

해외 학술지 전자원문 로컬호스팅 시스템 구현

Implementation of Electronic Document Local Hosting System of Overseas Journals

김광영, 김진숙, 박정훈, 김정환
한국과학기술정보연구원 해외정보실

Kwang-Young Kim(kykim@kisti.re.kr), Jinsuk Kim(jinsuk@kisti.re.kr),
Jung-Hoon Park(jhpark@kisti.re.kr), Jeong Hwan Kim(kimjh@kisti.re.kr)

요약

오늘날 인터넷과 전자출판 기술은 매우 발달되었고 고품질의 해외 학술정보를 이용하는 것은 매우 많은 이익을 가질 수가 있다. 국내 연구자들은 이런 고품질의 전자저널 원문에 대한 요구가 증대되고 있다. 또한 국가 상호간에 전자저널들을 안전하게 보존 및 서비스하기 위해서는 서로 간에 디지털 아카이빙의 활동이 필요하다. 따라서 본 논문에서는 전자저널 국가컨소시엄 참가기관 이용자들에게는 One-Stop 전자원문 무료 서비스와 중소기업, 개인이용자들에게도 저렴한 전자원문 유료 서비스 제공하며 국가적인 전자정보자원의 장기보존체계를 마련을 하기위한 전자원문 로컬호스팅 시스템을 구현하였다.

■ 중심어 : | 전자원문 | 로컬호스팅 | 관리시스템 |

Abstract

Today, As the internet and electronic publishing technology was powerful and there are many benefits to using various academic information which has an electronic document of high quality. Many researchers had much required to see a full-text of electronic document. There was an requirement of digital achieving activities between countries for each other electronic document in order to safely preserve and services. In this paper, we have implemented an electronic document local-hosting system to provide free service to KESLI member institutions and provide pay-per-view service to individual users and small/medium size companies who are not member of KESLI and build national long-term preservation system of electronic information resources.

■ keyword : | Electronic Document | Local-hosting | Management System |

I. 서론

오늘날 학술지를 포함한 과학기술분야의 정보량은 지난 30년 동안 꾸준히 증가하고 있다[1]. 과거에는 학술정보 서지 정보들을 중심으로 구축이 되었지만 현재 사용자들은 전자원문을 보고자하는 욕구가 증대되고

있다. 그러나 학술지의 구독 단가도 전자매체의 등장 이후 잠시 정체하는듯 하였으나 계속 상승함으로써 학술지 구독을 위한 도서관에 요구되는 소요 비용도 매년 급격히 증가하고 있다. 또한 학술지 원문제공서비스의 본질적인 가치인 정보 및 정보서비스는 다양한 가치 속성 및 무형자산과 공공재의 성격을 지니고 있어서 정량

적이고 객관화된 가치를 평가하기가 쉽지 않다[2].

사용자들의 이런 욕구를 충족시키기 위해서 한국과학기술정보연구원(KISTI)에서는 전자저널 국가컨소시엄(KESLI)의 전자저널 공동구매를 통하여, 세계 유수한 전자저널의 구독단가를 최대한 낮추고 인쇄저널로 구독하지 않는 저널이라도 컨소시엄에 가입함으로써 전자원문(full-text)까지 이용할 수 있도록 노력을 하고 있다.

디지털문헌의 아카이빙(digital archiving)은 말 그대로 디지털문헌을 안전하게 보존하는 활동이며, 시간이 경과되어도 그것에 접근할 수 있고 진본을 유지할 수 있도록 하는 모든 행위들을 말한다. 보존문제가 부각하게 된 배경은 정보기술의 발전으로 인하여 디지털 정보의 생산과 유통이 급격하게 증가하였기 때문이다[3]. 우리는 아직 디지털 보존의 사각지대에 있으며 유네스코 한국위원회와 정보트러스트센터가 주최하고 국립중앙도서관이 후원한 디지털 유산 보존을 위한 포럼에서 보존에 대한 인식 제고를 촉구하는 논의가 이루어진 바 있고, 국립중앙도서관에서 온라인 디지털자원 보존사업을 추진하고 있으나 아직 국가지식자원 관리 차원에서 보존의 문제를 인식하고 있다는 흔적은 엿보이지 않는다고 기술하였다[4]. Paper to Digital이라는 시대적 조류에 부응하여 역동성과 부가가치 창출이 용이한 디지털 정보자원을 개발하고 체계적인 보존과 활용을 위한 디지털 아카이빙 시스템 구축은 반드시 필요하다. 또한 국가적인 전자정보자원에 대한 장기보존체계 마련을 위해서도 디지털 아카이빙 시스템 구축은 필요로 한다.

해외 전자원문의 데이터베이스 구축과 서비스를 하기 위해서는 일관된 관리절차 확립과 지속적으로 입수되는 자료들에 대한 효율적인 관리 및 표준화에 근거한 신속한 정보서비스 체계를 구축해야 할 필요성이 있다 [5]. 국내 학술논문의 전자원문들은 한국과학기술정보연구원(KISTI)에서 수집하고 서비스를 하고 있으므로 국내의 이용자들은 보다 빠르게 접근을 할 수가 있다. 반면 해외 학술지들의 전자원문은 대부분 출판사들이 소유를 하고 있으므로 사용자가 해외 전자원문을 보기 위해서는 결국 해당 출판사로 접근해야 한다. 만약 이

런 출판사가 사이트 운영을 중단하거나 작은 규모의 출판사가 없어질 경우에 그 전자원문을 즉시 볼 수가 없다. 물론 국내에서 소장하고 있는 인쇄학술지에 대해서 원문 복사 신청하여 볼 수 있지만 자료에 대한 빠른 접근 원하는 사용자에게는 불편한 점들이 발생한다.

따라서 본 논문에서는 출판사와 협약을 통해서 입수된 해외 전자원문을 사용자에게 빠르고 다양한 서비스를 제공하며 전자저널 구독이 중단될 경우에도 과거 구독분에 대해서는 사용자들에게 지속적으로 서비스를 제공과 대용량의 전자원문들을 안전하게 보존하는 보존센터의 역할도 수행할 수 있는 전자원문 로컬호스팅 시스템을 개발하고자 한다.

II. 국내외 관련 연구

1. 국내 동향

NSDL은 한국과학기술정보연구원(KISTI)에서 수행하는 국내의 논문, 특허, 보고서 등의 다양한 디지털자원을 서비스하는 포털사이트이다. 현재 구축건수는 약 9천만 건의 논문, 특허, 보고서, 동향, 표준, 사실정보 등을 제공하고 있으며 11,807종의 전자저널의 전자원문을 이용할 수 있다. 그중에 해외학술지는 약 4천6백 건을 구축하고 있다.

국내의 경우, 디지털 콘텐츠에 대한 아카이빙은 시작 단계이며, 현재 국회도서관, KERIS, KISTI 등의 국가기관과 북토피아, DBPLA 등의 개인 기업체에서 디지털 콘텐츠를 저장 관리하고 있다. 최호남(2005)은 디지털 아카이브는 장기 보존 및 접근에 대한 체계적 접근이 없는 상태로 진정한 의미의 아카이브로 보기는 어려운 실정이라고 기술하였다[6]. 국립중앙도서관, 국가 기록원 등은 장기보존 전략을 가지고 있으며 현재 KISTI에서도 디지털 자원의 항구적인 접근을 위한 전략을 수립하고 있다[7][8].

2. 국외 동향

국가대표 도서관이나 기억조직(memory institute)에서 디지털 정보자원의 중요성을 인식하고 정부 혹은 국

가의 지지를 바탕으로 자료의 수집을 시작하는 형태로 미국의 NDIIPP(National Digital Information Infrastructure and Preservation Program), 영국의 DCC(Digital Curation Centre)와 JISC(Joint Information Systems Committee)를 대표적인 사례로 꼽을 수 있으며 그 외에도 호주 국립도서관을 중심으로 한 PANDORA (Preserving and Accessing Networked Documentary Resources of Australia), 일본의 NII-REO(National Institute of Informatics - Repository of Electronic Journals and Online Publication)의 디지털 아카이빙 사업 등이 있다[7].

네덜란드 국가도서관(National Library of the Netherlands, KB)[9]에서는 e-Depot을 중심으로 매체 노후화 및 소프트웨어와 하드웨어 플랫폼의 빠른 진보에 대비하여 디지털 객체에 대한 장기적인 접근을 보장하는 것을 목적으로 자국의 디지털납본자료, 웹 아카이브, 디지털화 자료는 물론 전 세계 디지털발간물을 아카이빙하고 있고 2003년부터 2008년 12월 현재까지 e-Depot이 수집한 1,200만 건 이상의 학술지 대부분이 Elsevier, Springer 등의 국제적 출판사의 발간물들이다.

Portico[10]는 2002년 앤드류멜론재단(Andrew W. Mellon Foundation)의 재정지원으로 JSTOR가 시작한 Electronic-Archive Initiative에서 출발하였다. 전자학술지의 장기적 아카이브를 제공하는 비영리서비스로 2005년부터 새로운 전자 아카이빙 서비스가 시작되었고 앤드류멜론재단은 물론이고, 최근에는 미국 의회도서관(Library of Congress, LC)의 재정지원을 받고 있다. 참여 출판사가 제공하는 전자형태로만 발간되는 (born digital) 전자학술지, 인쇄 및 전자형태가 병행 발간되는 학술지, 원래 인쇄본을 디지털화한 학술지 모두를 아카이빙 대상에 포함시키고 있다.

LOCKSS(Lots of Copies Keep Stuff Safe)[11]는 스탠포드대학 도서관(Stanford University Libraries)에 본부를 두고 있는 국제적 프로그램으로 도서관이 자관 소유의 디지털콘텐츠를 저렴한 비용으로 쉽게 수집·보존할 수 있도록 디지털보존 툴(공개 소프트웨어)과 각종 지원 서비스를 제공하고 있다. LOCKSS 회원도서

관은 현재 약 400개 학술출판사가 발간하는 전자학술지 2,700종의 보존권한을 획득과 아시아/태평양지역에서는 다음의 중국, 뉴질랜드, 싱가포르 도서관이 회원으로 참여하고 있다.

Journal@rchive[12]는 일본 과학기술진흥기구(科學技術振興機構, Japan Science and Technology Agency, JST)에서 과학기술정보 발신·유통시스템(J-STAGE)의 아카이브 사이트로 2005년부터 실시하고 있는 전자 아카이브 사업이다. 일본의 연구 성과에 대한 국제 발신 능력 강화와 중요한 지적 재산의 보존을 목적으로 한다. 일본 자국내 학협회가 발간하는 학술지를 장간호부터 전자화하여 공개함으로써 메이지 시대부터 일본의 연구 성과를 열람할 수 있고, 세계에 영향을 준 저명한 논문도 많이 수록하고 있다. 전자 아카이브 대상지는 일본학술회의 등 관련 기관의 협력으로 선정하고 있으며, 자연과학분야 뿐만 아니라 인문, 사회과학분야 학술지도 대상으로 함으로써 2005년 74개, 2006년 65개, 2007년 58개, 2008년 181개, 2009년 266개가 선정되었고 2010년도 이후에도 추가로 선정하여 범위를 확대해 나갈 예정이다.

Technical Reports Archives and Image Library(TRAIL)[13]는 미국의 애리조나대학(University of Arizona)이 주도하는 Greater Western Library Alliance(GWLA)는 Center for Research Libraries(CRL) 등과 협력하여 1975년 이전에 발행된 미국 연방정부의 기술보고서를 수집하고 디지털화하여 아카이브를 구축한 후 자유롭게 활용시키는 것을 목적으로 하고 있다.

Information Bridge[14]는 미국 에너지부(Department of Energy, DOE)가 자체적으로 또는 연구비를 지원하여 생산한 연구보고서 전자원문(full-text) 및 서지정보 약 21만 건을 일반에게 무료로 제공하고 주제 분야는 물리학, 화학, 재료학, 생물학, 환경학, 에너지기술, 컴퓨터 및 정보학, 재생산에너지 등과 주로 1991년 이후에 생산된 것이지만 그 이전에 생산된 연구보고서의 디지털화 작업이 지속적으로 진행되고 있다.

NASA Technical Reports Server(NTRS)[15]는 NASA의 기술문헌을 학생, 교사, 연구자, 일반인이 이용할 수 있도록 서비스를 제공하고 NASA가 자체 생산

하거나 재정지원으로 생산되는 과학기술정보가 늘어남에 따라 계속 증가하고 있고 단일 인터페이스를 통해서도 다른 3가지 컬렉션(NACA Collection, NASA Collection, NIX Collection)에 포함된 정보를 검색하는 것이 가능하다.

이와 같이 국외에서는 국가대표 도서관, 대학이나 연구기관 등을 중심으로 많은 프로젝트들이 진행 및 개발되고 있다. 또한 안전하게 디지털 자원을 보존하기 위해서 국가나 기관 간에 협약을 통해서 디지털 자원을 상호 교환하여 보존을 할 필요성이 높아지고 있다.

III. 전자원문 로컬호스팅

본 논문에서는 사용자에게 대용량의 전자원문을 빠르게 접근할 수 있도록 체계적인 전자원문 시스템을 구축하며 디지털 전자원문들을 영구적인 접근을 위한 효율적인 보존 시스템을 개발하고자 한다. 따라서 본 논문에서는 해외출사와 협약에서부터 전자원문 입수, 관리, 보존 및 서비스까지 하나의 플랫폼기반의 시스템을 구현 및 개발하였다.

1. 전자원문 수집 및 분석

1.1 전자원문 프로세스

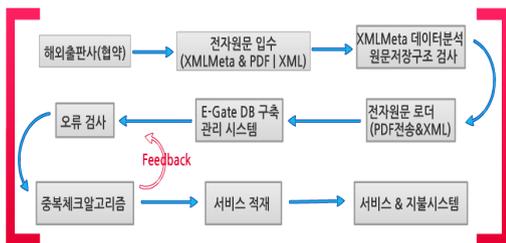


그림 1. 전자원문 프로세스

[그림 1]과 같이 전자원문의 입수에서 서비스 단계까지 프로세스를 보면 아래와 같다.

첫째, KESLI는 해외출판사들과 상호 협약을 통해서 전자원문 입수 준비 작업을 수행한다. 또한 필요에 따라서 서지 메타데이터 상호 교환 또는 원문전체를 상호

교환할 수가 있다.

둘째, KESLI는 해외출판사와 협약이 체결되면 출판사들로부터 서지 메타데이터와 전자원문 전체를 입수한다. 즉 출판사의 제공하는 학술 메타데이터와 PDF 또는 XML 등의 대용량의 전자원문을 입수한다.

셋째, 데이터 분석기는 입수된 전자원문과 서지 메타데이터를 분석하여 전자원문과 서지 메타데이터 간의 매핑 처리를 수행한다. 즉 입수된 서지 메타데이터와 전자원문 간에 상호 연결 정보를 생성한다.

넷째, 데이터 분석기는 매핑된 전자원문과 서지 메타데이터를 이용하여 학술콘텐츠 관리시스템(E-Gate)에 필요한 요소항목들을 메타데이터에서 추출하여 저장한다. 그리고 전자원문은 전자원문 서버로 전송하며 로컬 호스팅 관리 시스템은 저장된 전자원문들을 체계적으로 관리한다.

다섯째, 학술콘텐츠 관리시스템은 등록된 서지 정보에 대해서는 오류 검사나 중복 체크를 한 후에 신규 추가 또는 갱신 처리를 수행하며 전자원문 서비스와 연결을 하기위해서 전자원문의 경로 정보와 로컬 존재 유무 정보를 관리한다. 출판사로부터 보존용으로 입수된 전자원문의 경우에는 다크(dark)아카이빙용으로 저장 관리하며 유사시에 라이트(light) 아카이빙용으로 전환하여 서비스할 수 있도록 처리한다.

다크 아카이빙은 유사시를 대비하여 출판사로부터 전자원문을 입수하여 보존만 할 수 있으면 직접 서비스를 할 수 없는 원문을 의미한다. 라이트 아카이빙은 출판사로부터 협약을 통해서 전자원문을 입수하여 보존 및 직접 서비스를 할 수 있는 원문을 의미한다.

여섯째, 모든 오류, 중복 등의 검사 수행이 끝날 경우 서비스에 적재하여 사용자에게 전자원문 서비스를 제공한다. 일반 이용자나 기관 이용자들은 검색 서비스 통해서 색인 및 적재된 전자원문들을 검색 및 열람할 수가 있다.

1.2 메타데이터와 전자원문 매핑

해외출판사로부터 전자원문과 메타데이터를 입수 받을 경우에 디렉터리별로 구성된 서지 메타데이터 및 전자원문 파일들을 입수한다. 또한 입수된 전자원문과 서

지 메타데이터의 연결을 하기 위해서는 입수 처별로 메타데이터와 전자원문을 분석 및 매핑 규칙을 만들어 연결한다. 아래의 [그림 2]와 같이 입수 처별로 메타데이터와 전자원문들 간의 관계를 분석하여 정규표현식이나 와이드카드 등과 같은 방식을 이용하여 매핑 처리한다.

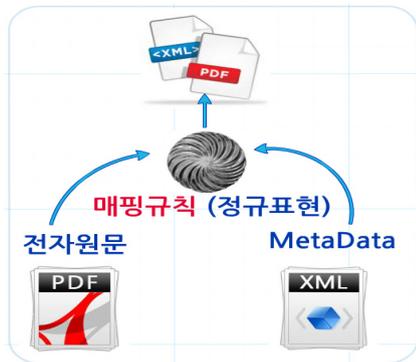


그림 2. 메타데이터와 전자원문 매핑

입수된 데이터들은 각 출판사별로 규칙이 상이함으로 각 출판사 별로 매핑 규칙들을 만들어 환경 설정에서 저장 관리한다. 입수된 데이터들의 디렉터리는 각 저널식별자, 권, 호, 및 전자원문 식별자 등으로 구성되어 있으므로 원형을 그대로 유지하도록 해야 한다. 그러나 비체계적으로 구성된 경우에는 체계적으로 관리할 수 있도록 새로이 구성한다.

1.3 입수 메타데이터 항목 분석

입수된 메타데이터는 기존의 학술콘텐츠 관리시스템(E-Gate)에서 자동으로 메타데이터들을 분석하여 필요한 정보들을 추출하였지만 전자원문 로컬호스팅 서비스를 하기 위해서는 추가적인 정보들을 추출해야한다. 각 출판사별로 서지 메타데이터의 형식은 다양하기 때문에 각 출판사별로 서지 메타데이터 및 전자원문정보를 추출할 수 있는 분석기의 역할은 매우 중요하다. 즉 XML, SGML, TXT 등 다양한 포맷으로 서지 메타데이터가 입수된다.

본 논문에서는 [표 1]과 같이 기존의 학술콘텐츠 관

리시스템(E-Gate)에서 사용하고 있는 DC(Dublin Core) 메타데이터 분석기를 확장하여 로컬호스팅 서비스에서 필요로 하는 "원문 URL" 항목을 추가적으로 할당을 하였다. 즉 전자원문의 경로 정보는 전자원문 로컬호스팅 서비스를 위해서 필수 항목이다. 따라서 위의 항목에 논문의 경로를 추가하여 사용자가 전자원문을 클릭할 경우 전자원문 열람할 수 있도록 처리해야 한다.

로컬호스팅 관리 시스템은 전자원문의 입수처별, 저널별로 체계적으로 관리하여 전자원문 서비스를 제공하기 위해서 전자원문이 로컬 서버에 존재 여부 필드를 추가하였다. 또한 유사시에는 다크 아카이빙을 라이트 아카이빙으로 전환처리하기 위해서 정보원별로 아카이빙 체크를 할 수 있는 필드를 추가하였다.

표 1. Springer 메타데이터 예

항목명	예) Springer 입수항목	DC XML 엘리먼트
발행자명	PublisherInfo/PublisherName	DC.Publisher
발행지역	PublisherInfo/PublisherLocation	DC.Publisher.place
발행자홈페이지	PublisherInfo/PublisherURL	DC.Publisher.homepage
...
Electronic ISSN	Journal/JournalInfo/JournalElectronicISSN	DC.Identifier.electronicssn=ISSN
계재지명	Journal/JournalInfo/JournalTitle	DC.Relation.isPartOf.title
...
논문URL	article-meta/self_uri@content-type	DC.Identifier.article

2. 로컬호스팅 시스템

[그림 3]과 같이 전자원문 로컬호스팅 시스템은 입수-관리-검증-통계 등의 전체 과정을 체계적으로 관리할 수 있도록 구성하였다. 또한 입수부터 이력관리를 통해서 중간에 발생할 수 있는 오류 등에 대해서 지속적으로 모니터링 처리를 수행한다.

전자원문 로컬호스팅 시스템은 크게 외부시스템, 수집기, 관리/검증 및 통계/모니터링으로 나누어진다. 로컬호스팅 시스템의 각 세부적인 기능별 설명은 아래와 같다.

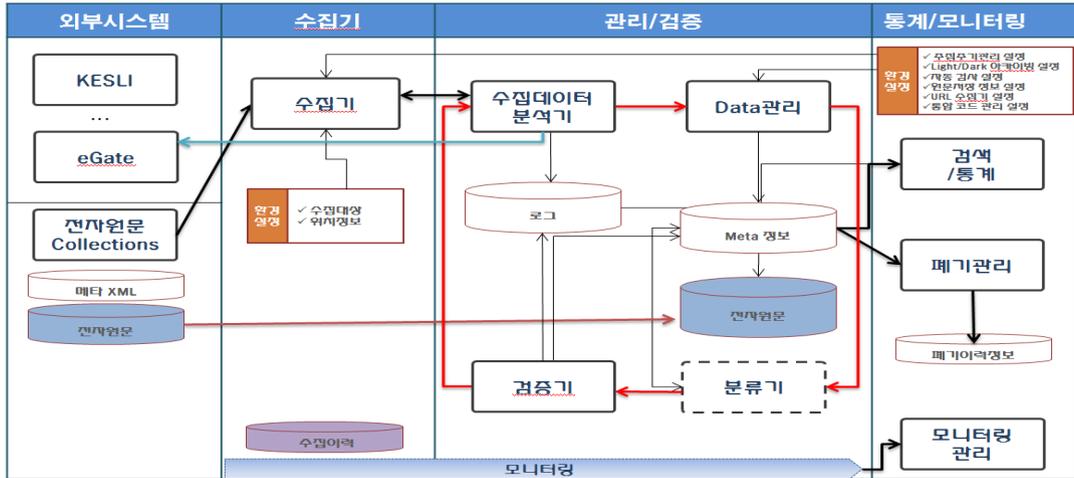


그림 3. 로컬호스팅 시스템 프로세스

2.1 외부시스템

외부시스템은 로컬호스팅 관리 시스템과 밀접한 관계를 가진다. 즉 출판사와 협약을 담당하고 있는 KESLI와 협약된 전자원문을 서비스할 수 있도록 라이선스 정보와 연계 등이 필요하다. 로컬호스팅 관리 시스템은 학술 데이터 정보를 입수-관리하는 학술콘텐츠 관리 시스템(E-Gate)과 밀접한 관계를 가지고 있으며 상호 유기적으로 연결한다.

2.2 수집기

수집기는 오프라인으로 입수된 CD나 대용량 데이터를 관리자가 직접 업로드를 수행하며 적재된 데이터들은 자동으로 오류나 중복 처리 등의 이력정보 관리를 수행한다. 수집기는 온라인으로 수집 가능한 전자원문과 서지 메타데이터들은 자동으로 수집하는 역할도 수행한다. 또한 입수되는 전자원문에 대한 위치정보 및 수집대상 정보 등을 관리한다. 이렇게 수집된 데이터들은 주기적으로 수집데이터 분석 시스템으로 전자원문과 메타데이터를 넘겨주게 된다.

2.3 수집 데이터 분석기

수집 데이터 분석기는 출판사로부터 입수된 논문의 서지 메타데이터를 분석한다. 수집 데이터 분석기는 학

술콘텐츠 관리 시스템(E-Gate)에서 필요로 하는 서지 메타데이터 정보와 검색 서비스에 필요로 하는 메타데이터 정보들을 자동으로 추출하여 학술콘텐츠 관리 시스템에 추가/갱신을 수행한다. 해외출판사에서 전체 메타데이터를 입수 받을 경우에 기존에 구축된 학술콘텐츠 관리 시스템에서 누락 또는 신규 데이터들이 포함되어 있기 때문에 이와 같은 작업을 동시에 수행한다. 이렇게 함으로써 기존의 학술콘텐츠 관리 시스템도 최신의 서지 데이터 정보를 유지한다.

수집 데이터 분석기는 로컬호스팅 관리 시스템을 위해서 전자원문의 경로 정보, 정보원, 입수날짜, 제목 등을 추출한다. 수집 데이터 분석기는 각 출판사별로 입수되는 다양한 형식의 서지 메타데이터를 분석해야 한다. 또한 입수된 전자원문을 서비스하기 위해서 전자원문 서버로 원문 파일 전송한다. 수집 데이터 분석기는 전자원문과 메타데이터 매핑 정보를 활용한 전자원문 접근 경로 설정에 따라 자동으로 분류되어 관리할 수 있도록 한다.

2.4 Data관리

Data관리 시스템은 전자원문들의 데이터들을 관리한다. [표 2]와 같이 전자원문 데이터 관리를 위한 제어번호, 정보원, 경로 의 다양한 정보를 관리한다. 또한 데이터 관리 시스템은 분석기에서 분석한 내용의 오류나 중

복 등을 체크할 수 있다. 제어번호는 기존의 학술콘텐츠 관리 시스템과의 메타데이터와 연계 정보이다. 정보원은 출판사의 정보를 표시하며 경로는 전자원문 서버에 물리적인 전자원문 위치 정보이다.

표 2. 원문 파일 관리 정보

항목	예시	비고
제어번호	5121311	관리를 위한 고유 번호
정보원	Springer	출판사
경로	디렉토리/파일	전자원문 경로
...

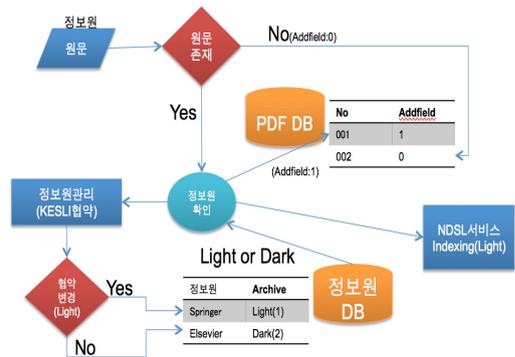


그림 4. Data관리 흐름도

[그림 4]와 같이 데이터 관리 시스템은 입수된 전자원문이 로컬 서버에 저장되어 있는지를 표시하기 위해서 PDF DB의 Addfield에 값을 1로 세팅한다. 또한 입수된 전자원문을 서비스하기 위해서는 반드시 다크 (dark) 아카이빙용인지 아니면 라이트(light) 아카이빙용인지를 구분 처리해야한다. 협약이 된 라이트 아카이빙용의 전자원문들은 서비스 가능하지만 서비스 협약은 되지 않고 보존 다크 아카이빙용으로 입수된 전자원문들은 로컬호스팅 서비스를 수행 할 수가 없다. 하지만 유사시에는 보존용으로 입수한 다크 아카이빙용 전자원문들에 대해서도 자동으로 서비스 가능하게 전환을 할 수 있도록 처리한다. 라이선스 내용들은 KESLI 라이선스 관리 시스템과 연계하여 로컬로 가지고 있는 여부, 로컬호스팅 가능한지 여부 등의 정보를 확인하여 서비스 여부를 결정한다.

협약이 변경될 경우에 수백만 건의 PDF DB 필드를 변경하지 않고 [그림 4]에서 정보원 관리를 통해서 해당하는 정보원만 변경할 수 있도록 하였다. 예를 들면 Elsevier가 협약을 통해서 Light로 전환이 될 경우에는 정보원 DB만 Dark→Light로 전환 후에 NDSL서비스에서만 다시 색인을 수행할 수 있는 시스템을 구성하였다.

2.5 검증기

검증기는 로컬호스팅 전자원문 서비스 중에 발생할 수 있는 오류를 최소화하기 위해서 필요하다. 검증기는 전자원문의 오류 여부를 자동으로 검사하기 위해서 전자원문의 파일 사이즈, 정확한 물리적인 경로 여부를 검사한다. 또한 관리자가 직접 샘플링을 통해서 전자원문 오류 여부 검사를 수행할 수 있도록 하였다.

[그림 5]와 같이 자동검사일 경우에는 직접 PDF파일 경로 정보를 이용하여 자동으로 전자원문의 파일 크기 검사와 열어보고 오류가 발생하는지 여부를 체크하여 오류가 발생할 경우에 오류DB에 저장한다. 만약 관리자가 수동 검사를 수행하고자 할 경우에는 직접 샘플링하여 검사를 수행할 수 있도록 하였다.

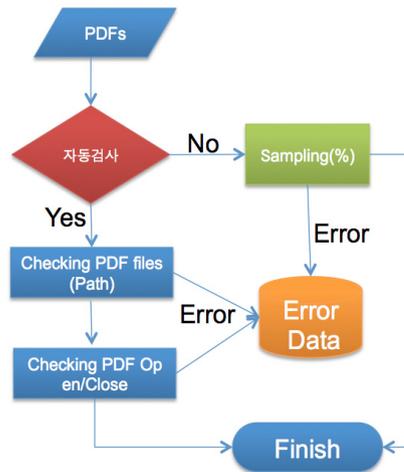


그림 5. 전자원문 검증기 흐름도

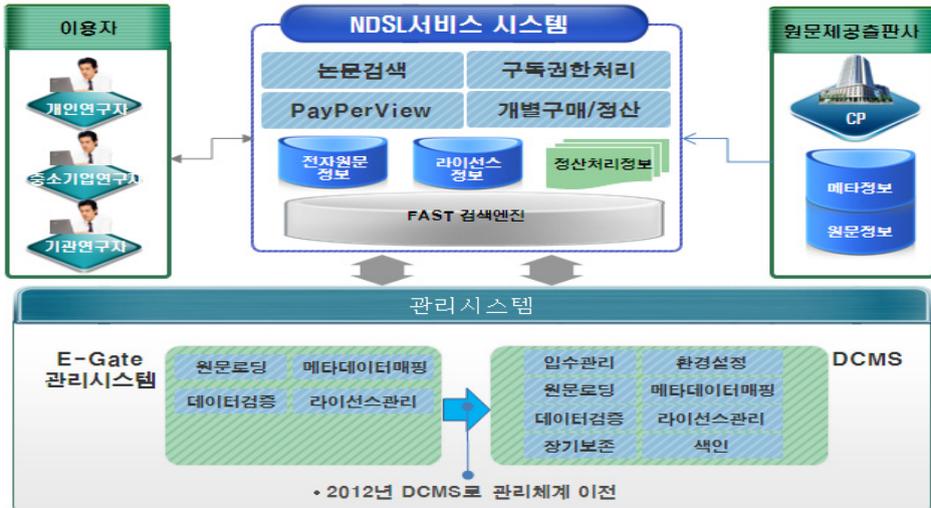


그림 6. 서비스와 연계

2.6 통계/모니터링 관리

통계기능은 입수된 전자원문을 년/월/일별로 통계를 수행하여 전체 전자원문의 건 수 등의 정보들을 조회하는 기능이다. 모니터링 관리 시스템은 입수에서부터 전체과정을 모니터링 및 수집 이력관리를 담당하게 되며 입수에서부터 검증까지 전체 과정에서 오류가 발생할 경우에 대해서 모니터링을 수행한다.

2.7 폐기관리

폐기관리는 학술콘텐츠 관리 시스템(E-Gate)의 라이선스와 밀접한 관계를 유지한다. 만약 라이선스 협약이 완료될 경우 해당되는 전자원문들은 자동으로 다크(dark) 보존 아카이빙으로 처리하여 로컬호스팅 서비스를 수행을 하지 않도록 한다. 반대로 다시 협약이 될 경우에는 다크 보존 아카이빙에서 라이트(light) 아카이빙으로 전환을 수행하여 로컬호스팅 서비스를 수행하도록 한다.

2.8 환경설정

환경설정은 전자원문 저장 정보, 전자원문의 자동 검사 설정, 수집 주기 관리 설정, 로컬호스팅 관리시스템에 필요로 하는 통합코드 관리, 사용자 관리 등의 다양한 환경설정 정보들을 관리한다.

3. 서비스와 연계 시스템

[그림 6]와 같이 로컬호스팅 시스템에서 수집된 전자원문은 필요한 정보들을 학술콘텐츠 관리 시스템(E-Gate)에 등록 처리한다. 일반이용자 또는 기관협력 이용자일 경우에는 이용자 라이선스 정보를 통해서 카드정산 또는 후불정산을 통해서 로컬호스팅의 전자원문들을 열람한다. 만약 사용자가 로컬호스팅에 없는 전자원문을 선택할 경우에는 콘텐츠 제공자(content provider)의 사이트로 직접 연계를 하여 전자원문을 볼 수 있도록 처리한다.

3.1 전자원문 서비스 로직

로컬호스팅 관리 시스템의 전자원문은 최종적으로 NDSL을 통해서 서비스가 된다. 따라서 일반 사용자의 경우에는 NDSL 검색서비스를 통해서 검색한 논문에 대해서 전자원문을 열람하고자 할 경우에 구독권한 여부를 통해서 개별구매/정산 후에 전자원문을 열람을 할 수가 있다. PayPerView는 라이선스 미확보 기관 이용자를 위해 예약금제도를 도입하고 예약금은 이용자가 소속해있는 기관단위로 정산 처리한다. 또한 핸드폰소액결제, 신용카드 등 다양한 결제수단을 제공해야한다.

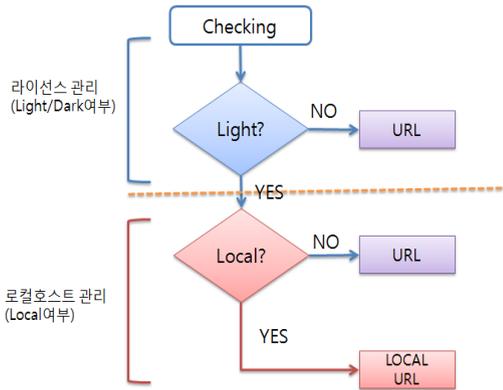


그림 7. 전자원문 서비스 로직

[그림 7]와 같이 사용자가 전자원문을 요청할 경우에 해당하는 전자원문을 출판사로 연결할 것인지 아니면 로컬호스팅에서 직접 서비스 할지 여부를 결정하는 방식이다. 즉 로컬호스팅 관리 시스템에서는 기관에서 보유하고 있는 전자원문들 중에서 라이트(light) 아카이빙 용과 다크(dark) 아카이빙용으로 분리되어 관리한다.

만약에 라이트 아카이빙용이지만 현재 구축단계에서 누락 되었거나 구축 중일 때는 해당 출판사의 URL 정보로 연결한다.

3.2 지불 시스템 연계

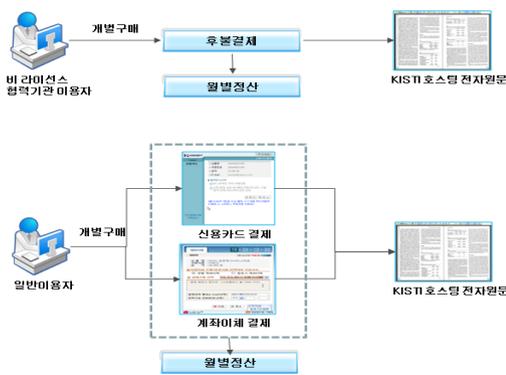


그림 8. 지불 시스템

[그림 8]과 같이 일반이용자와 협력기관 이용자가 라이선스가 있는 로컬전자원문들을 열람을 위해서 각각 다른 프로세스로 처리한다. 일반 이용자들은 신용카드

나 계좌이체 등의 결제를 통해서 라이선스가 있는 전자원문 로컬호스팅 서비스를 받을 수가 있다. 그러나 협력기관 이용자는 기관별로 월별정산을 통해서 열람을 처리함으로 바로 전자원문 로컬호스팅 서비스를 제공 받을 수 있다.

3.3 통계 서비스

통계 서비스 시스템은 서비스되고 있는 전자원문에 대해서 판매 통계 관리기능이다. 즉 논문별 이용현황 중에서 가장 많이 이용된 논문 등의 정보를 확인할 수가 있으며 월별/일련 판매현황, 논문 판매현황, 회원별 판매현황 등과 같이 다양한 통계 서비스를 제공한다.

IV. 결론

오늘날 정보의 생산량은 점점 많아지고 정보의 질이 높은 선진국에서 만들어지는 학술정보 원문에 대한 국내의 요구가 증대되고 있다. 과거 학술정보의 서지 정보들을 중심으로 구축이 되었지만 현재는 사용자들은 전자원문을 보고자하는 욕구가 증대되고 있다.

본 논문에서는 대용량의 전자원문을 입수에서 관리까지 체계적으로 처리할 수 있는 로컬호스팅 관리 시스템을 구축하여 사용자에게 빠르고 안전하게 전자원문 접근을 할 수 있도록 서비스 체계를 구축하였다. 또한 전자원문 로컬호스팅 관리 시스템은 다크 아카이빙용의 전자원문 확보를 통하여 유사시에 다른 출판사에서 전자원문들을 서비스 하지 못할 경우에 자동으로 서비스 전환을 통해서 사용자에게 지속적으로 고품질의 전자원문들을 서비스를 할 수 있는 체계를 마련하였다. 또한 전자원문 로컬호스팅 관리 시스템은 입수부터 서비스까지 전체 과정들을 관리 및 모니터링할 수 있는 체계로 구축함으로써 관리의 효율성을 높였다.

향후 연구과제로는 고품질의 전자원문들을 활용하여 사용자에게 다양한 서비스를 제공할 필요성이 높다. 예를 들면 전자원문XML에서 그림, 표 및 참고문헌 등의 다양한 정보들을 추출 및 구축하여 지금 보다 한 단계 더 높은 서비스를 제공해야 한다. 또한 전자원문들을

영구 보존하기 위한 메타데이터, 마이그레이션 등의 과정들도 추가적으로 연구가 필요로 하다.

참고 문헌

[1] Fernandez-Cano, Antonio, Manuel Torralbo, and Monica Vallejoa, "Reconsidering Price's Model of Scientific Growth:An Overview," *Scientometrics*, Vol.61, No.3, pp.301-321, 2004.

[2] 정영미, "해외학술지 원문제공서비스의 경제적 가치 평가 방법론 및 모델 개발", *정보관리학회지*, 제25권, 제4호, pp.243-267, 2008.

[3] 이수상, "디지털 아카이빙의 워크플로우와 보존 처리 기술에 관한 연구", *한국도서관정보학회지*, 제35권, 제3호, pp.119-138, 2004.

[4] 설문원, *국가 디지털 아카이빙 체제 구축에 관한 연구*, 한국과학기술정보 연구원, 2005.

[5] 류범중, 이용봉, "해외 기술정보 메타 데이터베이스 구축 및 관리시스템에 관한 연구", *한국문헌정보학회지*, 제36권, 제3호, pp.201-213, 2002.

[6] 최호남, 이용봉, "해외 전자저널의 디지털 아카이브 구축 전략에 관한 연구", *한국문헌정보학회지*, 제39권, 제2호, pp.161-183, 2005.

[7] 광승진, *수명주기 기반 디지털 콘텐츠 아카이빙 정책 연구*, 한국과학기술정보 연구원, 2010.

[8] 정영임, 최호남, 최선희, "아카이빙 데이터의 활용성 증진을 위한 전략연구: 국내외 학술논문을 중심으로", *정보관리학회지*, 제27권, 제1호, pp.185-206, 2009.

[9] <http://www.kb.nl/hrd/dd/index-en.html>

[10] <http://www.portico.org/digital-preservation>

[11] <http://www.lockss.org/lockss/Home>

[12] <http://www.journalarchive.jst.go.jp>

[13] <http://trail.gwla.org>

[14] <http://www.osti.gov/bridge>

[15] <http://ntrs.nasa.gov>

저자 소개

김 광 영(Kwang-Young Kim)

정회원



- 2011년 : 충남대학교 문헌정보학과(박사)
- 2001년 ~ 현재 : 한국과학기술정보 연구원 선임연구원

<관심분야> : 정보검색시스템, 개인화 검색시스템, 디지털도서관, 아카이빙 시스템

김 진 숙(Jinsuk Kim)

정회원

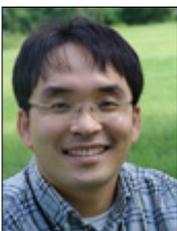


- 1995년 : KAIST 생물학과(석사)
- 2002년 : KAIST 전산학과(석사)
- 2002년 ~ 현재 : 한국과학기술정보 연구원 선임연구원

<관심분야> : 정보검색, 생물정보학, 비자언어처리 및 검색알고리즘

박 정 훈(Jung-Hoon Park)

정회원



- 2003년 : 충남대학교 문헌정보학과(석사)
- 2002년 ~ 현재 : 한국과학기술정보 연구원 선임연구원

<관심분야> : 메타데이터 레지스트리, 정보검색시스템, 디지털정보자원, 모바일웹

김 정 환(Jeong Hwan Kim)

정회원



- 2009년 : 충남대학교 문헌정보학과(박사수료)
- 2006년 ~ 현재 : 한국과학기술정보연구원 책임연구원

<관심분야> : 컨소시엄, 정보자료 공동수집 및 공동활용, 디지털 정보자원 개발 및 아카이빙 시스템