계는 대부분 학술 또는 출판 관련 작품(Work) 정보에 DOI 식별자가 발급 되어 디지털 콘텐츠의 유통에서 많은 기업들이 활용하고 있다.

DOI 식별체계는 IDF(International DOI Foundation)가 총괄 운영하며 등록관리기관(RA : Registration Agency)이 DOI 등록관리 기능을 담당하고 있다. 주목할 점은 2007년 3월 중국 내에서 과학기술 분야의 학술정보를 서비스하는 기관인 Wanfang Data가 IDF로부터 DOI RA로 새롭게 선정된 점이다. 저작권 분야에서 주목 받지 못하던 중국이 국제 출판계의 권리자들이 주도하는 IDF로부터 인정 받았다는 사실은 중국이 다양한 영역에서 새롭게 변화하고 있다는 사실을 우리에게 보여주고 있다. 중국 회사인 Wanfang Data는 DOI RA 기관이며, 안정된 수익모델 기반을 가지는 CrossRef와 유사한 서비스를 제공할 것으로 보인다. 이로써 DOI 식별체계는 미국, 유럽, 호주 및 아시아 권에서 활용되는 식별체계 자리를 잡게 되었다.

2.2.2 UCI 식별체계

현재의 UCI(Universal Content Identifier) 식별체계는 문화콘텐츠 식별체계 COI(Content Object Identifier)와 통합 운영되는 진정한 의미의 국가표준 식별체계라 할 수 있다. 디지털 식별체계는 식별자를 표현하는 구문구조, 메타데이터 체계, 변환 서비스 그리고 운영 정책 등의 구성요소를 가진다[5]. UCI 식별자가 식별하는 대상은 FRBR 모델에서의 작품(Work), 표현(Expression), 디지털 콘텐츠(Digital Manifestation), 실물 콘텐츠(Physical Manifestation) 그리고 실물 객체(Physical Item) 등이다[1].

작품과 표현은 모두 추상적인 대상이다. 음악 분야를 예로 들면 이영훈 작사 및 작곡의 ‘옛사랑’은

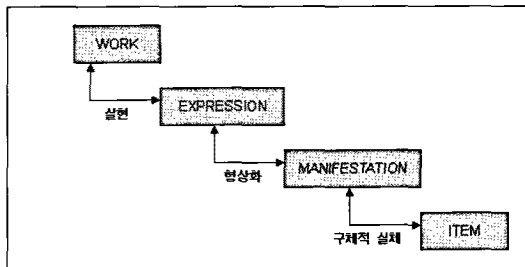


그림 1. IFLA에서 정의한 관리대상 FRBR 모델

작품(Work)이며, 이 곡을 가수 이문세가 부른 것을 표현(Expression)이라 한다. 이문세가 스튜디오 또는 라이브 공연장에서 부른 곡을 녹음한 마스터 레코딩까지 표현(Expression)이라고 할 수 있다. 이를 CD 또는 MP3 파일로 제작했을 때 CD는 실물 콘텐츠(Physical Manifestation) 그리고 MP3 파일은 디지털 콘텐츠(Digital Manifestation)라고 할 수 있다. 마지막으로 동일 곡을 수록한 많은 CD 중 특정 CD 하나를 관리할 필요가 있을 때 이를 실물 객체(Physical Item)로 식별할 수 있다.

UCI는 디지털 식별자가 발급될 경우 식별 대상에 대한 메타데이터가 함께 관리되는데 이를 ‘UCI 응용 메타데이터’라고 한다. UCI 응용 메타데이터 중에는 반드시 포함 시켜야 하는 항목들이 있는데 이를 ‘UCI 식별 메타데이터’라고 한다. UCI 식별자를 발급하고 관리하는 등록관리기관(RA)은 UCI 식별자 발급 시에 ‘UCI 식별 메타데이터’ 항목들은 총괄기구에 전송해야 하는 의무사항을 가진다[16].

현재의 UCI 식별 메타데이터 항목으로는 Content-based Filtering에 활용하기에 한계를 가진다. 영화 콘텐츠에 대한 추천을 위해 User Comments, Plot Outlines, Synopsis, Plot Keywords 등의 문화 콘텐츠에 대한 메타데이터 항목을 정의하여 콘텐츠 추천에 활용하는 연구가 진행된 바 있다[17]. 4장에서는 UCI 메타데이터 확장 방안을 제시한다.

표 2. UCI 식별 메타데이터 항목

항목명		정 의
영문	한글	
UCI	UCI	식별대상에 발급된 디지털 식별자
Identifier	기존식별자	UCI 이외에 발급된 타 식별자
Title	타이틀	알려져 있는 식별 대상의 이름
Type	유형	FRBR 모델에 기초한 식별 대상의 유형
Mode	표현형태	지각적 형태
Format	형식	식별 대상의 형식
Contribution	기여	식별 대상에 기여한 기여자 (Container)
Contributor	기여자	기여자 이름(Contribution의 하위요소)
Role	역할	기여자 역할(Contribution의 하위요소)

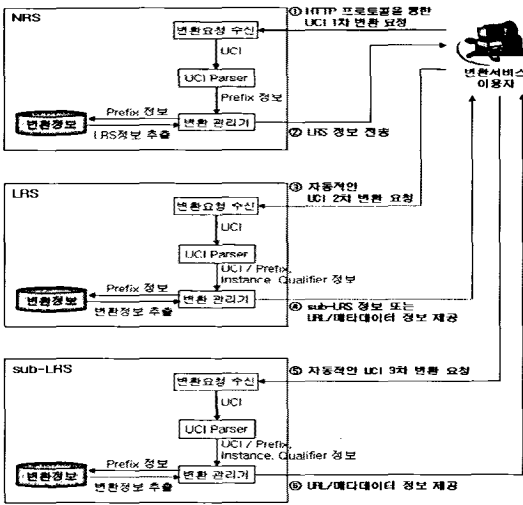


그림 2. UCI 변환 서비스 흐름도

변환 서비스는 유무선 인터넷 환경에서 식별자를 클릭하면 해당 UCI 식별자를 발급한 등록관리기관에서 응답 결과를 제공하는 서비스를 말한다. 일반적으로 해당 콘텐츠에 대한 메타데이터 내용을 보여주거나 콘텐츠를 바로 구매할 수 있는 장바구니 페이지로 이동하는 서비스를 제공한다. UCI 식별자 요청에 대한 변환 서비스의 결과 페이지는 해당 콘텐츠에 대한 선호도 정보를 쉽게 얻을 수 있는 창구로 활용될 수 있다.

3. 대규모 콘텐츠 추천을 위한 UCI 식별체계 개선 방안

UCI 식별체계는 구문구조, 메타데이터 체계, 변환 서비스 그리고 운영 정책 등의 구성요소를 가지는데, 콘텐츠 추천에 활용도를 높이기 위해서는 기존 메타데이터 체계 및 변환 서비스에 대한 개선이 필요하며, UCI 식별자를 이용하여 콘텐츠 추천을 요청하는 구문구조의 정의도 필요하다. 따라서 여기에서는 UCI 국가표준 디지털 식별체계의 메타데이터와 UCI 변환서비스 시스템을 개선하여 대규모 콘텐츠 추천을 제공할 수 있도록 한다.

3.1 대규모 콘텐츠 추천을 위한 UCI 메타데이터 체계 개선 방안

대규모 콘텐츠에 대한 효과적인 추천을 위해서는

추천에 필요한 콘텐츠 메타데이터 항목이 잘 정의되어야 한다[14]. UCI 메타데이터 체계는 UCI 식별자를 발급하고 관리하는 등록관리기관(RA)에서 정의하는 UCI 응용 메타데이터와 UCI 응용 메타데이터 내에 반드시 의무적으로 포함하여야 하는 UCI 식별 메타데이터로 구성되어 있다. 현재의 UCI 식별 메타데이터는 표 2와 같이 UCI, 기존식별자, 타이틀, 유형, 표현 형태, 형식 및 기여자 등과 같은 아주 기본적인 항목만을 정의하고 있다. 이러한 메타데이터 항목은 기존의 추천 연구에서 활용되었던 메타데이터 항목을 포함하지 않고 있기 때문에 UCI가 식별하는 콘텐츠에 대한 추천 정보를 서술하는 데에 명확한 한계를 가진다.

콘텐츠 추천에 필요한 콘텐츠 메타데이터에는 요약, 키워드, 장르, 연령구분, 평점, 리뷰 등이 있다. 요약 및 키워드 항목은 전통적인 Content-based Filtering 기법에서 주로 활용되는 항목으로써 이를 이용하여 콘텐츠 간의 유사성을 측정하였으며[18], 장르 및 연령구분 항목은 영화 및 음악 등의 멀티미디어 콘텐츠 추천에 많이 활용되었다[19,20]. 평점 및 리뷰 항목은 Collaborative Filtering 기법에서 사용자간 또는 콘텐츠간의 유사도 및 예측값을 생성하기 위하여 주로 활용되었다[2,8].

콘텐츠 추천에 UCI 식별체계가 효과적으로 활용되기 위해서는 메타데이터 확장이 필수적인데, UCI 등록관리기관에서 임의로 정의하는 응용 메타데이터를 확장하는 방안은 바람직하지 않다. 이는 향후 등록관리기관 간의 상호 호환성에 문제가 발생하는 주요 원인이 되기 때문이다. 따라서 모든 UCI 등록관리기관에서 의무적으로 준수하는 UCI 식별메타데이터를 개선하는 것이 바람직하다. UCI 식별메타데이터 항목인 식별자(UCI), 기존식별자(Identifier), 제목(Title), 유형(Type), 표현형태(Mode), 형식(Format), 기여자(Contributor)로는 대규모 콘텐츠 추천 및 콘텐츠를 활용하고자 하는 사용자에게 만족할 만한 정보를 제공하여 주지 못하기 때문에 UCI 메타데이터를 활용한 콘텐츠 온라인 유통 활성화에 저해 요인이기도 하므로, UCI식별체계의 질적 성장을 위한 메타데이터 항목으로의 개선이 필요하다. UCI 식별메타데이터는 다양한 장르의 콘텐츠에 대하여 UCI 식별자를 발급하므로 범용적이면서 필수항목으로서의 가치가 있는 항목의 선택이 중요하며 너무 많은 항목

의 추가는 저해 요인으로 작용하므로 기존의 Content-based Filtering과 Collaborative Filtering 기법에서 중요하게 활용한 항목을 적용하여 개선하고자 한다. 다음 메타데이터 항목은 Content-based Filtering, Collaborative Filtering 기법을 활용하기 위해 UCI 메타데이터 포함되어야 할 항목들이다.

- 제목(Title): 콘텐츠 제목이며, 현재 UCI 식별 메타데이터에 정의되어 있다.
- 장르(Genre): 콘텐츠가 속한 장르정보이며, 대중소 등과 같은 세부적인 장르 분류가 필요하다.
- 연령 등급(Age Rate): 연령제한 없음, 아동물, 청소년물, 성인물 등과 같은 연령 등급정보를 의미한다.
- 요약(Summary): 콘텐츠에 대한 요약정보로 Abstract, Synopsis 등과 같은 정보를 의미한다.
- 키워드(Keyword): 콘텐츠와 관련된 키워드를 의미한다.
- 기여자(Contributor): 저자, 작사, 작곡, 가창, 감독, 출연자, 스태프 등과 같이 콘텐츠에 기여한 다양한 인물을 의미한다. 기여자는 현재 UCI 식별 메타데이터에 정의되어 있다.
- 사용자 리뷰(User Review): 콘텐츠에 대한 사용자들의 평가를 관리하는 항목이다.

이상의 메타데이터 항목들은 콘텐츠를 표현하고, 평가하고, 추천하기 위한 기본적인 정보라 할 수 있다. 제목과 기여자 항목은 현재 UCI 식별 메타데이터 항목으로 정의되어 있기 때문에 나머지 항목들의 UCI 식별 메타데이터로의 확장이 필요하다. 표 3은 기존 UCI 식별체계 메타데이터를 확장한 항목들을 나타낸다.

표 3. UCI 식별 메타데이터 확장 항목

항 목 명		기존 UCI 식별체계 지원 유무
영 문	한 글	
Title	제목	O
Genre	장르	X
Age Rate	연령구분	X
Summary	요약	X
Keyword	키워드	X
Contributor	기여자	O
User Review	사용자 리뷰	X

3.2 UCI 변환 서비스 개선 방안

Collaborative Filtering 방식은 다수의 사용자 평가 정보를 수집하는 것이 중요한데, UCI 변환 서비스를 활용하면 이러한 정보를 더욱 효과적으로 수집할 수 있다. UCI 식별자가 식별하는 대상에 대한 변환 서비스 결과 페이지에 해당 콘텐츠에 대한 선호도 정보를 입력하는 부분을 포함하면, UCI 변환 서비스가 일어날 때마다 콘텐츠에 대한 선호도 정보를 수집할 수가 있다. 한편, UCI 변환 서비스 요청 시에 해당 콘텐츠와 관련된 콘텐츠의 추천을 명시적으로 요청할 수 있도록 다음과 같이 확장할 필요가 있다.

$$f(uci) = \text{list of recommendations without user profile} \quad (1)$$

$$g(uci, user) = \text{list of recommendations with user profile} \quad (2)$$

(1)의 경우는 사용자 프로파일 정보와 관계없이 UCI가 식별하는 콘텐츠와 관련된 콘텐츠의 추천을 요청하는 경우이며, (2)의 경우는 사용자 정보를 제공하여 사용자 프로파일을 활용하여 콘텐츠의 추천을 요청하는 경우를 의미한다. 다음은 변환 서비스 요청 시에 사용되는 실례를 보인 것이다.

(1) I100-1234::CR

(2) I100-1234::CR:<User>

위에서 'I100-1234'는 UCI 식별자를 의미하며, '::'은 뒤에 변환 서비스 결과에 콘텐츠 추천 요청이 있음을 의미하는 구분자를 의미한다. 'CR'은 Contents Recommendation의 약자로 추천 콘텐츠 정보를 변환 서비스 결과로 표시해 달라는 것을 의미한다. (2)에서는 사용자 정보가 추가로 전달되는 형태를 표시한 것으로 ':'는 뒤에 사용자 정보가 온다는 것을 의미하는 구분자의 의미를 가진다.

이상의 콘텐츠 추천 요청은 변환 서비스 외에 API를 통해 지원될 수도 있다. 현재 정의되어 있는 UCI API는 등록관리기관에서 사용하는 메타데이터 등록 및 UCI 발급을 위한 API와 등록관리기관에서 등록된 UCI 응용 메타데이터 중에서 UCI 식별 메타데이터를 총괄기관에 전송하는 API가 정의되어 있는데, 콘텐츠 추천을 요청하는 API를 추가로 제공하여 확장할 필요가 있다.

4. 개선된 UCI 메타데이터를 이용한 운용시스템 설계 및 구현

본 장에서는 3장에서 제안한 개선된 UCI 메타데이터를 운용할 수 있는 시스템을 설계하고 구현한다.

4.1 개선된 UCI 메타데이터를 이용한 운용시스템 설계

UCI 운용시스템은 등록관리시스템인 RA 시스템(Registration Agency)과 총괄기구시스템(Registration Authority System)인 ROOT 시스템으로 구성된다. 그림 3은 이러한 UCI 운용시스템 구성도이다. 등록관리 시스템은 기관, 단체 또는 개인인 등록자(Registrant)로부터 자원을 등록 받고 총괄기구가 제공하는 UCI API를 이용하여 메타데이터를 전송하며, UCI 변환 요청에 대하여 LRS(Location Resolution Server) 2차 변환 서비스를 제공한다. 총괄기구시스템(Registration Authority, Root System)은 식별체계 운영에 대한 전반적인 권한과 보급 확산의 책임을 가진 최상위 기구이다. 등록관리 기관에서 발급한 UCI 메타데이터를 수집하여 영구 보존하며 UCI 식별자에 대응하는 결과를 제공하기 위하여 NAS(Name Resolution Server) 1차 변환 서비스를 제공

한다.

그림 4는 등록관리시스템(RA System)과 총괄기구시스템(Root System)의 소프트웨어 구성도이다. 등록관리시스템은 메타데이터관리, 1차변환서비스, 검색, 통계 및 웹서비스 기능을 제공하며, 총괄기구시스템은 2차변환서비스 및 등록관리시스템과의 연동기능을 제공한다. 변환서비스에 Content-based Filtering 방식과 Collaborative Filtering 방식을 적용하기 위해서는 각 장르에 따라 변환서비스를 제공하는 RA System에 UCI 메타데이터 확장 항목을 적용하여 UCI 변환서비스 결과페이지를 디자인하여야 한다. 또한 UCL API 기능 중에는 Root System에 기존메타데이터 항목을 등록(Metadata Register)하는 기능을 포함한다. 따라서 메타데이터 확장항목을 등록하기 위해서 UCL API를 확장한다.

4.2 확장된 UCI 메타데이터를 이용한 운용시스템 구현

확장된 UCI 메타데이터를 이용한 운용시스템을 구현하기 위하여 한국콘텐츠진흥원(KOCCA)의 UCI 시스템을 활용하였다. 그림 5는 UCI 시스템의 하드웨어 구성도를 나타낸다. UCI 시스템은 보안을 위해 IDC의 방화벽 안쪽에 위치하고 있으며, 총괄관리를 위한 시스템들은 안정적인 속도의 서비스를 제공하

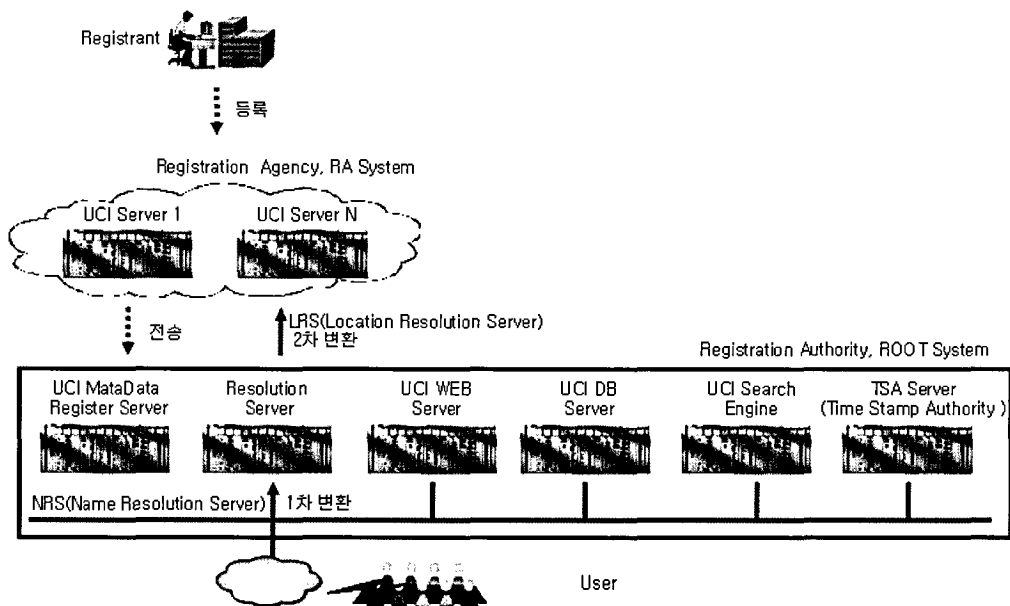
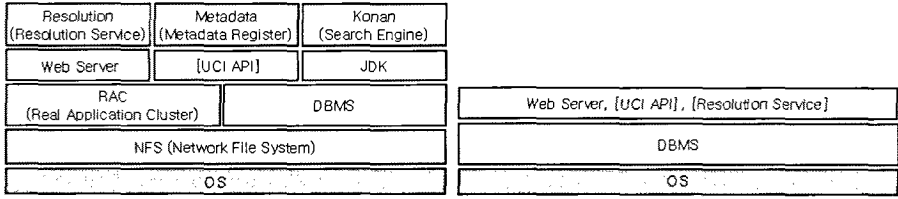


그림 3. UCI 운용시스템 구성도



(a) Root System

(b) RA System

그림 4. UCI 운용시스템 소프트웨어 구성도

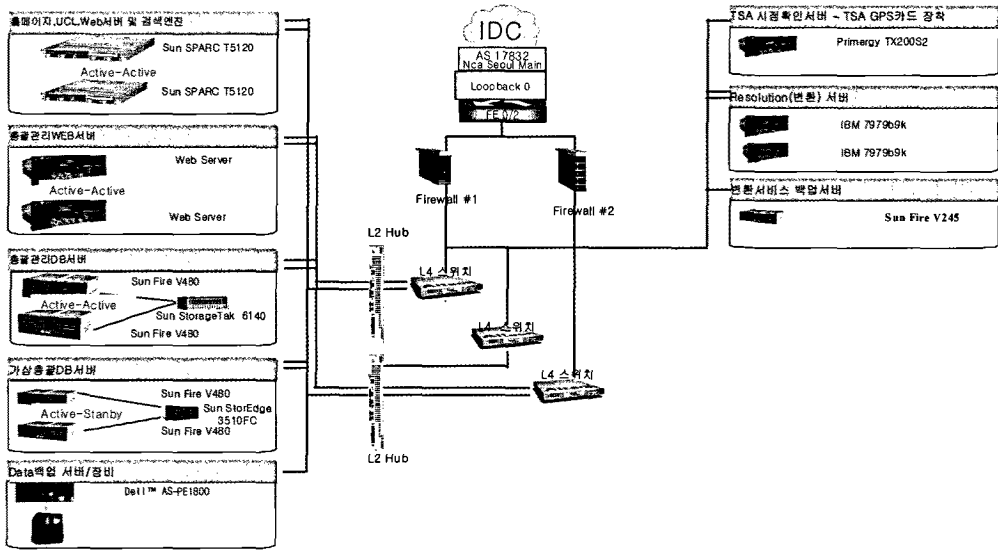


그림 5. UCI 운용시스템 하드웨어 구성도

기 위하여 L4 스위치에 연결되어 있다.

표 4는 총괄기구시스템과 등록관리시스템의 소프트웨어 구성 환경을 나타낸다. 총괄기구시스템과 등록관리시스템 모두 플랫폼에 독립적인 JVM을 기반으로 Servlet 및 JSP를 이용하여 구현하였다. 총괄기구시스템은 Unix를 운영체제로 활용하였으며, 등록관리시스템은 Unix, Linux, Windows Server 등 다양한 운영체제를 활용하여 구축하였다.

4.3 개선된 UCI 메타데이터를 이용한 운용시스템 활용 사례

여기에서는 앞 절에서 구현한 운용시스템을 이용하여 Content-based Filtering 방식에 UCI 식별체계가 효과적으로 활용되기 위한 메타데이터 개선 사례를 보인다. 먼저 표 5는 기존 UCI 메타데이터의 예이다. UCI 식별체제는 콘텐츠를 식별할 수 있는 '1100-1234' 식별번호와 '45321'이라는 기존식별자로

표 4. UCI 운용시스템 소프트웨어 구성 환경

구 분	총괄기구시스템	등록관리시스템
OS	Unix	Unix, Linux, Windows 2000, Windows 2003 Server
JVM	JVM 1.2 이상	JVM 1.2 이상
Web Server	Apache 2.0 이상	Apache 2.0 이상, IIS 5.0 이상
WAS	Servlet 2.3 이상, JSP 1.2 이상	Servlet 2.3 이상, JSP 1.2 이상
DBMS	Oracle	Oracle, Informix, Sybase, SQL-Server

표 5. UCI 기존 메타데이터 항목의 예

번호	항목명	내 용
1	UCI	I100-1234
2	Identifier	45321
3	Title	해운대(Haeundae, 2009)
4	Type	Digital
5	Mode	Visual & Audio
6	Format	AVI
7	Contributor	감독 윤제균, 배우 설경구, 배우 하지원, 배우 박중훈, 배우 엄정화

구성되어 있으며, 기존식별자는 관리측면에 따라 콘텐츠를 등록하는 등록자가 관리하는 관리번호 또는 다른 식별체계 DOI, KOI, ISNB에서 발급한 식별자 등이 될 수 있다. 콘텐츠 제목을 관리하는 Title 항목과 콘텐츠 디지털 정보의 유형을 파악할 수 있는 'Digital'과 콘텐츠의 표현형태인 'Visual & Audio', 파일형식인 'AVI', 콘텐츠의 권리정보를 제공하는 Contributor 항목으로 역할인 '감독', '배우'와 기여자인 '윤제균', '설경구', '하지원', '박중훈', '엄정화'로 구성되어 서비스되고 있다.

위와 같이 기존의 UCI 식별체계는 콘텐츠의 기본적인 항목만을 정의하고 있기 때문에 추천을 위해서는 많은 한계를 가진다. 콘텐츠를 서비스 받는 사용자 관점에서 '해운대(Haeundae, 2009)' 라는 제목 항목만으로는 해당 콘텐츠의 정보를 이해하기 어려우며, 전문서적 및 논문 분야라면 더욱 한계를 드러낼 수 있다. 이는 Content-based Filtering 방식인 콘텐츠

추천에 포함된 텍스트 정보 또는 콘텐츠와 관련된 메타데이터 정보를 분석하여 콘텐츠 간의 유사성 및 사용자 선호도를 바탕으로 콘텐츠 정보 검색을 통하여 콘텐츠를 제시할 수 없는 이유이기도 하다.

표 6은 본 논문에서 제안한 UCI메타데이터 확장 항목의 예를 보인다. 기존 메타데이터 항목을 보완하고 Content-base Filtering 방식을 적용하기 위한 확장항목으로 Genre, Age Rate, Summary, Keyword 등이 있다. 콘텐츠의 장르 항목으로 '모험, 드라마', 연령등급인 '12세', 콘텐츠의 요약정보를 제공하는 Summary, 정보검색에 사용되는 Keyword 항목을 이용하여 추천이 가능하다. User Review항목은 Collaborative Filtering 방식을 적용하기 위한 것으로 다수의 사용자가 평가한 내용 및 평점을 수집하고 변환 서비스 결과로 제공하기 위한 항목이다.

4.4 확장된 UCI 메타데이터를 이용한 운용시스템 결과

UCI 기존 변환 서비스 사례와4.1절의 UCI 메타데이터를 이용한 운용시스템 구현에서 설명한 Content-based Filtering방식과 Collaborative Filtering방식을 적용한 메타데이터 확장에 대한 UCI 변환 서비스 결과 페이지이다.

• 기존 UCI 변환서비스 사례

그림 6은 기존UCI 변환서비스 사례로 UCI 식별체계에서 정의한 필수 메타데이터 항목인 UCI, 제목, 기존식별자, 유형, 표현 형태, 형식, 기여자 정보 등을 제공하며 RA 등록관리 기관에서 정의한 응용메타데

표 6. UCI 메타데이터 확장 항목의 예

번호	항목명	내 용
8	Genre	모험, 드라마
9	Age Rate	12세
10	Summary	2004년 역사상 유례없는 최대의 사상자를 내며 전세계에 엄청난 충격을 안겨준 인도네시아 쓰나미. 당시 인도양에 원양어선을 타고 나갔던 해운대 토박이 만식은 예기치 못한 쓰나미에 휩쓸리게 되고, 단 한 순간의 실수로 그가 믿고 의지했던 연희 아버지를 잃고 만다. 이 사고 때문에 그는 연희를 좋아하면서도 자신의 마음을 숨길 수 밖에 없다. 그러던 어느 날, 만식은 오랫동안 가슴 속에 담아두었던 자신의 마음을 전하기로 결심하고 연희를 위해 멋진 프로포즈를 준비한다.
11	Keyword	해운대, 윤제균, 재난영화, 쓰나미
12	User Review	평점 : 9점 1) 마음 한구석이 찡한 감동이 있는 한국 대표영화.[9점] 2) 설경구 땀에 흥행한다.[9점] 3) 웃음과 감동이 있는 영화. 정말 재미있게 봤습니다.[9점]

제목	해운대(Haeundae, 2009)		
UCI	1100-1234	서비스횟수	1000
기존식별자	45321		
유형	Digital	표현 형태	AudioVisual 형식 AV
기여자	감독 윤재균, 배우 송경구, 배우 하지원, 배우 박종호, 배우 임정화		

그림 6. 기존 UCI 변환서비스 사례(예:1100-1234)

제목	해운대(Haeundae, 2009)										
UCI	1100-1234	서비스횟수	1000								
기존식별자	45321										
유형	Digital	표현 형태	AudioVisual 형식 AV								
기여자	감독 윤재균, 배우 송경구, 배우 하지원, 배우 박종호, 배우 임정화										
장르	모험, 드라마	연령등급	12세								
요약	2004년 역사상 유례없는 최대의 사상자를 내며 전세계에 엄청난 충격을 안겨준 인도네시아 쓰나미. 당시 인도양에 횡랑어선을 타고 나갔던 해운대 투박이 만석은 예기치 못한 쓰나미에 휩쓸리게 되고, 단 한 순간의 실수로 그가 알고 의지했던 연인 아버지를 잃고 만다. 이 사고 때문에 그는 연회를 좋아하면서도 자신의 마음을 숨길 수 밖에 없다. 그러던 어느 날, 만석은 오랫동안 가슴 속에 담아두었던 자신의 마음을 전하기로 결심하고 연회를 위해 멋진 프로포즈를 준비한다.										
키워드	해운대, 윤재균, 재난영화, 쓰나미										
사용자 리뷰	<table border="1"> <tr> <td>평점주기</td> <td>★★★★☆</td> </tr> <tr> <td>리뷰</td> <td>등록</td> </tr> </table> <p>사용자 리뷰 리스트</p> <table border="1"> <tr> <td>평점</td> <td>9.0</td> </tr> <tr> <td>리뷰</td> <td>1).마음 한구석이 찢힌 감동이 있는 한국 대표영화 (90점) 2).송경구 명에 흠뻑한다 (90점) 3).웃음과 감동이 있는 영화. 정말 재미있게 봤습니다. (90점)</td> </tr> </table>			평점주기	★★★★☆	리뷰	등록	평점	9.0	리뷰	1).마음 한구석이 찢힌 감동이 있는 한국 대표영화 (90점) 2).송경구 명에 흠뻑한다 (90점) 3).웃음과 감동이 있는 영화. 정말 재미있게 봤습니다. (90점)
평점주기	★★★★☆										
리뷰	등록										
평점	9.0										
리뷰	1).마음 한구석이 찢힌 감동이 있는 한국 대표영화 (90점) 2).송경구 명에 흠뻑한다 (90점) 3).웃음과 감동이 있는 영화. 정말 재미있게 봤습니다. (90점)										

그림 7. UCI 변환서비스 개선 사례(예: 1100-1234::CR 또는 1100-1234::CR:(User))

이터 정보를 추가로 제공하나 코어 메타데이터 항목만을 표시하였다.

• 확장된UCI 변환서비스 개선 사례

그림 7은 변환서비스 개선사례로 기존 UCI 변환서비스 항목인 제목, UCI, 기존식별자, 유형, 표현형태, 형식, 기여자 정보와 Content-based Filtering 확장항목인 장르, 연령등급, 요약, 키워드와 Collaborative Filtering 확장항목인 사용자 리뷰 및 사용자 리뷰리스트 정보를 제공하는 개선 사례이다. 사용자 리뷰에서는 평점주기 및 리뷰 쓰기로 다수의 사용자가 평가한 내용을 수집하고, 사용자 리뷰 리스트에서는 수집정보에 대한 평점 및 사용자 리뷰를 서비스 결과로 제공한다.

이는 변환서비스를 사용하는 많은 사용자들로부터 수집한 정보와 각 콘텐츠의 관심사에 대한 평균적인 평가를 통하여 고객들의 선호도와 관심 표현을 바탕으로 하여 선호도, 관심에서 비슷한 패턴을 가진 사용자들을 식별해 비슷한 취향을 가진 사용자들에게 서로 콘텐츠들을 교차 추천하거나 분류된 사용자의 취향이나 생활 형태에 따라 관련 콘텐츠를 추천하는 형태의 서비스를 제공할 수 있게 된다.

5. 결론 및 향후 연구방향

UCI 식별체계는 디지털 네트워크 환경에서 콘텐츠를 식별하기 위한 국가표준 식별체계의 기능을 가지며, 구문구조, 메타데이터 체계, 변환 서비스 그리고 운영정책 등으로 구성되어 있다. 본 논문에서는 Content-based Filtering과 Collaborative Filtering 등의 콘텐츠 추천 기법에 UCI 식별체계가 효과적으로 활용될 수 있도록 기존 UCI 식별 메타데이터 확장 방안과 UCI 변환 서비스의 확장 및 변환 서비스 요청 시 콘텐츠 추천을 명시하여 요청하기 위한 구문구조 확장 방안을 제시하였다. UCI 식별 메타데이터의 확장은 Content-based Filtering을 통한 추천 및 Collaborative Filtering 방법에서 대상 콘텐츠에 대한 추천정보를 서술을 위해 필요하였고, 변환 서비스의 확장은 Collaborative Filtering 방법을 활용하기 위한 사용자 평가 정보의 수집과 콘텐츠 추천을 받을 수 있는 환경을 제공할 수 있도록 하였다. 또한, 현재 UCI 총괄기구에 수집된 전체 등록관리기관 UCI 식별 메타데이터는 항목의 특성 상 내부 관리용으로 활용이 제한되지만, 식별 메타데이터 확장이 이루어지면 의미 있는 통합 검색 서비스가 제공될 수 있다. 이를 통해 특정 분야의 개별 등록관리기관이 보유하고 관리하는 콘텐츠에 대한 추천이 아닌 다양한 장르의 UCI 등록관리기관 전체가 관리하는 콘텐츠에 대한 추천을 통하여 사용자에게 개인 맞춤형 서비스가 될 수 있게 된다. UCI가 국가표준 식별체제로 활용됨에 따라 유통되는 모든 디지털콘텐츠가 UCI 운용시스템에 등록될 것으로 기대되고 있다. 따라서 향후에는 UCI 등록관리시스템과 총괄기구시스템이 확장성을 지원할 수 있도록 하는 기능 및 방법에 대하여 연구할 예정이다.

참고 문헌

[1] 이수정, 이형동, 김형주, "사용자 경향에 기반한 동적 추천 기법: 영화 추천 시스템을 중심으로," 정보과학회논문지, 제31권, 제2호, pp. 153-163, 2004.

[2] 이병욱, 송희석, 강신철, 남수현, 최준연, "유비쿼터스 환경에서의 상황인식 기반 VOD 콘텐츠 추천시스템," Telecommunications Review, 제

- 17권, 제6호, 2007.
- [3] S. Ahn and C.K. Shi, "Exploring Movie Recommendation Sytem Using Cultural Metadata," *IEEE Computer Society*, pp. 431-438, 2008.
- [4] F. Ricci and Q.N. Nguyen, "Acquiring and Revising Preferences in a Critique-Based Mobile Recommender System," *IEEE Intelligent Systems*, pp. 22-29, 2007.
- [5] Claypool M, Gokhale A, Miranda T, Murnikov P, Netes D, and Sartin M, "Combining Content-Based and Collaborative Filters in an Online Newspaper," *Proceedings of ACM SIGIR Workshop on Recommender Systems*, 1999.
- [6] Shardanand U and Maes P, "Social Information Filtering: Algorithms for Automating 'Word of Mouth'," *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 210-217, 1995.
- [7] Rong Hu and Yansheng Lu, "A Hybrid User and Item-Based Collaborative Filtering with Smoothing on Sparse Data," *ICAT*, pp. 184-189, 2006.
- [8] Jun Wang, Arjen P. Vries, and Marcel J.T. Reinders, "Unified Relevance Models for Rating Prediction in Collaborative Filtering," *ACM Transactions on Information Systems*, Vol.26, No.3, 2008.
- [9] 석중호, "식별체계기반 디지털콘텐츠 유통체제 구축방안 연구," 정보관리학회지 제20권 제4호, pp. 195-210, 2003.
- [10] Norman Paskin, "E-Citation : Actionable Identifiers and Scholarly Referencing," *Learned Publishing*, pp. 159-168, 2000.
- [11] 양준식, 조원희, 박광희, 유승훈, 김덕환 "유비쿼터스 컴퓨팅 환경에서 텔레매틱스 서비스를 위한 개인화된 추천 시스템의 설계 및 구현," 한국컴퓨터정보학회지, 제15권, 제2호, pp. 127-132, 2008.
- [12] 정경용, 조선문, "내용기반 필터링을 위한 프로파일 학습에 의한 선호도 발견," 한국콘텐츠학회논문지, 제8권, 제2호, pp. 1-8, 2008.
- [13] Balabanovic M and Shoham Y, "Fab: Content-Based, Collaborative Recommendation," *Communications of the Association for Computing Machinery*, pp. 66-72, 1997
- [14] 최인복, 박태근, 이재동, "소비자의 감성과 소비 유형을 이용한 협업여과기반 콘텐츠 추천 기법," 정보처리학회논문지, 제15-D권, 제3호, pp. 421-428, 2008.
- [15] 최인복, 이재동, "이웃크기를 이용한 사용자기반과 아이템기반 협업여과의 결합예측 기법," 정보처리학회논문지, 제16-B권, 제1호, pp. 55-62, 2009.
- [16] 이재식, 박석두, "장르별 협업필터링을 이용한 영화추천시스템의 향상," 한국지능정보시스템학회논문지, 제13권, 제4호, pp. 65-78, 1998.
- [17] 김지섭, 남제호, "방송콘텐츠 저작권 식별관리를 위한 UCI 표준식별체계와 내용기반 식별정보의 상호연계 연구," 방송공학회논문지, 제14권, 제13호, pp. 288-298, 2009.
- [18] 한국정보사회진흥원, "UCI 명세서 v2.2," 2007.
- [19] Popescul A, Unger L, Pennock D, and Lawrence S, "Probabilistic Model for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments," *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufman, San Francisco, pp. 437-444, 2001.
- [20] Manos Papagelis and Dimitris Plexousakis, "Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents," *Engineering Applications of Artificial Intelligence* 18, pp. 781-789, 2005.
- [21] UBCIM Publications, "Functional Requirements for Bibliographic Record," *New Series*, Vol.19, 1998.



나 문 성

1982년 3월~1989년 2월 조선대
학교 전자공학과 공학사
2001년 3월~2003년 8월 한양대
학교 경영대학원 석사
2006년 9월~2009년 9월 단국대
학교 컴퓨터과학과 박사
수료

1996년 9월~1997년 3월 운전시뮬레이터 개발
2005년 3월~2007년 2월 단국대학교 정보통신대학원
겸임교수
2000년 12월~현재 한국콘텐츠진흥원 제작지원 본부장
관심분야 : 멀티미디어 콘텐츠, 가상화, 하드웨어 시스템
구현



이 재 동

1985년 인하대학교 전자계산학
(학사)
1991년 Cleveland State
University(석사)
1996년 Kent State University
(박사)
1997년 3월~현재 단국대학교 컴

퓨터학부 컴퓨터공학전공 교수
2009년 8월~현재 단국대학교 미디어콘텐츠기술센터
센터장
2009년 7월~현재 한국문화콘텐츠기술학회 학회장
2006년 4월~현재 국가지정 단국대학교 CT연구소 소장
2002년 11월~현재 농협중앙회 전산고문
2007년 2월~2009년 3월 Dream economy leader 포럼위원
2006년 7월~2007년 12월 민관확대 콘텐츠 정책협의회
위원
2005년 8월~2006년 8월 문화관광부 KOCCA CT포럼/
전략기획 운영위원/분과위원장
2005년 3월~2009년 6월 단국대학교 콘텐츠&컨버전스
기술연구소 소장
2005년 1월~2006년 12월 전국대학정보화 협의회 이사
2004년 7월~2006년 6월 단국대학교 정보통신원 원장
(C.I.O)
2004년 1월~2006년 6월 (사)이러닝 산업협회 이사
관심분야 : Ubiquitous Computing, Contents & Enter-
tainment Technologies, (Mobile) Internet
Technology/Applications, Cloud Computing