

학술정보의 효율적 전주기 관리 및 서비스 체제 구축에 관한 사례 연구

조성남*, 서태설*, 김완중*
*한국과학기술정보연구원
e-mail:chosn@kisti.re.kr

A Case Study on The Establishment of Efficient Management of Academic Information and Services.

Sung-Nam Cho*, Tae-Sul Seo*, Wan-Jong Kim*
*Korea Institute of Science and Technology Information

요 약

본 연구에서는 학술정보의 효율적 관리 및 서비스 체제를 수립하고 활용성을 강화하기 위해 논문 자동 수집 및 구축, 논문 투고 및 심사, 전자출판, XML 자동 구축에 이르기까지 전주기 기능 구현에 방안을 제시하고 이를 구현한 사례를 소개하고자 한다. 기존 Peer-Review 시스템은 데이터 수집 및 변환파트와 투고 및 심사, 전자출판 파트, XML 구축 파트 등으로 분리되거나 별도로 운영되어 왔다. 본 연구에서는 이러한 기능을 통합적으로 구현하기 위해 KISTI의 Peer-Review 시스템인 ACOMS와 XML 기반 학술정보 서비스 시스템(KpubS) 기능 및 프로세스를 통합한 구현 방안을 제시하고 이를 시스템으로 구현하였다. 그러나 아직까지 원문 PDF를 XML로 완전 자동 변환할 수 있는 기술은 없는 상태로써 반자동 변환 기능을 탑재하였다. 본 연구 내용은 다양한 수집 채널을 통해 입수된 학술정보 및 XML 자료에 대한 입력 체계를 제공하고, XML 본문으로부터 메타데이터를 자동 식별 및 추출하여 학술정보 DB를 구축·저장하며, 구축된 DB 및 XML 기반의 전자출판 기능 제공과 함께 웹서비스 단계까지의 전주기 공정에 대한 프로세스 모델 및 시스템에 관한 것이다. 이에 데이터베이스 구축을 위한 시간 및 비용을 절감할 수 있을 뿐만 아니라, 시맨틱 태깅 및 수식, 표, 그림 정보 제공 등 이용자에게 다양한 형태의 서비스를 제공할 수 있게 되었다.

1. 서론

지적활동 및 지식표현의 가장 대표적인 학술지는 모든 분야의 학문 발전과 지식 전파에 많은 기여를 해왔다. 매년 전 세계적으로 연구자에 의해 수많은 논문이 투고되고, 투고된 논문은 에디터에 의해 심사되고 학술지를 통해 발간되고 있으나, 이 엄청난 량의 논문의 투고, 심사, 발간 등의 작업을 일일이 수동으로 처리하는 데는 한계가 있기에 시스템으로 처리할 수 있는 Peer-Review System을 개발하여 활용하기 시작하였으며, 우리나라에서도 10년 전부터 공공기관을 중심으로 Peer-Review System을 개발하여 무상으로 학회에 보급하여 왔다. 최근에는 여러 민간업체에서 수익을 목적으로 Peer-Review System을 개발하여 사업화하고 있다. Peer-Review System이 발전하면서 기존에 서지로 발간되던 학술지가 전자저널로 출판되면서 활용성과 접근성이 강화되어 왔다. 일반적으로 전자출판은 메타데이터와 원문으로 구성되어져 있고 원문은 PDF 형태로 출판된다. 그러나 최근 PDF의 문제점과 한계를 극복하기 위한 일환으로 데이터베

이스 이식성이 높고 웹에서 다양한 표현이 가능한 XML 형태로 구축하여 서비스를 하고 있다. XML의 여러 가지 장점에도 불구하고 제작비가 비싸다는 큰 단점이 있다. 해외에서는 원문 장당 1달러 정도의 제작비가 소요되고 국내에서는 편당 5 만 원 내외 수준이다. XML 구축 고비용 문제를 위해 한국과학기술정보연구원에서는 원문 PDF를 XML 자동 변환 시스템(KPubS-Work)을 개발하여 서비스하고 있으나 PDF 특성 및 현존 기술의 한계를 감안할 때 100% 자동 변환에는 한계가 있기에 일부 수동 조치가 필요하지만 1 시간 정도 교육을 받으면 초보자도 능히 처리할 수 있는 작업이다. 그러나 학회 대부분은 인력 및 인식 부족으로 XML 변환 시스템을 활용하기 보다는 XML 전문 구축 업체 및 기관에 의뢰를 통해 구축하고 있는 실정이다.

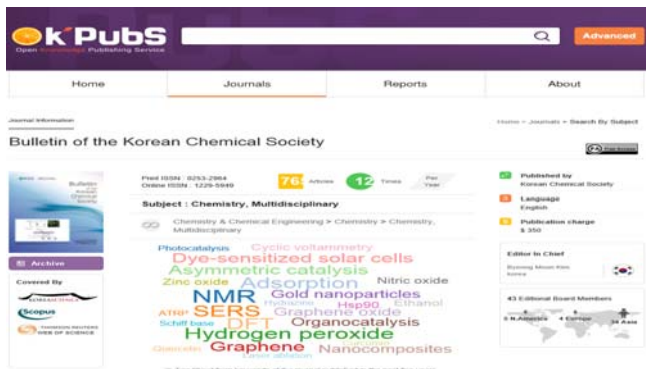
이와 같이 논문 투고 및 심사, DB구축, 전자출판을 위한 Peer-Review System과 향후 활용도 및 부가가치가 높은 XML의 변환 시스템이 별도로 운영되고 있기에 작업 소요시간도 길어지고 활용도도 떨어지는 등 효율적이지 못하다.

이에 본 논문에서는 논문투고심사시스템과 XML 기반 학술정보 구축 시스템을 통합하여 학술정보에 대한 종합적이고 체계적인 수집, Peer-Review, 구축, 관리 및 서비스가 가능한 전주기 프로세스와 구현 방안을 제시하고, 한국과학기술정보연구원에서 제공하고 있는 ACOMS 4.0(온라인 논문투고심사시스템)과 XML 기반 학술정보 서비스(KPubS), 그리고 KPubS-Work(XML변환시스템)를 통합적으로 구현한 사례를 소개하고자 하며, 아울러 XML 구축 및 활용을 보다 더 용이하도록 하기 위한 방안으로 논문투고심사시스템 프로세스에서 게재 확정된 논문에 대한 최종 논문 제출 시 XML 기반 입력시스템 제공을 통해 자동으로 XML로 변환과 동시에 PDF로도 저장한다. 이 기능은 향후 개발하여 서비스 예정이며 학회별로 선택적으로 사용이 가능하다. 이로써 전자출판 전 단계 및 완료 단계 모두 XML 구축이 가능하고 외부 원문 PDF 데이터도 XML 변환이 가능해짐에 따라 XML 구축이 한결 효율적으로 구축되고 서비스되는 기반을 마련하였다.

2. 관련연구

2.1 XML

XML은 W3C가 SGML의 한계를 극복하기 위해 제안한 웹문서를 구조화한 마크업(Markup)언어이다.[1] 우리가 흔히 웹 형식을 지정할 때 사용하는 HTML과 SGML보다 업그레이드된 형태로 볼 수 있다. XML은 특정 하드웨어나 S/W에 종속되지 않고 독립적으로 변환 및 생성이 가능하고 이질 데이터 간 연계가 쉽다. 또한 구조화는 물론 개방성과 확장성으로 인해 시각화, 통계, 검색, 계층화 등 정보표현을 다양하게 할 수 있다. 실제적으로 본 연구와 연관 연구로 구축된 XML 기반 학술정보 시각화 서비스 사이트(<http://open.kpubs.org>)에서는 XML 데이터를 통해 태그 클라우드 및 LOD(Linking Open Data)를 구축하여 텍스트 위주의 학술정보 서비스를 시각화로 시범 서비스하고 있다.[2] 시각화로 표현 시 정보전달이 더 직관적이고 이용자에게 친숙하다는 장점이 있다.[3]



(그림 1) 학술지 정보의 다면적 브라우징

(그림 1)은 실제로 XML 기반 학술정보 데이터를 통해 다면적 브라우징을 표현하였다.[4] 시각화를 특히 논문 학술정보는 대부분 PDF형식으로써 검색이나 표 및 그림 표현 등

에 한계가 있으나 XML은 이를 극복하여 본문 내용은 물론 표나 그림에 대한 검색도 가능하다.

2.2 Peer-Review System

학술지에 대한 논문 투고 및 심사 업무가 예전에는 이메일을 통해 수동적 방법으로 이루어졌으나 컴퓨터 및 웹 기술이 발전하면서 대부분은 온라인을 통해 이루어지고 있다. 매년 500개 이상 저널이 생성되고 3백만 개 이상 논문이 투고됨에 따라 온라인 논문투고심사시스템이 없다면 큰 혼란이 예상된다.[5] Ulrich의 데이터베이스에 따르면, 전 세계의 약 50,000 여종의 학술지 중 절반 이상은 peer-review 과정을 거치고 2012년 STM 보고서에 따르면, 중국과 동아시아에서 출판된 논문과 신규 학술지가 수적 측면에서 꾸준히 증가하고 있음을 알 수 있다.

전 세계에는 5,000~10,000개의 학술지 출판사가 있지만 그 중 소수의 영리 목적의 출판사만이 학술지 출판 사업을 지배하고 출판시장의 상당부분을 차지하고 있으며, 세계 출판계의 상위 1~4위를 차지하는 Elsevier, Springer, Wiley, Taylor & Francis 등 대형 출판사들이 출판하는 절반 이상의 논문에 대한 이익률을 차지하고 있다. 국내에서는 학회정보화사업 일환으로 한국과학기술정보연구원에서 2001년도에 ACOMS 3.0을 개발하여 학회에 무료로 지원하고 있으며 2010년도에 국내 학술지의 글로벌화를 위해 영문 버전 ACOSM 4.0을 개발하여 보급함으로써 해외 학술인의 참여 및 활용이 가능하게 되었다. 공공기관인 한국연구재단에서도 논문투고심사시스템 JAMS 2.0을 무료로 보급하고 있으며 학술지 평가 시에 온라인 논문투고심사시스템 활용 유무를 항목으로 두고 있다. 또한 민간에서도 영리를 목적으로 논문투고심사시스템을 유료로 보급하고 있다. (그림 2)는 주요 국내 및 해외 보급처를 보여주고 있다.

국내현황	국외현황
<ul style="list-style-type: none"> • 민간사업자를 중심으로 시스템을 개발하여 학회에 판매 또는 대여 • ㈜휴로닉스: 이공계열 10여개 학회 • ㈜네띠앤: 이공계열 20여개 학회 • ㈜OnPCS: 이공계열 30여개 학회 • ㈜M2Community: 의학계열 30여개 학회 • 메드링.인포링: 의학계열 50여개 학회 • KinNet: 이공계열 10여개 학회 • 비영리 기관으로서 학회에 지원: KISTI-ACOMS, 연구재단 JAMS 	<ul style="list-style-type: none"> • 해외의 경우, 대규모 출판사를 중심으로 시스템 판매 • 논문투고부터 심사, 게재 결정, 편집과정, 웹저널 생성, 책자 발간을 모두 시스템으로 처리 • Thomson Scientific사 SCHOLARONE: 2,200 학술지 • Aries Systems Co.의 Editorial Manager: 150개 학회 • Elsevier EES: 350여 만 • Cornell Univ. dPubs, British Uni. OJS • 일본의 J-STAGE: KISTI와 같이 국가기관이 학회를 지원

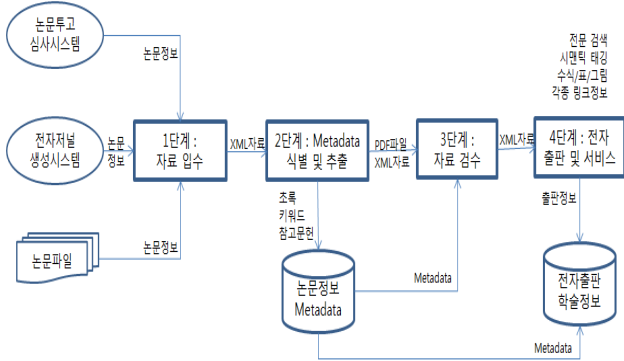
(그림 2) 국내외 Peer-Review System

기존 Peer-Review 시스템은 논문 투고부터 심사, 게재 결정, 편집과정, 웹저널 생성, 책자 발간을 모두 시스템으로 처리하고 있으나 외부로부터 데이터 수집을 통한 학술정보 DB 구축, 원문 PDF 파일의 XML 자동 추출 및 변환 기능은 별도로 분리되어 운영됨에 따라 본 연구에서는 학술정보 수집, 투고 및 심사, 전자출판, XML 자동 구축까지 학술정보 전주기 프로세스를 수립하여 시스템으로 구축한 사례를

보여주고자 한다.

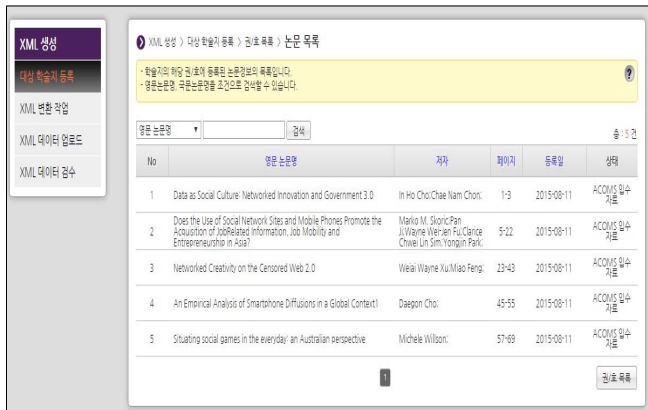
3. 학술정보 전주기 관리 및 서비스 구현

학술정보 자료 처리는 (그림 3)과 같이 학술정보를 입수하고, 메타데이터를 추출 및 저장하고, 검수단계를 거친 후 전자출판 및 서비스 단계를 거친다.



(그림 3) 자료 처리 흐름도

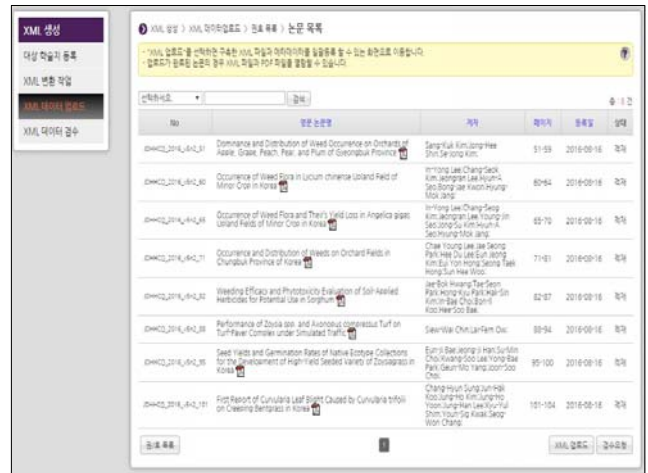
1단계인 자료 입수 단계에서는 학술단체별로 보유하고 있는 논문투고관리시스템 또는 전자저널 생성시스템을 통하여 실시간으로 정보를 수집하거나, 학술정보 입수시스템을 통하여 논문 파일에서 메타데이터와 원문을 등록하는 형태로 학술정보를 수집한다.



(그림 4) ACOMS에서 입수한 논문 목록 화면

(그림 4)는 한국과학기술정보연구원(KISTI)이 학술단체에 무료로 제공하는 논문투고관리시스템인 ACOMS를 통해 입수된 논문정보 목록 화면이다.

메타데이터 식별 및 추출 단계에서는 자료 입수 단계를 거친 학술정보를 대상으로 JATS1.0 표준을 따르는 전문 XML 파일을 업로드 할 수 있도록 구성하고 업로드된 XML 파일을 파싱하여 <FRONT> 태그 내에서 제목, 저자, 초록, 키워드 등의 정보를 추출하여 논문정보 데이터베이스에 저장하고, <BACK>태그 내에서 참고문헌 정보를 추출하여 참고문헌정보 데이터베이스에 저장한다. (그림 5)는 업로드된 XML 파일에서 메타데이터를 추출 후 추출된 정보를 확인하는 화면이다. 모든 정보를 확인한 후에는 3단계인 검수단계로 자료를 이관하기 위하여 검수요청 과정을 수행한다.



(그림 5) 메타데이터 추출 완료 후 확인 화면

검수 단계에서는 데이터베이스에서 추출된 메타데이터를 확인하고 JATS1.0 스타일시트를 적용한 XML 파일 미리보기 기능을 제공하여 내용적 오류가 있는지 확인할 수 있도록 구성하였다. (그림 6)와 같이 모든 내용 검토가 완료되면 검수완료 기능을 통하여 서비스 대상으로 처리하거나 문제가 있는 경우 검수반려 기능을 통하여 전단계로 이관되도록 기능을 구성하였다.



(그림 6) 검수 반려 및 완료 처리 화면

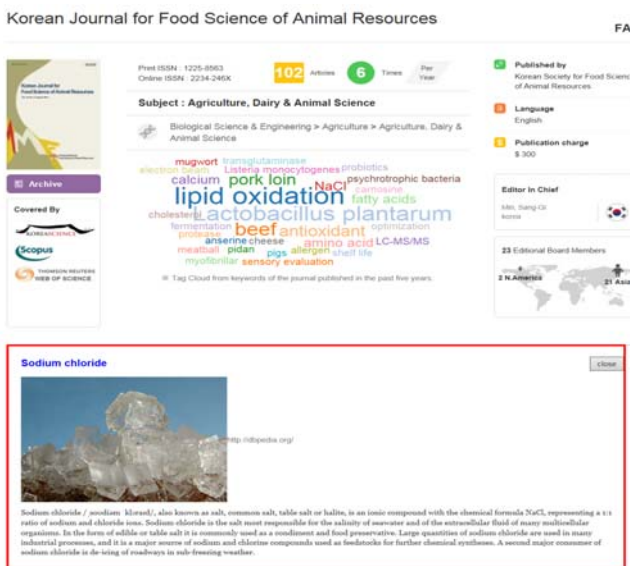
전주기 학술정보 관리시스템의 최종 단계인 전자출판 및 서비스 단계에서는 전 단계를 거쳐 구축된 학술정보를 학술지, 권/호, 논문정보 순서로 열람할 수 있도록 구성하였다. 논문 전문에 대하여 검색할 수 있도록 검색 기능을 제공하고, 전문 XML을 파싱하여 표 및 그림정보에 대한 내부 링크를 제공하고 참고문헌의 경우 DOI 정보가 있으면 해당 원문에 접근할 수 있도록 링크를 제공하였다. (그림 7)은 논문 정보 상세를 열람하는 화면으로 제목, 저자, DOI, 저작권 관련, 라이선스 정책 정보를 확인할 수 있도록 구성하였고 EPUB, PubReader 파일을 실시간으로 생성하도록 구성하였다. 또한, 정보공유를 지원하기 위하여 각종 SNS API를 통해 연계할 수 있도록 페이지를 구성하였으며, 조회 및 다운로드수를 알 수 있도록 Metrics 페이지를 구성하고 인용정보를 알 수 있도록 Cited By 페이지를 구성하였다. 논문 열

람의 편의를 제공하기 위하여 목차를 통한 해당 페이지 링크를 제공하고 이미지 재사용이 가능하도록 다운로드 기능을 제공하였다.



(그림 7) XML 기반 논문정보 서비스 화면

부가적으로, XML 기반 원문을 통해 문자 위주 서비스에 서 탈피한 학술정보 시각화 표현이 가능하다. (그림 8)과 같이 논문의 초록과 본문에서 모든 단어들을 태깅하여 빈도수를 통해 태그 클라우드를 생성하였고 태그클라우드에서 특정 단어를 선택하면 LOD(Linking Open Data)를 통해 외부 사전 서비스하고 자동 연계되도록 구축하여 서비스하고 있다. 아직까지는 영문 단어만 가능하고 한글 단어는 추후에 서비스될 예정이다. 또한 태그 클라우드에서 LOD 연계 사례로서 빨간색 박스 부분이 선택한 단어에 대한 설명이다.



(그림 8) 태그 클라우드와 LOD 연계

본 연구에서는 ACOMS 4.0에서 생성된 학술지 메타데이터 및 PDF를 XML 변환 시스템 KPubS-Work 시스템으로 불러들여 PDF를 XML로 변환하지만 서두에 언급한 바와 같

이 기술의 한계로 인해 전면 자동 변환이 불가능함에 따라 일정 수작업이 필요하다. 이를 개선하는 방안으로 논문투고 시스템에서 게재 확정된 논문의 최종 논문 제출 시 파일로 제출하지 않고 XML 입력기를 통해 저자가 직접 항목에 입력하게 되면 시스템은 입력 데이터를 토대로 XML 자동 구축은 물론 PDF까지 자동 생성하게 된다. 이 경우 저자가 직접 입력하기 때문에 비용은 전혀 들지 않지만 저자의 협조 및 수고가 필요하다. 해당 기능은 2017년도 사업을 통해 시스템에 반영할 예정이다. 해당 기능이 개발되면 이미 전자출판된 학술지에 대한 일괄 XML 변환(Batch)은 물론 Peer-Review System을 통해서 입수된 논문의 전자출판 시점에 XML 생성이 가능해진다. 본 연구를 통해 분리되어 있는 Peer-Review System과 XML자동 변환기를 하나의 시스템으로 통합하고 XML 구축의 효율성을 향상시키기 위해 Peer-Review System에 XML 입력기능을 포함함으로써 학술정보의 투고, Peer-Review, 편집, 전자출판, XML 구축, 학술정보 시각화까지 전 과정을 일원화하였다.

4. 결론 및 향후 연구

본 연구에서는 학술정보 수집 및 관리, 활용성을 강화하기 위해 논문투고심사시스템과 XML 변환 및 구축 시스템을 결합하여 논문 투고 및 심사 기능 제공, 원문 데이터의 수집, XML 자료로부터 모든 메타데이터 항목을 자동 추출하여 데이터베이스를 구축하고 XML 자료를 기반으로 본문 서비스를 제공하는 기술을 제공함에 있다. 이를 통해 학술정보의 수집, 데이터베이스 및 XML 구축을 위한 시간 및 비용을 절감할 수 있을 뿐만 아니라, XML을 통해 시맨틱 태깅 및 수식, 표, 그림 정보 제공 등 이용자에게 다양한 형태의 서비스를 제공하는 등 학술정보의 가용성 및 활용성을 크게 향상시킬 수 있었다..

향후 연구로는 논문정보 PDF 파일에서 전문 XML 파일을 생성할 수 있는 시스템을 구성하여 일부 자료 처리과정의 수동적 과정을 반자동화할 수 있는 기능 설계 및 구현에 대한 연구가 필요하다고 본다.

참고문헌

- [1] W3C, <https://www.w3.org>
- [2] 조성남, 서태설, “학술정보 시각화 서비스 개발에 관한 연구”, 문헌정보학회지, 50(2) 2016, pp.183-196.
- [3] 서태설, 이혜진, “전자저널 기술현황과 국내 학술지 전자저널 구축방안”, 정보과학회지, 28(10), 2010, pp.79-89.
- [4] <http://open.kpubs.org>
- [5] 한국과학기술정보연구원, “학술정보 시각화 서비스 방안 연구”, 2015