

전자저널 기술현황과 국내 학술지 전자저널 구축방안

한국과학기술정보연구원 | 서태설 · 이혜진

1. 서론

인터넷과 디지털 기술의 발전으로 학술정보의 생산, 출판, 유통, 활용에 이르는 학술활동이 디지털 기반으로 변화함에 따라, 학술지는 급속히 전자저널화되고 있다. 이에 따라, 출판사는 학술지의 유통정책을 개선하고, 학술논문의 가독성, 온라인 접근성을 향상시키기 위한 노력을 하고 있다.

새로운 학술발견의 공유를 목적으로 1665년에 처음 생겨난 학술지(scientific journals)는 18세기부터 급격히 증가하여, 현재는 수만 종의 학술지가 전 세계에서 발간되고 있다. 학술지는 새로운 연구결과를 동일 분야 연구자에게 공개함으로써 과학발전에 기여하는 역할을 하고 있다.

그런데 학술지를 출판하여 배포하는 데에는 많은 비용이 들기 때문에 학회들의 경우는 회원들에게만 국한해서 배포한다거나 일반인에게는 유료로 판매할 수밖에 없는 것이 현실이다. 한편 Elsevier나 Wiley, Springer 같은 대형 상업 출판사들이 생겨나 학술지의 인쇄 및 배포 사업을 하게 됨에 따라, 학술지도 시장원리에 의해서 발간 및 배포되는 시대가 되었다.

이런 가운데 학술지 출판사들은 패키지 형태로 학술지 구독계약을 요구하고 있어서 구독기관들이 필요하지 않은 학술지까지 구독해야만 하는 상황도 벌어지고 있다. 그에 따라 학술지의 구독료가 급격히 상승함으로 말미암아 구독을 포기하는 경우도 생겨나고 있다.

한편, 국내 학술지는 아직까지 종이 중심의 발행 수준에 머물러 있고, 발간 및 출판 과정이 체계화되어 있지 않으며, 소규모 학회들의 경우 인쇄형에서 전자형으로 전환하는 방법에 대한 지식이 부족하여 전자저널 서비스에 많은 어려움이 있다. 이로 인해 우리나라의 과학기술 학술논문정보를 국제적으로 배포하고 가시성을 높이는데 있어 장애가 되어 국제적 인지도 및 활용도는 상대적으로 낮다.

지식 경쟁력은 단순히 양적 규모가 아니라 질적으

로 영향력이 높은 지식을 얼마나 많이 가지고 있느냐에 달려 있다. 이를 위해서는 국내 과학기술 학술논문의 효율적인 발간 체계를 구축함과 아울러, 국제적 접근성 향상과 경쟁력을 강화하기 위한 국제 표준을 준수한 전자저널 구축 방안을 마련하고 추진해야 할 것이다.

2. 국내외 전자저널 구축 동향

2.1 전자저널의 필요성

인터넷의 확산과 함께 연구자들이 물리적 공간의 제약을 받지 않고 보다 쉽고 빠르게 정보에 접근하기를 원하면서 전자저널은 가장 보편적인 정보획득 수단으로 자리매김하고 있다. 이에 따라 도서관들은 전자저널의 구입 비중을 늘리고 있다.

한편 Web of Science¹⁾, SCOPUS²⁾와 같은 국제색인 데이터베이스의 등재 평가기준에 전자저널에 대한 평가가 거의 필수 항목으로 포함되게 되었다. 전문가들의 분석에 따르면, SCIE 등재평가 시 별도의 도메인을 갖는 전자저널이 있는지 여부가 포함된다고 한다. SCOPUS의 경우에는 최근 학술지 평가에서 인쇄본 대신에 PDF를 보내고 학술지 홈페이지 주소를 요청하고 있다.

이 밖에 최근 상업출판사는 대부분의 학술지를 전자저널 형태로 제작하고 있으며, 전자저널의 발간 형식은 XML이 표준으로 자리잡아가고 있다. 국제색인 데이터베이스 가운데 XML로 자료를 만들어 보내는 것을 요청하는 곳으로는 PubMed Central, BioMed Central, Oxford University Press, MedKnowledge, Hindawi, Bentham, Frontiers, PLoS 등이 있다. 1996년 XML이 W3C에 의해서 차세대 웹의 표준 언어로 받아들여진

1) Thomson-Reuters에서 운영하는 자연과학, 사회과학 및 인문예술학 Citation 및 참고문헌정보 데이터베이스로 학술지 10,000 여종을 대상으로 한다.

2) Elsevier에서 운영하는 Citation 및 참고 문헌 데이터베이스로 14,000 종의 학술지를 다루고 있다.

표 1 국내 학술지 원문서비스 현황

구분	한국연구재단 (2009년8월 현재)	KISTI (2008년12월 현재)	KERIS (2009년12월 현재)	상업 DB (2009년9월 현재)	합계 (중복제외)
학술지종수	1,729종	640종	544종	2,272종	3,438종

이후 2000년에 등장한 PMC XML은 이제 전자학술지의 표준으로 자리매김하고 있다. 따라서 이제 학술지는 PMC XML 형식으로 만드는 것이 필요하다.

또한, 학술논문이 전자화됨에 따라 개별 학술논문마다 전자 식별자를 붙이는 것이 확산되고 있다. 그 대표적인 것이 DOI(Digital Object Identifier)인데, 이는 디지털 객체에 부여되는 유일한 수문자열 스트링을 의미한다. 학술 출판사는 CrossRef³⁾에 메타데이터와 해당 서비스 URL을 등록하고 부여 받는 전자원문 서비스 위치에 대한 식별자를 통해, 서비스 URL이 바뀌어도 전자원문으로의 직접적인 링크가 끊어지지 않게 된다.

2.2 국내 학술지의 전자화 현황

우리나라에서 발행하는 학술지는 총 3,438종으로 추산된다. 이는 한국연구재단 등재지 및 등재후보지, KISTI의 학회마을과 KERIS의 RISS4U를 통해 원문 서비스되고 있는 학술지, 상업적 DB 회사에서 원문서비스가 되고 있는 학술지 중 중복을 제외한 숫자이다(표 1 참조).

전체 학술지 중 학술지 발행기관의 홈페이지에서 원문서비스를 하고 있는 학술지는 48.3%인 1,681종이다. 원문서비스가 되는 학술지의 원문파일 유형은 PDF나 TIF 형식의 이미지 파일이 주를 이루었다. 의학분야 외에는 XML로 학술지를 구축하는 학술지는 거의 없는 실정이다.

다음으로 우리나라 학술지 중에는 DOI와 같은 전자 식별자를 부여하고 있는 학술지가 많지 않다. 기본적으로 DOI가 무엇인지에 대한 이해가 없거나 이해한다고 하여도 학회 스스로가 그것을 추진할 여력이 없는 경우가 많은 상태이다. 최근 대한의학학술지편집인협의회와 한국과학기술정보연구원이 주도하여 국내 학술지에 DOI를 부여하기 시작했다.

2.3 오픈액세스(OA) 운동과 전자저널

전자저널이 보편화되면서 학술지 구독의 새로운 패턴으로 컨소시엄 형태의 공동구매가 가능하게 됨에 따

라 이전에 없었던 새로운 문제가 제기되고 있다. 저널을 패키지형태로 구매(번들 계약)해야 하기 때문에 원하는 저널이 포함된 패키지를 구매하기 위해서 원하지 않는 더 많은 저널을 구매해야 하는 문제가 발생하고 있다. 가장 큰 문제점은 구매자인 도서관들이 전자저널을 소장할 수 있는 권한도 재복사 서비스(reprint) 권한도 없기 때문에 지속적인 이용을 위해서는 매년 재계약을 해야만 한다는 점이다.

에초에 학술적 발견의 공유를 위해 등장한 학술지가 이제는 시장의 상품처럼 유통되게 됨에 따라 학술적 발견에 대한 정보의 공유가 제한되고 있는 것이다. 이를 해결하기 위해서는 학술지 출판 및 유통방식에 변화를 주어야 한다는 지적이 나오고 있다. 그 중에서 가장 대표적인 것이 학술지 오픈액세스(open access) 운동이다. 오픈액세스란 모든 학술논문이 누구에게나 무료로 공개되어야 한다는 개념으로 2000년대에 들어서면서부터 시작된 운동이다. 오픈액세스 학술지에서는 출판사와 저자가 논문의 발간 및 유통에 들어가는 비용을 지불하도록 되어 있기 때문에 구독자가 정보를 무료로 접근할 수 있는 것이다. 기본적으로 오픈액세스 학술지는 전자저널이어야 한다.

오픈액세스를 요구하는 선언들이 세계 각국에서 이어졌는데, 대표적인 것에는 Budapest Open Access Initiative(BOAI)와 SPARC(the Scholarly Publishing and Academic Resources Coalition)이 있다. BOAI에서는 모든 학문분야에서의 연구논문이 인터넷 상에서 공개 사용가능하도록 하는 원리, 전략, 코멘트를 진술하였고, 기관별 또는 학문분야별 레포지토리에서의 셀프 아카이빙(self-archiving)과 오픈액세스 저널의 두 가지 상호보완적인 전략을 사용할 것을 권고하고 있다.

이러한 개념으로 운영되고 있는 사이트의 유형으로는 OAJ(open access journal)와 기관 또는 분야별 레포지토리(repository)가 있다. 대표적인 것으로는 물리학분야의 arXive, 미국의 NLM(National Library of Medicine)에서 운영하는 PubMed Central(PMC), BioMed Central(BMC), PLoS(Public Library of Science) 등이 있다.

2.3 전자저널 구축 주요 사례

2.3.1 미국 PMC(PubMed Central)

PMC는 미국 NIH(National Institute of Health)의 오

3) CrossRef는 전 세계적으로 책임감 있는 출판사들이 참여하고 재정을 분담하여 설립한 비영리 독립기관이며, 학술출판물에 대한 공식적인 DOI 링크 등록기관으로서 출판사간 인용에 대한 링크 시스템을 운영하고, 학술지 및 메타데이터의 기탁, 이용자별 서비스의 제공을 담당한다.

표 2 PMC 등재 기준

level		기준
Stage 1	Scientific Quality Check	Medline 등재여부, SCI/EMBASE 등 서지데이터베이스 등재여부, 편집장의 지명도 등
Stage 2	Technical Requirement	XML DTD tagging 여부
Stage 3	Agreement	OA 국제적 개념에 맞는 copyright, CCL 정의 여부

프엑세스 논문 국가기록보관소로서 NIH에서 지원하는 연구자금 내에 오픈엑세스를 위한 비용을 포함시킴으로써 연구자에게 위임하는 형태로 운영되고 있다. NIH에서 받은 연구자금으로 기고한 논문은 무조건 PMC을 통해 공개하도록 제도화되어 있다.

PMC에 등록되기 위한 절차는 표 2와 같이 크게 세 단계로 나뉜다. 여기에 논문을 기탁하기 위해서는 논문의 전문을 학술지 논문 DTD(Document Type Definition)에 따르는 XML이나 SGML(Standard Generalized Markup Language) 형식으로 올리도록 하고 있다. 이 과정이 모두 통과되면, PMC에 등록되게 된다.

2.3.2 일본 J-STAGE

일본과학기술진흥기구(JST: Japan Science and Technology Agency)는 1999년부터 일본 학회의 전자출판을 지원하고 자국의 학회 논문을 국제적으로 유통되도록 하기 위한 무료 플랫폼인 J-STAGE를 운영하고 있다. 현재 약 570개 학술지 26만 편 이상의 논문을 서비스 중이다. J-STAGE를 통해 제공되는 서비스는 표 3과 같다. 2010년부터 J-STAGE 3가 개발될 예정인데, 여기에는 XML 전자저널 생성기능이 포함될 예정이다.

일본화학회는 2002년부터 원문데이터를 J-SATGE에서 서비스하고 있으며, 그 전까지는 자체적으로 마크업 언어(markup language)를 이용해 원문데이터를 제작해왔었다.

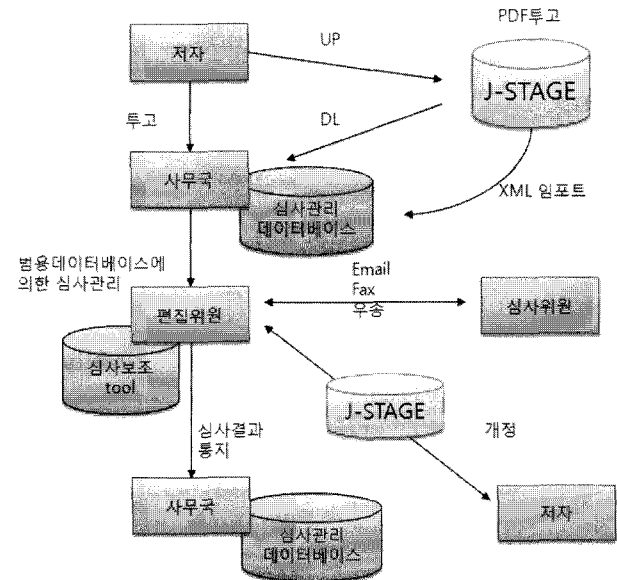


그림 1 일본화학회의 전자투고관리 사례

표 3 J-STAGE를 통해 제공되는 논문 투고 및 출판 상세 서비스 내용

기능	설명
편집/평가 표준 지원 시스템 논문 제출/접수 기능 편집/평가 지원 기능 절차 관리 기능	논문의 편집 및 평가를 지원하고 그와 관련된 절차를 관리하는 기능 온라인상에서의 논문 제출/접수 기능 편집자 및 평가위원 배정, 동료평가 결과 공지 기능 논문의 현재 상황 보여주기 기능
전자저널 출판 지원	FrameMaker+SGML 템플릿 지원 서지정보, 참고문헌 생성 툴 지원 각종 툴 정보 제공
전자저널 갱신 시스템 갱신지원 시스템 접근 기능 검색 기능 정기회원 인증 전자부속데이터 신규정보 관련아티클 연계 특성 접속통계	전자저널의 갱신정보 논문을 갱신하기 위한 업로드 데이터 업로드 기능 서지정보, 원문 정보에 접근하기 위한 기능 날짜, 제목, 저자명, 키워드, 원문으로부터 논문검색기능 회원내 대한 IP나 패스워드로 인증 전자점부데이터 접근 기능 새로운 논문에 대한 정보를 추가시키기 위한 기능 오기, 취소, 주석 등과 같은 논문내에 관련된 정보 연계 학회 접근 통계
참고문헌링킹생성기능 참고문헌연계 인용문헌연계	참고문헌정보를 기반으로 자동으로 링킹생성 인용된 참고문헌을 기반으로 참고문헌 연계 인용문헌정보와 참고문헌 연계

2.3.3 대한의학학술지편집인협의회

우리나라의 학술논문의 국제화는 대한의학회 산하 대한의학학술지편집인협의회(의편협; KAMJE)가 1997년에 미국의 PubMed와 같은 형태의 KoreaMed를 구축한 것이 효시라고 할 수 있다. 2010년 현재 163종의 국내 의학분야 학술지에 수록된 약 17만 건의 학술논문은 영어로 서비스하고 있으며, 2008년부터는 PubMed Central에 대응되는 Synapse를 구축하여 서비스하고 있다. 의편협은 정보 서비스 사업 외에 자체 개발한 인용지수(KOMCI)를 이용한 학술지의 정기적인 평가, 편집인 교육, 연구출판윤리 확립 등의 학술지 품질 향상 사업도 병행하고 있다.

2.3.4 한국과학기술정보연구원

과학기술분야의 경우 한국과학기술정보연구원(KISTI)에서는 2007년 12월부터 KoreaScience라는 사이트를 오픈하여 2010년 현재 260여종의 국내 과학기술분야 학술지에 게재된 논문 20만 편 이상을 서비스하고 있다. KoreaScience는 현재 WorldWideScience.org에 연계되어 전 세계적으로 서비스되고 있으며, 2008년부터는 의편협과 KISTI가 CrossRef의 Co-Sponsoring Member로 가입하여 각각 의학 분야와 과학기술분야의 학술지 DOI 기탁사업을 펼치고 있다.

3. XML 전자저널 구축 프레임워크

3.1 XML 전자저널 요소 기술

3.1.1 XML

전자적 문서를 표현하는 표준 언어는 1986년 SGML(Standard Generalized Markup Language)을 대표로 시작하여 최근 XML(eXtensible Markup Language)까지 많은 발전을 이루었다.

SGML은 1986년, 국제표준화기구(ISO)에서 정한 국제표준 언어로서 다양하고 복잡한 문서에 대하여 무제한적인 방법으로 표현이 가능하지만 인터넷 웹브라우저에서는 정상적인 문서의 조회가 불가능하다. 무제한적이고 다양한 표현을 정확하게 해석할 수 있는 웹브라우저의 개발이 불가능하기 때문에 전용 뷰어를 사용해야만 조회가 가능하다.

HTML(Hyper Text Markup Language)은 1992년, SGML을 모태로 하여 만든 인터넷 전용 언어로 인터넷에서

의 처리기술이 우수하고 웹브라우저에서 쉽게 조회할 수 있으나, 홈페이지 등에서 자주 사용되는 제한적인 표현방법만 채택하였기 때문에 학술논문같이 복잡한 문서 등의 원활한 작성 및 사용이 용이하지 않다.

1996년 W3C에서 제안한 XML은 기존의 마크업 언어인 SGML과 HTML의 한계를 극복하여 문서를 작성하는 사람이 직접 태그를 정의하여 사용할 수 있도록 한 확장된 언어이다. HTML의 경우 웹상에서의 표현을 중점으로 둔다면, XML은 정보 전달을 중점을 두고 있다고 할 수 있으며, 문서나 자료를 구조화하기 위해 사용된다.

문서의 구조 및 내용은 XML 태그로 표현되며, XML 문서의 경우 웹을 통해 전달받은 정보는 응용 프로그램을 통해 쉽고 정확하게 정보를 처리할 수 있도록 데이터를 조직화, 구조화시켜 다양한 분야에서 응용할 수 있어 그 효용성이 아주 높다. 이런 이유로 최근 국제 상업출판사는 대부분 XML 형태로 학술지를 제작하고 있다.

3.1.2 XML DTD

DTD(Document Type Definition)는 XML 문서를 표현함에 있어서 XML 문서의 태그를 정의하는데 필요한 작성 규칙이다. W3C에서 권고한 문법대로 작성된 XML 문서를 잘 짜여진(well-formed) 문서라고 하며, DTD 형식에 맞게 짜여진 문서를 유효한(valid) 문서라고 한다. 이렇듯 문서가 잘 짜여졌는지 혹은 유효한지를 검사하기 위해 Validator를 활용하게 되는데 바로 XML DTD에 정의된 룰에 따라 판단한다.

XML문서에 가장 첫줄에는 <?xml...>이라는 위와 같은 코드가 와야 하며, 그 다음에 바로 해당 XML 문서가 따르는 DTD의 경로를 지정해주게 된다. "PUBLIC"은 국제적으로 공인된 단체에서 공개한 DTD이며, "SYSTEM"은 특정 기관에서 사용하기 위해 공개한 DTD라는 것을 의미한다. 그 다음에 나오는 "-"는 ISO와 같은 국제공인단체일 경우 사용하며, "+"는 그 외일 경우 사용한다.

확장 가능한 XML의 경우, 다양한 형태로 나타날 수 있기 때문에 표현에 대한 정의를 문서로 나타내어 정확히 할 필요가 있다. 새로운 마크업 언어를 개발하면 그에 따른 응용프로그램 역시 개발되어야 하는데 문

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE article PUBLIC "-//NLM//DTD Journal Publishing DTD v2.3 20070202//EN"
"http://dtd.nlm.nih.gov/publishing/3.0/journalpublishing3.dtd">
```

그림 2 XML 문서 내에서 DTD 경로 지정 예

서를 처리해야 하는 프로그램의 경우 마크업 언어로 작성된 문서를 처리하기 위해서 DTD에 정의된 구조를 이해하도록 프로그래밍 되어야 한다.

3.1.3 XSL

XML 자체는 단순히 응용프로그램 간 또는 사람과 응용프로그램 간에 데이터 전달용으로 사용된다. 이것은 XML 문서가 응용 프로그램이 바로 사용할 수 있는 적합한 형태를 가지고 있다는 것과 다른 의미이다. 따라서 각각의 응용프로그램은 외부로부터 받은 XML 문서를 내부적으로 사용되는 포맷으로 변환하는 과정이 필요한데 이에 대해서 W3C에서 XML 문서를 다양한 포맷으로 변환시키는 변환규칙을 작성할 수 있도록 개발한 언어가 바로 XSL(eXtensible Stylesheet Language)이다. 즉, 논리적 구조만 가지고 있는 XML instance(문서)를 외부로 보여지게 하기 위해서 사용하는 포맷팅 언어로 고도로 구조화된 정보를 포맷팅하기 위해 필요한 강력한 포맷팅 능력을 제공한다.

3.1.4 XSLT

XSLT(eXtensible Stylesheet Language Transformation)는 XML을 또 다른 구조를 갖는 XML, HTML, TEXT로 변환하기 위해 사용하는 기술이다. 따라서 HTML로 변환을 하고 싶다면 반드시 XSLT를 활용해야 한다. XSLT는 XSL의 일부로 사용할 수도 있고, 하나의 객체로서 사용될 수도 있다. 여기서 중요한 점은 XML syntax를 따르는 기술이어야 한다. 이것은 SVG, Voice ML, WML 등으로 변환할 때 사용가능하다는 의미이기도 하다. 예를 들어, 하나의 주소록을 XML로 만들어서 파일이나 DB 형태로 저장하고 있다고 가정했을 때, XSLT를 이용하여 HTML 문서로 게시할 수도 있고, WML로 변환하여 핸드폰 문자서비스도 보낼 수 있으며, VoiceML로 변환해서 음성으로 전달할 수도 있고, SVG로 변환해서 그래픽으로 만들 수 있는 것이다.

3.2 전자저널 구조 분석 : NLM DTD를 중심으로

3.2.1 NLM DTD의 개요

NLM의 NCBI(National Center for Biotechnology In-

formation)에서 저널 콘텐츠의 정보교환을 목적으로 만든 NLM DTD는 다양한 요소(element)와 속성(attribute)을 정의한 XML Schema module을 제공하며, 메타데이터뿐만 아니라 전문(full-text) 기술도 가능하여 저널 논문 이외에도 사설, 단행본, 리뷰 등과 같은 데이터의 텍스트, 그래픽 콘텐츠 기술까지 가능하다. 현재 PMC XML DTD는 버전 3.0까지 발전하였으며, 버전이 올라갈 때 마다 포함하는 tag 수가 늘어서 현재 154개의 tag를 사용하고 있다.

NLM DTD의 특징은 다음과 같다.

- XML Schema module set으로 구성되어 있어 물리적 파일로 분리가 가능하며, 다양한 schema와의 조합도 가능하다.
- 저널 아티클 정보의 교환뿐 아니라 문서 저장, 생성, 출판을 위한 schema 구성도 가능하다.
- 파라미터 엔티티를 정의하여 DTD의 의도에 맞게 수정하여 재정의가 가능한 포용성 있는 데이터 모델이다.

NLM DTD는 사용용도에 따라 크게 저널을 포함한 모든 문헌을 보존하고 교환이 가능하도록 설계된 Archiving and Interchange Tag Set, 저널 기사의 콘텐츠 및 구조와 퍼블리싱에 최적화되어 있는 Journal Publishing Tag Set, 논문 편집을 위해 설계된 Article Authoring Tag Set, NCBI에서 발행하는 단행본을 기술하기 위해 설계된 NCBI Book Tag Set으로 구별된다. 본고에서는 전자저널의 구축뿐만 아니라 추후 웹 퍼블리싱까지 고려하여 저널의 출판과 아카이브를 위한 호환을 제공하는 Journal Publishing Tag Set을 채택하였다.

3.2.2 NLM DTD 3.0의 구조

Journal Publishing Tag Set(이하, NLM DTD라고 칭함)는 크게 XML 선언부와 최상위 요소인 Article element로 구성되어 있다. XML 선언부는 예를 들어, 그림 3처럼 article 태그를 통해 해당 XML 문서의 유형과 언어구분, DTD 버전, 참조 subschema set을 정의한다.

최상위 요소인 <article> 요소는 논문의 메타정보를 포함하는 <front>, 본문정보를 기술하는 <body>, 논문

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE article PUBLIC "-//NLM//DTD Journal Publishing DTD v2.3 20070202//EN"
"http://dtd.nlm.nih.gov/publishing/3.0/journalpublishing3.dtd">
<article article-type="research-article"
dtd-version="3.0" xml:lang="en"
xmlns:mml="http://www.w3.org/1998/Math/MathML"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
```

그림 3 NLM DTD 선언부

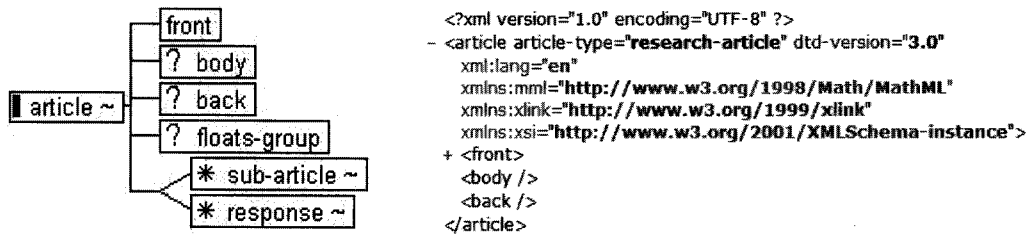


그림 4 Article element 구조

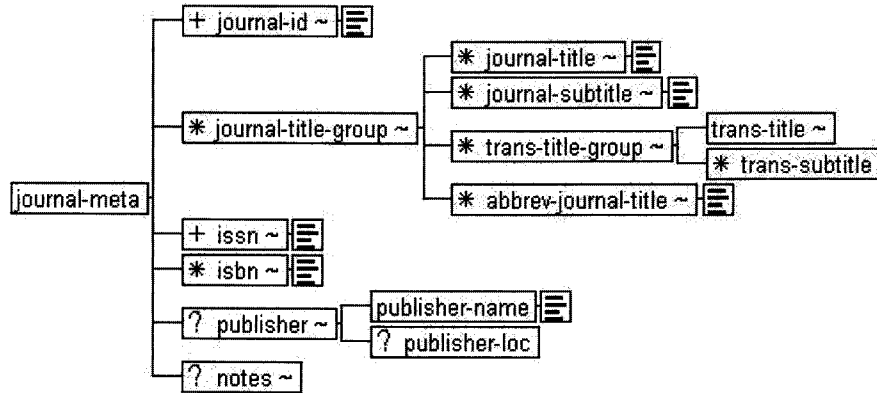


그림 5 journal-meta element 구조

배경 자료를 기술하는 <back>, 논문의 하위 데이터 셋 (예. 화학구조, 그림, 표 등)을 기술하는 <float-group>, 논문 내에 포함된 다른 논문을 기술하는 <sub-article>, 주석을 기술하는 <response> 요소들로 구성되어 있다. <article> 요소의 구조와 시퀀스, 출현빈도, 이를 적용한 XML instance를 보면 그림 4와 같다.

2 레벨의 논문 메타정보에 해당되는 <front> 요소는 저널의 메타정보를 기술하는 <journal-meta>, 논문의 메타정보를 기술하는 <article-meta>, 주석 및 저자노트를 포함하는 <notes> 요소들로 구성되어 있다.

<journal-meta> 요소는 크게 저널의 식별값을 기술하는 <journal-id>, 제목을 기술하는 <journal-title-group> 외에 <issn>, <isbn>, <publisher>, <notes>로 구성되어 있다. <journal-meta>의 구조, 시퀀스, 출현빈도는 그림 5와 같다.

<article-meta> 요소는 논문의 메타정보를 기술하는 요소로서 하위 요소로 <title-group>, <contrib-group>, <pub-date>, <volume> 등으로 구성되어 있으며, 주요 요소들은 표 4와 같다.

논문의 본문을 기술하는 <body> 요소는 포함되는 텍스트 구성 및 표, 그림에 대한 정보를 기술할 수 있도록 구조화되어 있다. 콘텐츠는 주로 절 정보를 기술할 수 있는 <sec> 요소와 문단을 기술할 수 있는 <p> 요소를 사용하며, <fig>, <table-wrap> 등의 요소를 이용하여 표와 그림을 기술할 수 있다. 이뿐만 아니라

표 4 <article-meta> 주요 하위 요소

요소명	설명	출현빈도
<title-group>	기사 제목	1
<contrib-group>	저자 그룹	>=0
<pub-date>	출판 일자	>=1
<volume>	권	0 or 1
<issue>	호	0 or 1
<issue-title>	특정호에 부여된 제목	0 or 1
<supplement>	부속서	0 or 1
<fpage>	시작페이지	0 or 1
<lpage>	끝페이지	0 or 1
<permissions>	저작권 정보	0 or 1
<abstract>	초록정보	>=0
<trans-abstract>	다른언어로 기술된 초록정보	>=0
<kwd-group>	키워드 정보	>=0
<funding-group>	기금정보	>=0
<conference>	관련 컨퍼런스 정보	>=0
<counts>	그림, 표 등 서브셋 갯수	0 or 1

본문 내에 화학식, 동영상 등의 다양한 멀티미디어를 포함시킬 수 있도록 설계되어 있다. 그림 6은 <body> 요소의 구조와 출현빈도 및 XML instance 예제를 보여준다.

논문에는 본문뿐만 아니라 참고문헌, 감사의 글, 부록 등의 부속정보들을 포함하고 있다. 이들을 기술하기 위해서는 <back> 요소를 사용한다. 따라서 <back>



그림 6 <body>요소의 구조(좌)와 XML instance의 예(우)

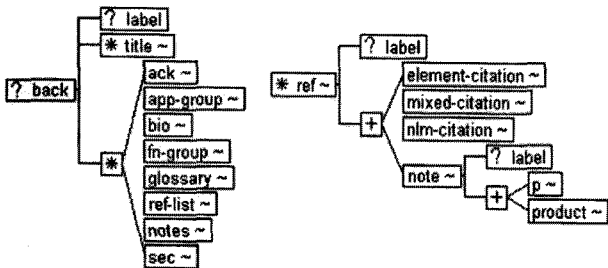


그림 7 <back>의 구조(좌)와 <ref>의 구조(우)

요소에는 감사의 글을 표현하는 <ack>, 부록정보를 표현하는 <app-group>, 저자의 약력을 소개하는 <bio>, 참고문헌을 기술하는 <ref-list> 등으로 구성되어 있다.

NLM DTD 3.0은 자원 유형별 상위 DTD 구성 요소의 모듈화를 위해 요소와 속성 집합 및 콘텐츠 모델을 기본적으로 파라미터 엔티티로 그룹화하여 기술하였고, 대부분의 요소는 개념 단위별로 그룹화하여 독립된 모듈로 정의하고 있다.

3.2.3 XML 전자저널 설계 규칙

전자저널 설계의 기본원칙은 W3C 권고안에 근거하여 관련 국제표준 예를 들어, XML, UTF-8, ISO/IEC 11179 part 5 for Naming, NLM DTD 등을 적극 수용하고, 무엇보다 전자저널의 범용성 확보를 전제로 한 국제표준을 철저히 준수하는 것이다.

XML 데이터의 엄밀성을 유지하기 위해 5가지 사항을 반드시 준수해야 한다. 첫째, 데이터를 명확하게 정의하고 둘째, 데이터 속성을 엄격하게 정의하며 셋째, XML에서의 선택사항을 최소로 유지한다. 그리고 넷째, 적용원칙을 구체적으로 기술하며 마지막으로, 표준화된 원칙을 체계적으로 정립 및 적용 관리한다.

또한, XML 데이터의 간결성/명료성을 유지하기 위해 4가지 사항을 반드시 준수해야 한다. 첫째, 복잡성을 회피하고, 둘째, 유지 및 변경 관리의 용이성을 확보한다. 셋째, 일반적인 어휘를 사용하며, 끝으로 약어나 축약어 사용을 절제하며, 예외사항에 대해서는 철저히 관리한다.

XML 디자인 룰은 헤더정보, 국가코드, 버전정보, 주석방법 등의 기본정보에 대한 원칙을 명시한다. XML 문서의 문자셋은 국외 사용자나 고어 및 특수문자 처리를 유연하게 하기 위해 유니코드 UTF-8을 사용해야 하며, W3C 스펙 또는 제안된 권고안에 근거해야 한다. 언어코드는 ISO639-1 표준, 일자와 시간은 ISO 8601표준을 반드시 준수해야 한다. 뿐만 아니라 표준화된 대소문자 협약을 준수하고 축약어와 약어 사용을 최대한 절제한다.

XML에 직접 사용해서는 안 되는 예약문자는 '&', '<', '>', '"', '''가 있다. 이는 아래와 같이 문자열 변

표 5 전자저널 XML 데이터 디자인 룰

Rule Description	status
XML의 encoding은 "utf-8"	MUST
언어구분을 위한 코드는 ISO639-1을 준수 (예, 한국어:KO, 영어:EN 등)	MUST
일자와 시간의 표준은 ISO8601을 준수 일자: YYYYMMDD(basic), YYYY-MM-DD 시간: hhmmss 또는 hh:mm:ss	OPTIONAL
임의적으로 데이터 확장 시 어트리뷰트 사용	RECOMMENDED
XML에서 예약문자를 포함한 특수문자는 입력방법을 준수	MUST
문서 유통시 XML 문서 내에 Doctype은 명시하지 않음 DTD를 이용한 유효성 검증 시 Pubmed validator를 이용	RECOMMENDED
Field 길이 제한은 정의하지 않음.	RECOMMENDED

표 6 약어 및 축약어 처리

Rule Description	status
일반적으로 사용되고 잘 알려진 경우 그대로 사용	RECOMMENDED
제안하는 약어나 축약어가 어플리케이션의 일차적인 이해당사자들이 이해할 수 있는 것임을 프로그래머가 인정한 경우	RECOMMENDED
출처가 신빙성이 없거나 확인 불가능한 경우 또는 해석에 있어 오류의 여지가 있을 경우 사용되어서는 안 됨	MUST NOT
Tag, element, attribute, type명에 축약어 사용 절제	RECOMMENDED NOT
메타데이터의 약어 및 축약어는 대문자로 사용	RECOMMENDED

환 과정을 통하여 저장해 주어서 처리한다. 만약 문자 열 변환 과정을 거치지 않으면 파서가 정상적으로 인식할 수 없게 된다.

과학·공학·수학 등의 분야에서 아주 작은 숫자나 아주 큰 숫자는 지수표기법으로 표기되는 경우가 많다. 이러한 지수 표기법의 숫자와 일반 숫자는 서로 구분할 수 있도록 표기할 필요가 있다. 또한 지수 표기법 중 일부(μ s, μ g 등의 단위)는 스펠링을 완전하게 표기하지 않으면 그 의미가 상실되게 된다.

표 7 XML 예약문자 및 처리방법

특수문자	입력	출력
&(ampersand)	&	&
<(less-than sign)	<	<
>(greater-than sign)	>	>
"(quotation mark)	"	"
'(apostrophe)	'	'

표 8 scientific notation 처리방법

특수문자	입력	출력
μ s	mus	microsecond
μ g	mug	microgram
°		degree
°C		degree C

표 9 특수문자 처리방법

특수문자	입력	출력
α	&agr;	alpha
β	&bgr;	beta
γ	&ggr;	gamma
ϵ	&egr;	epsilon
δ	&dgr;	delta
→	-->	-->
±	±	+/-
m^2	m²	m^2
10^{-6}	10(-6)	10^{-6}
10^6	10<inf>(6)</inf>	10^6
height ^{2,239}	height^{2,239}	height(2,239)
H ₂ O	H ₂ O	H ₂ O

XML 예약문자 이외에 그리스 문자, 위첨자, 아래 첨자와 같은 특수문자들이 있다. 표 9는 일반적으로 많이 사용되는 특수문자의 처리 예를 보여준 표이다.

3.3 전자저널 구축 프로세스

전자저널 구축 절차는 논문투고부터 발행까지의 절차를 살펴볼 필요가 있으며, 학술지에 따라 많은 차이가 있지만 본 고에서는 논문투고 및 전자저널 출판시스템의 공정 및 작업흐름은 크게 투고 → (심사의뢰)

→ 심사 → 승인 → 수정 및 편집 → 발행으로 구분하였다. 또한 웹 기반의 논문투고시스템을 통하여 생산한 논문을 전자저널 데이터베이스화하고 이를 전자출판 및 정보서비스까지 연동한 시스템을 고려하였다. 이는 투고자가 제출한 원문이나 이미 발행된 전자원문의 포맷을 변환하거나 직접 입력하는 방식의 이전 전자저널 구축 프로세스와는 달리 논문 투고 시점부터 DB화할 수 있는 체계가 될 수 있다. 즉, 그림 8과 같이 기존의 전자논문을 XML로 작성할 수 있는 워크벤치(work bench)를 모듈화 하여 학술지 발행 프로세스 어느 단계에서도 사용이 가능하도록 하는 방법으로 투고, 발행 시점에서 모듈을 장착하여 사용할 수 있게 하는 것이다. 이는 기존 시스템의 변경을 최소화하면서 XML DB를 구축할 수 있는 최적화방안이라고 볼 수 있다.

DB에 구축된 전자저널은 XSL을 통하여 전자저널 서비스의 스타일을 결정하고, 학술지 한 종에 대한 e-ISSN을 부여받아야만 비로소 학술지의 전자출판을 완성할 수 있다. 그림 8은 XML 기반의 전자 논문 변환 방식을 나타낸 것이며, 워크벤치는 XML DTD의 데이터 정의 룰에 따라 XML 파일을 손쉽게 생성할 수 있도록 하는 도구 역할을 하게 된다. 워크벤치를 통해 만들어진 XML 파일은 DB에 저장되어 서비스하게 되는 것이다. XML로 만들어진 논문은 구문, 표, 그래픽 등의 모든 객체를 추출하여 특성에 따른 다양한 변환과 상호 참조 등의 동적인 구조로 활용될 수 있다. 이외에도 목차, 색인을 자동생성하거나 본문, 참고문헌 등으로 별도 관리하는 것이 가능하며, 다양한 방식을 충족할 수 있는 전자파일 객체모델을 가질 수 있다.

4. XML 전자저널 구축 방안

4.1 전자저널 구축 방안

현재 우리나라 학술지 원문의 온라인 서비스 90%

정도로 매우 높은 것으로 조사되었지만, 그 면면을 살펴보면 대부분이 학회 홈페이지 내의 하부 메뉴로 되어 있거나, 정보서비스 사이트에서 통합적으로 서비스되는 형태가 대부분이다. 이는 개별 학술지마다 독립된 도메인을 갖도록 요구하고 있는 국제색인 DB의 등재기준과는 거리가 먼 것이며, 더구나 XML로 전자저널을 구축하고 있는 이공분야 학술지는 거의 없는 상황이다.

따라서 국내 학술지의 전자저널 구축은 기본적으로 독립된 도메인을 통해 최소한 논문 메타정보와 pdf 원문을 제공할 수 있어야 할 것이다. 그런 다음 e-ISSN을 부여하고, 여력이 되면 DOI 부착과 XML 전자저널 형태로 발전시켜 나가야 할 것이다.

의생명 분야의 학술지는 PubMed Central에 등재하기 위해서 XML 전자저널 구축이 필수적이지만, 과학기술 분야의 경우는 필수적인 것은 아니다. 하지만, 전세계적으로 XML 전자저널이 보편화되고 국제색인 등재 기관에서 요구하는 사항이므로 학술지의 국제화를 원하는 학술지의 경우는 과학기술 학술지일지라도 XML 전자저널 구축을 서둘러야 할 것이다.

XML 전자저널 구축 프로세스는 다음과 같은 단계를 따라 구축 및 서비스할 수 있다.

- ① 온라인 투고관리 시스템: 각 학회들은 논문 투고, 심사, 편집, 출판 프로세스를 자체적으로 운영하여 최종 원문 전자 파일과 서지 정보를 제출한다. 이 때 각 학회는 별도의 온라인 투고관리 시스템을 활용하여 수행할 수 있다.
- ② DOI 부여: CrossRef의 Co-Sponsoring 회원기관이 각 학회를 대신해서 DOI Prefix를 부여하며, 학회로부터 받은 원문 전자 파일과 서지 정보를 활용하여 랜딩 페이지를 구축한다. DOI 원문 랜딩 페이지는 개별 학술지에 구축할 수도 있지만, 보안 및 안정성을 고려하여 한 곳에 일괄적으로

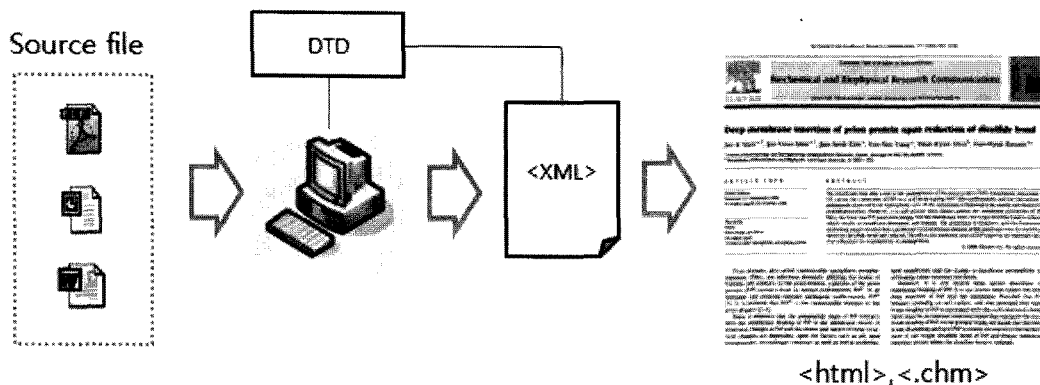


그림 8 XML 기반 전자저널 구축 프로세스

구축할 수 있다.

- ③ XML 변환: 학회로부터 받은 원문 전자 파일을 활용하여 XML 변환 작업을 수행한다.
- ④ 전자저널 사이트 구축: 학회가 지정한 도메인과 서버에 개별 학술지 전자저널 사이트를 구축하고, 국립중앙도서관에 e-ISSN을 신청한다.
- ⑤ XML 전자원문 포털 구축: 각 학술지의 XML 전자 원문을 통합하여 포털 사이트를 구축하여 국제적인 통합 검색이 이루어지도록 서비스 플랫폼을 구축한다.

4.2 XML 전자저널 구축 도구 활용 방안

현재 대부분의 학회에서는 XML 전자저널 구축을 고려하지 않고 인쇄 학술지 발간 후에 PDF 작성하여 서비스하는 방법을 채택하였기 때문에 XML 원문 작성이 쉽지 않은 형편이다. 이미지 형식의 PDF로 되어 있는 전자 원문의 경우는 더욱 어렵다. 이런 경우는 원문을 OCR로 읽어 들이거나 재입력을 해야 하는 번거로움이 있어서 XML 원문 전환은 거의 불가능하다. 다행히 텍스트 형식의 PDF인 경우는 수작업이나 전용 워크벤치를 이용해서 XML 원문 전환이 비교적 수월한 편이다. 이 경우도 문자 코드 문제, 그림 파일이나 수식 처리 문제 등이 난제로 남아 있다.

문자코드의 경우 출판 시 InDesign⁴⁾과 같은 범용 제품을 사용하게 된다면 손상 없이 XML 변환이 가능하다. 현재 국내에서는 전자저널 구축 기관에서 수작업 또는 자체적으로 만든 전용 워크 벤치를 활용하고 있으나, 해외에서는 eXtyle⁵⁾과 같은 범용 도구를 많이 사용하고 있다.

4.3 온라인 투고관리 시스템 활용방안

XML 원문 전환이 필요 없이 자동으로 XML 원문을 생성하기 위해서는 논문 투고 시점에서부터 XML로 투고를 하게 하는 방법이 있다. 그렇지만, 이 경우는 학술지 논문 투고자에게 많은 부담을 줄 수가 있

4) InDesign은 Adobe사가 온라인 출판과 오프라인 출판물의 경계를 허물 계획으로 개발한 인쇄 및 디지털 출판을 위한 전문 레이아웃 온라인 출판 소프트웨어이다. InDesign을 사용해서 출판을 하게 되면, 디자인 손상 없이 자동적이면서 인터랙티브하게 PDF 문서를 만들 수 있다.

5) eXtyle(Microsoft Word-based editorial and XML production software for publishers)는 Inera사가 개발한 마이크로소프트사의 Word 제품군용 플러그인 소프트웨어이며, 편집부서와 XML 웹페이지를 생산하는 부서들 사이에서 가교 역할을 수행한다. 주로 출판자 전용의 소프트웨어로 임의의 저자가 Word 파일을 이용하여 작업을 자동적으로 하거나 수동적으로 Word에서 작업을 수행할 수 있도록 디자인 되어 있다.

표 10 국내에서 사용 중인 주요 온라인 투고관리 시스템

시스템 또는 제공기관	사용 분야 또는 학회
KISTI-ACOMS	과학기술분야 100개 이상 학회가 사용
JAMS	연구재단등재 학술지 일부에 적용
(주)온피씨에스	기계, 금속, 전기, 전자 분야 약 10개 학회가 사용
(주)네티엔	이공 분야 약 10개 학회가 사용
(주)엠투커뮤니티	의학 분야 학회에 보급됨
인포랑	의학 분야 약 100개 학회가 사용
ScholarOne	엘스비어계열 학술지에 사용
OJS	일부 학회 사용
Manuscript Central	일부 학회 사용
J-STAGE	한일 공동 저널에 적용

기 때문에 사용하기 편리한 XML 전용 에디터를 개발하였을 때만 가능할 것이다.

현재 국내에서는 매우 다양한 형태의 온라인 투고 관리 시스템이 사용되고 있다. 이메일, 자체 개발 시스템, 해외 출판사 제공 시스템, 국내외 상용 시스템 등 다양하다. 현재 국내에서 활용되고 있는 온라인 투고 관리시스템의 현황을 표 10에 나타내었다.

이처럼 다양한 투고관리 시스템을 사용함으로써 말미암아 최종 논문 형식이 통일되지 못하고 있다. 결과적으로 XML 전환을 하기 위한 pdf 파일에 심각한 문제를 안고 있을 수가 있다. 따라서 XML 전자저널 구축 대상 학술지만이라도 가급적이면 동일한 파일 포맷의 전자 원문이 얻어지도록 표준화 작업이 필요하다.

5. 결론

앞서 살펴본 바와 같이 학술지의 전자화는 세계적인 추세이다. 학술지 출판사, 학회, 학술논문 서비스 기관 등이 모두 기존의 종이 매체에서 전자 매체 중심으로 급속히 전환하고 있다. 서구의 경우는 상업 출판사를 중심으로 전자저널 구축이 시작되었으나, 상업 출판사들의 가격 횡포에 대항하기 위한 공공 부문의 대응책으로서 전자화에 기반한 오픈 액세스 운동이 활발히 전개되고 있는 양상이다.

한편 상업 출판사가 열악한 한국, 일본, 중국 등 비영어권 국가에서는 자국의 학술지 국제화와 인지도 제고를 위해서 공공 부문이 나서서 학술지의 전자화를 추진하는 양상이다.

전자저널에서 중요한 것 중의 하나는 국제적인 표준을 따르는 것이다. 학술논문 서비스의 글로벌 표준으로 XML과 DOI가 확산되고 있다. 따라서 우리나라의 학술지도 국제 표준을 준수하여 XML 전문 구축과

DOI 식별자를 부착하여 서비스할 수 있는 체제를 갖추어야 할 것이다. 또한 국제색인 등재 기관들은 독립 도메인을 가진 전자저널 홈페이지 운영 여부도 중요 시하기 때문에 독립 도메인과 함께 e-ISSN도 부여받도록 하여야 한다.

참고문헌

[1] 김병규, 강무영, 최선희, 협회 전자저널 생성모델 연구 및 개발, 한국정보처리학회 제 27회 춘계학술대회, pp. 572-573, 2007.

[2] 노경란, 이해진, 2010, KISTI 지식리포트 「학술정보 센터의 새로운 서비스 모델: 오픈 액세스 출판」

[3] 이란주, 황신혜, “전자저널 효과적인 관리 및 이용자 서비스에 관한 연구: 대학도서관을 중심으로”, 정보관리학회지, 제20권, 제2호, pp. 135-156, 2003.

[4] 한종엽, “해양과학기술분야 전문학술지의 웹기반 논문투고심사시스템 및 전자저널 데이터베이스 연구”, 한국정보관리학회지, 제24권, 제1호, pp. 343-365, 2007.

[5] 과학기술정책연구원, “KoreaScience 글로벌 유통시스템 개발 및 구축 타당성 분석 연구”, 한국과학기술정보연구원, 2009.

[6] 서태설, “국내 지식정보의 글로벌 서비스 전략”, 한국비블리아학회 춘계학술발표회 자료집, 2009.

[7] 김성진, 정은경, 한민혜, “전자저널 컨소시엄을 둘러싼 학술커뮤니케이션의 쟁점과 대응동향”, 정보관리연구, 제39권, 제1호, pp. 27-52, 2008.

[8] 林和弘, “논문의 전자저널을 둘러싼 최근의 동향”, 科學技術動向, 7월호, 2009년.

[9] 차미경, 이나니, 김세주, “오픈 액세스 기반 학술 커뮤니케이션을 위한 정책 개발에 관한 연구”, 한국교육학술정보원, 2007.

[10] 한국정보관리학회, “기초학문자료센터 오픈액세스 시스템 구축 및 운영 방안 연구”, 한국학술진흥재단, 2007.

[11] NLM Journal Archiving and Interchange Tag Suite, [cited 2010. 3. 2]. <http://dtd.nlm.nih.gov/#id49712>

[12] PubMed Central. 2009. [cited 2010. 3. 12]. <http://www.ncbi.nlm.nih.gov/pmc/>

[13] M. Rahman and N. A. Mahmud, “Towards Developing a Theory of Knowledge Globalization”, Journal of Knowledge Globalization, Vol. 1, No. 1, pp.1-8, 2008.

[14] A. Creed and A. Zutshi, “The Wellhouse of Knowledge Globalization: IT and Virtual Community”, Journal of Knowledge Globalization, Vol. 1, No. 1, pp.29-42, 2008.

[15] John Willinsky, “The Publisher’s Pushback against NIH’s Public Access and Scholarly Publishing Sustainability”, PLoS Biology, Vol. 7, No. 1, pp. 20-22, 2009.

[16] Schonfeld, R. C., D. W. King, A. Okerson, and E. Fenton. 2004. The Nonsubscription Side of Periodicals: Changes in Library Operations and Costs between Print and Electronic Formats. <<http://www.clir.org/pubs/abstract/pub127abst.html>>.

[17] Willinsky, John “The Publisher’s Pushback against NIH’s Public Access and Scholarly Publishing Sustainability”, PLoS Biology, Vol. 7, No. 1, pp.20-22, 2009.

약력



서태설

1984 연세대학교 기계공학과 학사
 1986 한국과학기술원 기계공학과 석사
 2004 한국과학기술원 자동화/설계공학과 박사
 2006 영국 Cardiff Univ. MEC 방문연구원
 2010~ 과학기술연합대학원대학(UST) 조교수
 1986~현재 한국과학기술정보연구원 책임연구원

관심분야: 전자저널, XML, 메타데이터, 온톨로지
 E-mail : tsseo@kisti.re.kr



이혜진

2000 상명대학교 문헌정보학 학사
 2002 숙명여자대학교 문헌정보학 석사
 2004~현재 한국과학기술정보연구원 선임연구원
 관심분야: 전자저널출판, XML, 메타데이터, 오픈 액세스

Email : hyejin@kisti.re.kr