

XML 기반 학술지 전자출판 플랫폼(KPubS) 개발에 관한 연구

서태설*, 조성남*, 박선아*
*한국과학기술정보연구원
e-mail:tsseo@kisti.re.kr

A Study on Development of XML Journal Publishing Platform: KPubS

Tae-Sul Seo*, Sung-Nam Cho*, Sun-A Park*
*Korea Institute of Science and Technology Information

요 약

학술지는 책자 형태에서 전자 파일 형태로 급속히 변화하고 있다. 또한, 최근 전자 저널의 원문은 PDF에서 XML로 바뀌는 추세에 있다. 학술지 원문 XML은 JATS(Journal Article Tag Suite)라는 ANSI/NISO 표준을 따라 만드는 것이 보편화 되고 있다. 본 논문에서는 우리나라 학술지를 XML로 구축하고 국제적으로 서비스하기 위해서 만들어진 전자저널 출판 모델을 설계하고, 이를 개발·구현한 내용을 소개한다. 개발된 전주기 학술지 출판 플랫폼(KPubS)은 리눅스 기반 스프링 프레임워크와 MVC 패턴 및 시큐어 코딩 기법을 적용하여 개발하였으며, 반응형 웹 기술을 적용하였다.

1. 서론

최근 학술지는 책자 형태에서 전자 파일의 형태로 급속히 변화하고 있다. 전자저널은 PDF 원문을 근간으로 하는 학술지가 대세이나, 최근에는 원문을 XML DTD(eXtensible Markup Language Document Type Definition)로 하는 전자저널이 많아지고 있다. 따라서, XML 원문(full-text)을 구축하고 서비스하기 위한 기반 기술 개발이 시급하다[1].

학술지 XML DTD는 미국의 국립의학도서관(National Library of Medicine)에서 제안하여 의생명 분야 학술지 데이터베이스인 PubMed Central에 적용하면서 보편화되었다. 이것을 PMC XML DTD라고 하는데, 미국에서는 이것을 JATS라는 이름의 ANSI/NISO 표준으로 제정하였다[2].

우리나라에서도 의생명분야를 중심으로 XML 원문 구축이 활발하다. 현재 KoreaMed Synapse라는 사이트에서 약 140종의 의생명분야 학술지가 XML로 구축되어 서비스되고 있으나[3], 그 이외의 분야의 논문에 대한 XML 원문 구축과 보다 개선된 서비스가 필요한 실정이다.

본 논문에서는 XML 기반 국내 과학기술분야 학술지를 위한 전주기적인 학술지 출판 모델을 제안하고, 이를 반영하여 전자출판을 하는 플랫폼인 KPubS를 스프링 프레임워크(Spring framework)를 적용하여 개발하였다.

2. 기존 연구

학술지 XML 관련 연구는 크게 JATS DTD 구축 사

례 소개와 PDF에서 JATS XML로 변환하는 내용으로 나누어진다. Tachkawa 등[4]은 일본컴퓨터화학회논문지의 XML 구축 및 웹 출판 사례를 소개하였고, Nakanishi 등[5]도 일본 학술지의 XML 구축 사례를 연구하였다. 하지만, 지금까지의 연구는 학술지 논문의 XML 구축이라는 관점에서만 다루었고, 논문 투고에서 유통까지 학술지의 전주기적 관점에서 다룬 연구는 아직까지 없다.

3. 학술지 출판 모델

학술지 출판 모델은 해외에서 운영되고 있는 대표적인 사례를 벤치마킹하여 국내 상황에 맞게 재구성하여 만들어졌다[6]. 일반적으로 전자저널은 투고, 심사, 게재 확정, 편집, 인쇄, DB 구축, 웹 서비스 등의 일련의 과정을 거쳐서 독자에게 전달되어진다. 본 모델에서는 투고부터 인쇄까지의 과정을 창출(Creation) 단계, DB 구축을 아카이빙(Archiving) 단계, 웹 서비스를 서비스(Service) 단계로 구분하였고, 마지막에 확산(Promotion) 단계를 추가하였다.

먼저, 창출 단계에서는 논문의 품질을 높이는 단계로서 논문투고심사시스템(peer-review system)과 논문 파일의 품질이 중요하다. 본 모델에서는 하나의 논문투고심사시스템을 정하지 않고 이러한 요구 사항을 만족시키는 시스템을 선택하여 쓸 수 있도록 한다.

다음으로 아카이빙 단계는 논문 데이터의 표준적 관리를 하는 핵심 단계로, 서비스 단계와 확산 단계의 작업 효율을 극대화하기 위한 작업인 DOI(Digital Object Identifier) 관련 작업과 XML 원문 구축 작업이 루어진다.

논문 데이터 아카이빙을 위해서 논문의 메타데이터(title, author, abstract, keyword, 서지 사항 등)가 PDF 원문과 함께 저장된다. DOI 기탁을 위해서는 CrossRef에서 요구하는 XML 포맷을 준수한다. XML은 JATS에 따라 구축하였다. JATS는 학술지 기사의 내용과 메타데이터를 기술하는데 사용되는 XML 요소와 속성 용어들을 정의한다.

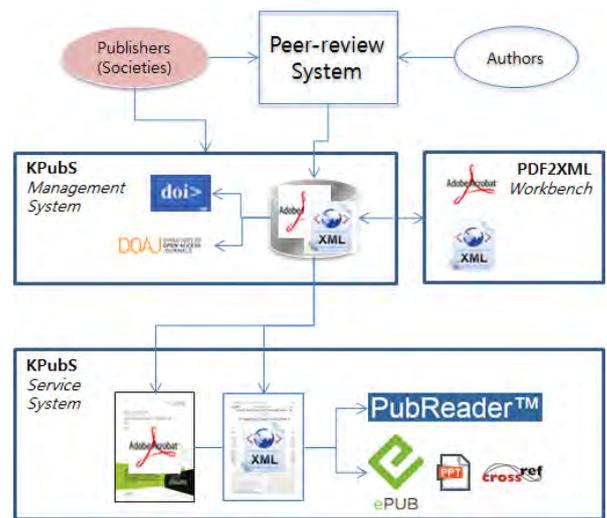
<표 1> Life-cycle Journal Publishing Model

단계	주요 수행 내용
창출 단계 Creation	<ul style="list-style-type: none"> • 논문투고심사시스템 연계 • 고품질의 논문 메타정보 • 고품질의 text PDF 원문 확보 • 고품질의 그림, 표 확보
아카이빙 단계 Archiving	<ul style="list-style-type: none"> • 아카이빙을 위한 데이터 구조화 • PDF 논문을 XML로 변환하는 기능 • DOI/DOAJ 기탁 XML 데이터 작성 기능 • 학술지 정보 연계 기능
서비스 단계 Service	<ul style="list-style-type: none"> • 개별 학술지의 전자저널 구축 • DOI landing page/OA repository 운영 • 서비스 색인 처리 • Pre/Post-print용 원고 제공 • 저작권 및 라이선스 정보 제공
확산 단계 Promotion	<ul style="list-style-type: none"> • WWS 연계 및 Google Crawling 허용 • 메타정보 제공 • 이용 및 인용 통계 제공

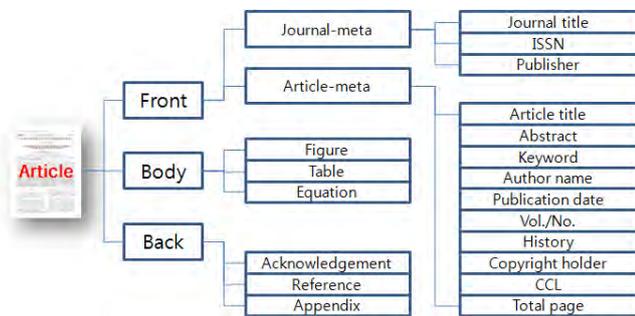
시스템은 자료를 수집, 가공, 업로드, 통계 모니터링 등을 수행하고, 서비스 시스템은 XML 원문 자료를 활용해서 검색, 브라우징, 링크, 다운로드 등의 서비스를 제공한다.

관리 시스템은 자료 수집을 논문투고심사시스템으로부터 전달받고, PDF 원문을 XML로 변환하는 작업기(workbench)와 연계되어 있으며, DOI와 DOAJ(Directory of Open Access Journals) 기탁(deposit)을 위한 메타데이터도 자동으로 구축하는 기능을 포함한다.

서버의 부하를 줄이고 개발 효율성을 향상시키기 위해 XML 변환기와 관리 시스템을 서비스 시스템과 독립적으로 분리 구성하였다.



<그림 2> Overview of KPubS System



<그림 1> JATS Data Model[7]

서비스 단계에서는 논문을 국제적으로 유통하기 위한 온라인 서비스 페이지를 구축하고 운영한다. 여기서는 DOI 랜딩 페이지와 오픈액세스 레포지토리 역할도 포함되어야 하며, 검색 효율 향상을 위한 색인 작업, pre-print 및 post-print 제공, 저작권 및 라이선스 정보 제공 등의 서비스를 포함하는 것이 필요하다.

마지막으로 확산 단계에서는 학술지의 이용과 인용을 높이기 위해서 국제적인 검색 서비스에 연계하는 것이 중요하다. Google Scholar에 색인되도록 하는 것과 이용 및 인용 통계 모니터링 기능이 여기에 포함된다.

4. 출판 플랫폼 개발

본 연구에서 개발한 KPubS 플랫폼은 크게 관리 시스템과 서비스 시스템으로 구성된다(<그림 2> 참조). 관리

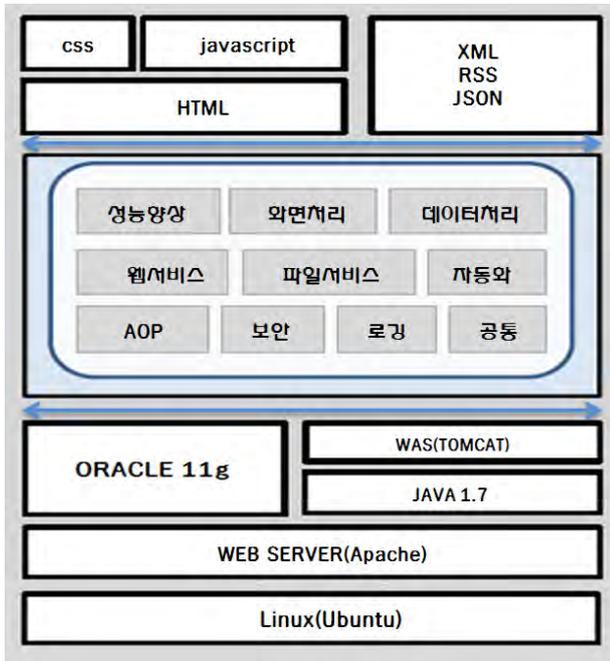
서비스 시스템은 XML 원문에 스타일시트를 적용하여 PDF보다 나은 UI와 함께 기존의 PDF 기반 서비스에서는 할 수 없었던 다양한 서비스를 제공한다. 원문을 PubReader로 보는 기능, e-PUB 형식으로 다운로드하는 기능, 본문 내의 이미지 파일을 PPT 형식으로 자동 다운로드하는 기능, 인용을 위한 서지정보 반출 기능 등이 여기에 해당한다. <그림 2>는 KPubS의 관리 시스템, PDF2XML 워크벤치, 서비스 시스템 간의 관계와 기능 구성을 보인다.



<그림 3> KPubS Pages using Responsive Web Technology

KPubS 플랫폼은 반응형 웹(responsive web) 기술을 적용하여서 모바일 기기와 PC를 포함한 모든 디바이스에서 최적화되게 하였다(<그림 3> 참조)

KPubS 플랫폼은 리눅스 기반 스프링 프레임워크와 MVC 패턴 및 시큐어 코딩 기법을 적용하여 개발하였으며, 확장성이 용이한 Maven을 채택하여 추가 모듈과의 연결이 가능하도록 설계하였다(<그림 4> 참조).



<그림 4> System Architecture of KPubS Platform

5. 결론

본 연구에서는 국내 학술지의 XML 기반 전자저널 출판 플랫폼이 개발되었다. 논문은 JATS XML DTD로 작성하였고, 관리 시스템을 통해서 원문과 메타데이터 입수와 XML 원문 구축 도구를 통해 PDF를 XML로 변환하는 작업을 일관작업으로 수행되도록 함과 아울러, PubReader와 e-PUB으로 원문을 다양하게 볼 수 있는 서비스를 구현하였다.

기존의 연구는 학술지 XML 구축에 한정되어 있지만, 본 연구에서 얻어진 출판 플랫폼은 논문의 투고에서 웹서비스까지의 전주기를 포함하기 때문에 원스톱으로 학술지 출판 작업을 수행할 수 있는 장점이 있다.

KPubS에는 현재 78종의 국내 학술지가 XML로 구축되어 안정적으로 서비스되고 있다[8].

향후 연구로 XML 원문을 활용한 시각화와 LOD(Linked Open Data)를 활용한 Semantic 서비스를 수행할 예정이다.

사사의 글

본 연구는 한국과학기술정보연구원의 “과학기술정보

서비스 및 개방형 체제 고도화(K-15-L02-C01-S04)” 사업의 지원을 받아서 수행되었다.

참고문헌

- [1] 서태설, 이해진, “전자저널 기술현황과 국내 학술지 전자저널 구축방안“, 정보과학회지, 28(10), 2010, pp.79-89.
- [2] ANSI/NISO Z39.96-2012 JATS: Journal Article Tag Suite Version 1.0), NISO.
- [3] <http://synapse.koreamed.org>
- [4] T. TACHIKAWA, H. SATO, K. NOZAWA, K.NAKAMURA, H. NAKANO, S. KURAUCHI, and H. GOTO, Journal of Computer Chemistry, Japan:Conversion of Japanese Word Manuscript into the XML Format and Web Publication Management System, J Comp Chem, 10(4), pp. 141-146, 2011.
- [5] H. Nakanishi, T. Naganawa, S. Tokizane, T. Yamamoto, Creating Journal Article Tag Suite extensible markup language from Japanese language articles and automatic typesetting using extensible stylesheet language transformations, Sci Ed, 2(2), pp. 63-72, 2015.
- [6] 국내정보실, “학술지 국제 출판 가이드”, 한국과학기술정보연구원, 2014
- [7] Tae-Sul Seo and Seon-Heui Choi, Data Models for Visualization Service of Scholarly Journal and Article Information, ICCS 2014, Jeju, June 2014.
- [8] <http://www.kpubs.org>