

JATS와 출판용 XML DTD를 이용한 학술지와 논문 메타데이터 요소 개발에 관한 연구*

A Study on Metadata Elements for Journal and Articles Using JATS and Publishing XML DTD

이 용 구(Yong-Gu Lee)**

〈 목 차 〉

- | | |
|----------------------|------------------------------|
| I. 서론 | 3. 학술지 출판용 XML DTD |
| 1. 연구의 필요성 및 목적 | III. 학술지와 논문을 위한 메타데이터 요소 개발 |
| 2. 연구방법 및 선행연구 | 1. 개체 관계 및 요소 선별 기준 |
| II. 학술지와 논문의 기술규칙 사례 | 2. 학술지와 논문 관련 메타데이터 요소 |
| 1. 목록규칙에서 학술지 기술요소 | IV. 결론 |
| 2. JATS | |

초 록

학술지와 학술지에 실린 논문의 메타데이터는 학술정보를 검색하는데 있어 매우 중요한 수단이 된다. 이 연구에서는 도서관계의 주요 목록 규칙, 논문 관련 미국 표준인 JATS, 해외의 유명 출판사의 출판용 XML DTD를 분석하여 학술지와 논문이 가지는 메타데이터의 기술요소를 파악하고자 하였다. 이들을 비교분석한 결과, 총 50개의 메타데이터 요소를 추출하였다. 학술지 관련 요소는 10개, 권/호는 7개, 논문은 22개, 그리고 저자 관련해서는 11개이다. 메타데이터의 선별적인 선택이 가능하도록 7개의 필수요소, 9개의 해당시 필수요소, 34개의 재량요소로 구분하여 제시하였다. 이 연구는 실제 사용되는 XML DTD를 이용하여 메타데이터 요소를 선별하였다는데 의미가 있다.

키워드: 학술지 메타데이터, 논문 메타데이터, JATS, 출판용 XML DTD, 메타데이터

ABSTRACT

In the library domain, the metadata concerning the journal and the articles published in the respective journal are the important attributes that can be used to search academic related information. In order to conduct this study, the American JATS standard and the XML DTDs of journal articles commonly used by well-known publishers were used. By comparing JATS standard to XML DTDs used by Elsevier and Springer, we were able to extract a total of 50 metadata elements(10 about the journals, 7 about the issues, 22 about the articles, and 11 about the authors). Additionally, we identified a total of 7 mandatory elements, 9 applicable mandatory elements, and 34 optional metadata elements. The implication of this study is that we selected metadata elements using XML DTDs that are used in the real world.

Keywords: Journal metadata, Article metadata, JATS, Publishing XML DTD, Metadata

* 이 논문은 2011년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 연구되었음(NRF-2011-327-H00020)

** 계명대학교 문헌정보학과 조교수 (yonggulee@kmu.ac.kr)

•논문접수: 2015년 6월 9일 •최초심사: 2015년 6월 9일 •게재확정: 2015년 6월 25일

•한국도서관·정보학회지 46(2), 367-392, 2015. [http://dx.doi.org/10.16981/kliiss.46.201506.367]

I. 서론

1. 연구의 필요성 및 목적

학술지(journal)는 일반적으로 연구자나 학자가 최신의 정보를 얻을 수 있는 중요한 정보 유통 경로 중 하나이다. 학술지는 단행본과 달리 저자가 생산한 최신 연구결과를 비교적 짧은 기간 안에 구독자인 이용자에게 전달할 수 있는 매체이다. 최근에는 학술지의 종수와 그 안의 콘텐츠도 점점 늘고 있는 추세인데, 오픈 액세스 저널(Open Access Journal) 운동과 같은 시도나 변화가 여기에 한몫한다.

학술지나 학술지에 게재된 논문 또는 기사(article)의 메타데이터는 해당 자원을 이용하는 데 있어 매우 중요한 도구가 되는데, 일반적으로 메타데이터가 특정 자원을 발견하거나 식별 하는데 필요하고, 자원의 관리 및 보존을 용이하게 해주기 때문이다.

최근 10여 년 동안 학술지나 논문 관련 정보를 제공하는 정기간행물 기사색인 서비스에 대한 연구는 시스템이나 서비스 자체에 대한 주제를 다루었으며, 학술지와 논문의 메타데이터 관련 연구들은 주로 데이터의 품질을 많이 다루었다.

하지만 메타데이터의 품질 관리만큼 중요한 것은 어떤 메타데이터 요소(elements)를 추출할 것인가의 문제이다. 이러한 문제는 메타데이터 스킴(scheme) 또는 스키마(schema)와 관련된다. 왜냐하면 아무리 정확하게 높은 품질의 메타데이터를 구축하고자 하여도 추출하고자 하는 특정 요소가 스키마에 정의되어 있지 않다면, 추출할 수 없기 때문이다.

일반적으로 메타데이터 스키마는 특정 목적을 위해 정의된 사용 규칙의 집합으로 정의할 수 있다. 다시 말해, 메타데이터의 스키마에서 정해진 규칙에 따라 필수적인 요소나 필요하다고 생각되는 요소를 원 자료로부터 추출하고, 이를 데이터베이스에 저장하거나 검색 시스템에 입력하여 서비스를 제공한다. 이때 원 자료에는 존재하는 정보인데 메타데이터 스키마가 이 정보에 대한 규칙이나 규정이 없어서 추출하지 않는다면, 그 메타데이터는 이용자에게 추출되지 않은 만큼의 정보를 전달하지 못하거나 해당 원 자료의 검색에서 실패를 가져오게 된다. 즉 다시 메타데이터 품질 문제와 관련이 된다.

한 전자저널 제공업체의 실제 메타데이터를 살펴보면, 학술지에 게재된 논문 원문에는 'JATS'와 같은 영어 약자로 된 키워드가 국문과 영문 키워드에 모두 쌍으로 있는데, 내려 받은 메타데이터의 키워드 요소에는 한번만 출현하였다. 만약 이 상황에서 영문이나 국문 키워드만을 추출하여 검색하거나 다른 기관과 메타데이터를 교환하고자 한다면, 정보 손실로 인해 원문의 정보가 정확히 검색되거나 전달되지 않는 오류가 발생할 수 있다. 이러한 오류는

해당 정보원을 기술하는 정확한 메타데이터 스키마를 개발한다면 해결할 수 있다. 즉 해당 업체가 학술지와 게재 논문 관련 자신들의 스키마에서 국문과 영문 키워드를 서로 다른 요소로 구조화 하면 해결할 수 있다.

학술지와 그에 게재되는 논문을 위한 메타데이터 스키마를 만들기 위해 출판용 XML DTD나 학술 콘텐츠의 메타데이터인 JATS(Journal Article Tag Suite) 등을 살펴 볼 필요가 있다. 현재 큰 규모의 상업 출판사와 학회는 학술지와 그에 게재된 논문을 출판하고 전자저널로 서비스하기 위해 자신들만의 XML(또는 SGML) DTD를 이용한다. 이러한 포맷이나 구조에서 학술지와 게재 논문의 속성을 표현하는 요소들을 포함하고 있어, 이를 이용하여 해당 메타데이터 스키마를 개발할 수 있다.

따라서 이 연구는 학술지와 게재 논문을 표현하는 다양한 XML DTD 사례를 제시하고, 이들이 갖고 있는 메타데이터 요소 중에 핵심적인 요소와 공통된 요소를 추출하는 귀납적 접근법을 적용하였다. 이 연구에서 사용한 사례로 주요 목록 규칙, 학술 콘텐츠용 메타데이터 JATS, 그리고 다수 기관의 출판용 XML DTD 등을 분석하였다. 이를 통해, 연구 목적으로 학술지와 게재 논문이 가지는 메타데이터의 기술요소와 최소한의 기술규정 및 기술을 위한 스키마를 개발하고자 하였다.

2. 연구방법 및 선행연구

앞서 기술한 연구 목적을 위해 다음과 같은 연구방법을 적용하였다. 우선, 이 연구에서는 학술지와 게재 논문을 기술하는 메타데이터에 대한 선행 연구를 파악하였다. 이를 위해 정기간행물 또는 연속간행물과 게재 논문을 대상으로 한 기사색인 서비스에 대한 국내 선행연구를 조사하였다.

둘째, 학술지와 게재 논문의 메타데이터를 기술하는 사례를 조사하였다. 이를 위해, 먼저 문헌정보학 분야의 목록 규칙들이 학술지와 게재 논문에 대해 어떠한 기술요소를 규정하고 있는지 살펴보았다. 또한 주로 학술지 논문의 콘텐츠에 대한 메타데이터를 기술하는 미국 표준인 JATS와 해외 유명 출판사와 학회가 출판을 위한 용도로 사용하는 학술지 및 논문의 XML DTD를 분석하였다.

셋째, 앞서 분석된 목록규칙의 기술요소, 다수의 출판용 DTD와 미국 표준인 JATS 등을 검토하여 이들에게 공통으로 사용되는 기술요소를 선별하고, 이들을 체계적으로 구성하여 학술지와 게재 논문의 메타데이터 기술요소와 하위요소 세트를 개발하였다. 또한 메타데이터의 각 요소별로 기본적인 기술규칙에 대한 설명, 필수 여부, 출처 등을 제시하였다.

최근 메타데이터 스키마나 요소 생성과 관련 다수의 연구들이 존재하는데, 주요 연구들을

살펴보면 다음과 같다.

최근 다양한 분야에서 메타데이터 스키마나 요소 개발과 관련된 연구들은 많이 있다(박옥남 2012; 김관준 2010; 김수정, 김용 2012). 그리고 연속간행물이나 정기간행물에 대한 서비스나 시스템에 관한 연구들도 다수 존재한다(윤정옥 2003; 이은철 외 2009). 하지만 학술지와 학술지에 실린 논문에 관한 메타데이터 개발에 관한 연구는 많지 않다.

국내 정기간행물 기사색인에 관한 한중엽(2000)의 연구는 연속간행물 기사 DB를 구축하기 위해 서지데이터 요소의 표준화 방안을 제시하였다. 이를 위해 국내외의 대표적인 서지 데이터베이스(10개 기관)의 데이터요소를 비교하고, ISO 690의 서지기술 요소를 분석하였다. 연구결과 기본적인 데이터 요소 17가지를 추출하였다. 이들 요소로 표제, 저자, 저자소속, 부수적 저자, 학술지명, 기사의 수록위치(권호와 페이지), ISSN, 발행지, 발행자, 발행일자, 본문언어, 판사항, 키워드, 주제분야, 초록(또는 내용목차), 소장기관, 원문보기 등을 제시하였다.

한성국과 이현실(2005)은 학술지 정보 공유의 활용 체계를 실현하기 위해서 상호운용성에 초점을 두고 학술지 정보 표준화를 위해 식별자 정보, 서지 정보, 소장정보 등 메타데이터의 특성별로 나누어 분석하였다. 식별자 정보는 ISSN, SICI, DOI 등을 고려 대상으로 하였으며, 서지 정보는 MARC 포맷과 더블린 코어(Dublin Core) 메타데이터를 조화시킨 형태가 필요하다고 주장하였다.

최근 이은철 등(2009)은 정보자원으로서 정기간행물 기사색인 서비스의 중요성을 인식하고, 이용자의 이들 서비스에 대한 요구사항과 국내외 정기간행물 기사색인 서비스의 분석을 수행하였다. 이를 통해 국내 정기간행물 기사색인 서비스 구축, 메타데이터 표준, 전거과일, 이용자 참여 서비스, 패킷 기반 다각적 정보탐색 기능, 식별체계 구축 등을 제안하였다. 특히, 국내에서 정기간행물 기사색인과 관련하여 메타데이터 표준에 대한 연구가 활발히 진행되지 않고 있는 실정에서, 이 서비스를 제공하는 기관이 조금씩 상이한 메타데이터 포맷을 유지하고 있음을 제시하고, 콘텐츠의 공유와 상호운용성 확보를 위해 표준화된 메타데이터를 제안하였다. 이들은 모두 27개의 메타데이터 요소를 포함하였다. 특히 기사색인 고유의 식별자를 제공하거나 원문의 식별자, 발행처 URL, 자료유형, 출처 분류, 주제어, 원문형태, 등재 여부 등의 메타데이터 요소를 추가함으로써 기사 색인 이용자의 접근점 및 활용도를 높이고자 하였다. 또한 이용자 참여 서비스를 제공하기 위하여 이용자 태그나 이용자 서평을 포함할 것을 제안하였다.

이들 선행연구는 실제 서비스 하는 기사색인 서비스 데이터베이스를 대상으로 데이터 요소를 분석하여, 대부분이 학술지나 게재 논문만을 대상으로 한 메타데이터 요소를 제시하였다. 선행연구와 달리 이 연구는 학술지와 논문 관련 미국의 표준과 실제 이들을 생산하는 출판사가 가지고 있는 메타데이터 요소를 활용한 점에서 다르다. 또한 학술지, 게재 논문 그리고 저자 정보까지 관련된 메타데이터를 제시하였다는 측면에서 차이가 있다.

II. 학술지와 논문의 기술규칙 사례

1. 목록규칙에서 학술지 기술요소

도서관의 목록규칙은 정보자원을 편목하거나 색인하는데 필요한 요소를 기술하는 규칙을 제시하기 때문에 엄밀히 말하면 메타데이터는 아니다. 하지만 이 연구에서는 목록규칙을 통해 어떠한 기술요소들을 추출해야 하는지 알아보려고 하였다.

이 연구에서는 학술지에 대한 주요 기술요소를 알아보기 위해 국제표준서지기술(ISBD)을 중심으로 주요한 목록규칙인 영미권의 목록규칙(AACR2R) 그리고 RDA(Resource Description and Access)를 살펴보았다.

이들 목록규칙 대부분은 학술지에 해당하는 연속간행물 자체에 대한 규칙을 제시하고 있으나, 학술지 게재 논문에 대한 기술요소나 어떻게 기술해야 하는지에 대한 규정은 없다. 대신 이들 논문에 대한 검색이나 서비스는 주로 정기간행물 기사색인 서비스나 서지 데이터베이스를 통해 이루어진다.

먼저 『영미편목규칙』(Anglo-American Cataloguing Rules: AACR)은 미국도서관협회 등이 주축이 되어 1967년에 초판이 나왔으며, 최신 버전은 2002년에 추가 개정된 AACR2R이다. 이 목록규칙은 도서와 팸플릿, 지도자료, 악보, 녹음자료, 영화자료, 전자자원 등을 주요 편목 대상으로 한다. 학술지인 연속간행물은 계속자원(continuing resources)으로 규정하고 규칙을 제시하고 있다. 이는 2002년에 대폭 개정이 되면서 명칭이 바뀐 것으로 연속간행물과 통합자원을 함께 포함하고 있다(Maxwell 2004).

연속간행물에 대해 AACR2R에서 제시하고 있는 목록규칙은 단행본과 동일하다. 즉 AACR2R는 연속간행물의 기본저록 및 부출저록의 관한 규칙과 기술을 단행본처럼 규정하고 있다(Maxwell 2004, 368-369). 다만, 도서관 현장에서 목록작성자가 전체 자료의 부분인 개개의 저작이나 논문을 검색하도록 서지 레코드로 기술하고, 이를 모체의 자료와 관련하여 분출(analytics)을 할 수 있도록 자체적인 규정을 제시하기도 한다(OCLC 2015).

국제표준서지기술(ISBD)은 목록기술의 국제적 표준화를 도모하며, 기계가독 형태로 서지 정보의 국제교환을 촉진하기 위한 목적으로 작성되었다. 따라서 ISBD는 KCR 4판과 AACR2와 같은 각국의 목록규칙에 영향을 미쳤다(문헌정보학용어사전 편찬위원회 2010, 59). 일반적인 원칙을 기술한 ISBD(G)의 주요 기술요소는 표제 및 책임표시사항, 판사항, 자료 특성사항, 발행 및 배포사항, 형태기술사항, 총서사항, 주기사항, 표준번호 및 입수조건 사항 등 8개이다. 마찬가지로 연속간행물에 해당하는 ISBD(CR)도 이들 8개의 기술 요소를 동일하게 반영하고 있다.

6 한국도서관정보학회지(제46권 제2호)

최근 서지레코드에 개체-관계 모형을 적용한 FRBR과 FRAD를 기반으로 AACR2의 후속판인 RDA가 개발되었다. RDA는 10부에 걸쳐 목록 규칙을 담고 있다. 주요 내용은 제1, 2, 3집단의 개체들의 속성(attributes) 또는 요소에 대한 기술 지침을 담고 있으며, 이들 개체들 간의 관계(relationships)를 어떻게 기록할 것인지 담고 있다(박진희 2009; 조재인 2009). RDA는 저작, 표현형, 구현형, 개별자료에 대해 각기 다른 속성과 관계에 대해 규정하고 있다. 연속간행물이나 이들에 실린 논문의 경우, 저작을 어떻게 정의하고 다른 개체들 간의 관계를 어떻게 규정할지에 대해서는 연구가 계속되고 있다.

앞에서 제시된 연속간행물에 대한 ISBD(CR)과 RDA에서의 기술규칙을 비교하여 살펴보면 <표1>과 같다(The Standing committees of the IFLA section on Cataloguing and the IFLA Section on Serial Publications 2002; The Joint Steering Committee for Development of RDA 2010). ISBD(CR)와 RDA는 연속간행물에 대한 규정이 매우 유사하게 제시되어 있음을 알 수 있으며, RDA의 경우 여러 장으로 흩어져 있음을 알 수 있다.

<표 1> ISBD와 RDA에서 연속간행물(학술지) 기술요소

기술요소	ISBD(CR)	RDA
표제와 책임표시사항	본표제 자료유형표시(optional) 대등표제 표제관련정보 책임표시	*2.3.2 본표제 3.2 매체유형 2.3.3 대등표제 2.3.4 표제관련정보 *2.4.2 본표제 관련 책임표시
판사항	판표시 대등판표시(optional) 특정 판에 관련된 책임표시 부차적 판표시 부차적 판의 책임표시	*2.5.2 판표시 2.5.3 대등판표시 2.5.4 특정 판에 관련된 책임표시 *2.5.6 어떤 판에 대해 이름이 나타나 있는 개정판에 대한 표시 2.5.8 어떤 판에 대해 이름이 나타나 있는 개정판에 대한 책임표시
권호차사항	권호차표시 연월차표시	*2.6.2 컷호의 권호표시 *2.6.4 마지막호의 권호표시 *2.6.5 마지막호의 연월차표시
발행, 배포 등 사항	발행지, 배포지 발행처, 배포처 발행년, 배포년 인쇄지, 제작지 인쇄처, 제작처 인쇄년, 제작년	2.7.2 제작지 *2.8.2 발행지 *2.9.2 배포지 2.7.4 제작처 *2.8.4 발행처 *2.9.4 배포처 *2.7.6 제작일자 *2.8.6 발행일자 *2.9.6 배포일자 *2.10.2 인쇄지 *2.10.4 인쇄처 *2.10.6 인쇄일

형태기술사항	특정자료표시와 수량 기타 형태세목 크기 말림자료(optional)	*3.4 수량 7.15 삽화 3.5 크기 27.1 관련 구현형
총서사항	총서, 하위총서의 본표제 총서, 하위총서의 대등표제 총서, 하위총서의 표제관련정보 총서, 하위총서의 책임표시 총서, 하위총서의 ISSN 총서, 하위총서 권호	*2.12.2 총서의 본표제 *2.12.10 하위총서의 본표제 2.12.10 총서의 대등표제 2.12.11 하위총서의 대등표제 2.12.4 총서의 표제관련정보 2.12.12 하위총서의 표제관련정보 2.12.6 총서의 책임표시 2.12.14 하위총서의 책임표시 2.12.8 총서의 ISSN 2.12.9 하위총서의 ISSN *2.12.9 총서 권호 *2.12.17 하위총서 권호
주기사항	주기사항	2. 구현형과 개별자료의 식별 3 수록매체의 기술 4 입수와 접근 정보의 제공 7 내용의 기술 25.1, 26.1, 27.1 저작, 표현형, 구현형, 개별자료 간 관계 기술의 규정들
표준번호 및 입수조건사항	표준번호(ISSN) 등록표제(약어 포함) 입수조건, 가격(optional) 이용자격(optional)	*2.15 구현형의 식별기호 2.3.9 등록표제 4.2 입수조건 2.15 구현형의 식별기호 4.2 입수조건

* 기술 수준에서 핵심요소(core element)를 나타냄

실제로 도서관이나 유사기관은 학술지에 게재된 논문을 검색하고 접근할 수 있도록 관리하고 있는데, 이는 주로 정기간행물 기사색인 서비스를 통해서 이다. 기사색인은 인쇄물 형태도 있지만 온라인 데이터베이스로 된 것도 있다. 인쇄물 형태는 참고정보원의 한 유형으로 정기간행물의 논문을 제목, 주제, 또는 저자명 순으로 나열하여 제공한다. 최근에는 웹 서비스를 통해 제공되는데, 먼저 데이터베이스에 학술지와 논문에 대한 주제나 키워드, 저자, 수록 학술지, 발행기관 등 관련 정보를 입력하고 해당 정보를 검색할 수 있다. 이들이 제공하는 정보는 논문의 표제, 저자명, 수록 학술지명, 정기간행물의 권호, 발행일, 해당 논문의 쪽 범위 및 수 등 이다. 또한 추가적으로 초록, 전문(full-text), 그리고 참고문헌 및 인용 관련 링크 등을 제공한다.

2. JATS

현재까지 메타데이터를 포함하여 논문의 전체 내용을 기술하거나 태깅하는 국제 표준은 없으며, 국가 표준으로는 JATS가 유일하다. JATS는 수년에 걸쳐 현장에서 시행착오를 거치면

서 발전하였고, 이를 국가 표준으로 전환하였다. 이러한 배경을 볼 때 JATS가 학술지에 게재 된 논문의 메타데이터를 비교적 잘 반영하였다고 판단된다.

JATS는 미국 NISO의 표준(Z39.96-2012)이다. JATS는 주로 학술지에 실린 논문 중심의 콘텐츠를 태깅(기술타깅)하기 위한 XML 요소와 속성 집합으로 정의된다. 뿐만 아니라 레터(letter), 사설, 도서 그리고 제품 리뷰 등과 같은 비논문 자료에 대한 텍스트 형식이나 그래픽 형식의 콘텐츠도 태깅이 가능하다. 출판사나 도서관, 기관 레포지토리 등에서 학술지의 논문과 같은 콘텐츠를 교환할 수 있도록 JATS는 해당 콘텐츠를 XML 형식으로 제공한다(National Information Standards Organization 2012).

JATS는 NLM(The National Library of Medicine)의 NCBI(The National Center for Biotechnology Information)로부터 시작되었다. 생의학 및 생명과학 분야의 학술지에 대한 무료 전문 아카이브 서비스인 PMC(PubMed Central)를 만들 때, NCBI는 PMC의 모든 콘텐츠를 관리하기 위해 공통된 형식이 필요했으며, 이를 위해 모든 콘텐츠가 하나의 DTD(Document Type Definition)를 갖도록 하였다. 이 DTD가 2003년부터 일반인에게 공개되어, NLM DTD가 되었다. 이후 NLM과 NISO는 이용자의 증가와 국가 차원의 콘텐츠 교환을 위해 이 DTD를 학술지 논문 모형(article models) 표준인 JATS로 만들었다(Beck 2011).

NLM DTD는 수년에 걸쳐 세 모형인 보존 모형, 출판 모형, 그리고 저작 모형을 발전시켰으며, 이를 바탕으로 2012년에 JATS(NISO Z39.96-2012) 버전 1.0을 제정하였다(National Information Standards Organization 2012). 따라서 JATS도 앞서 설명한 세 모형의 용도에 해당하는 Tag Suite를 가지고 있으며, 이들의 요소와 속성의 개수는 다음과 같다.

- Journal Archive and Interchange Tag Set: 요소 254개, 속성 135개
- Journal Publishing Tag Set: 요소 246개, 속성 134개
- Article Authoring Tag Set: 요소 197개, 속성 123개

JATS의 하위 모형 중에서 전체 요소를 포함하는 Journal Archive and Interchange Tag Set을 대상으로 학술지와 게재 논문에 대한 구조와 메타데이터를 살펴보았다. 먼저 JATS 표준으로 태깅된 논문의 사례 일부를 살펴보면, <그림 1>과 같다. 이 그림에서 알 수 있는 것은 최상위 요소가 <article>이며, 하위 요소로 <front>, <body>, <back> 등이 있다. 이 그림에는 나와 있지 않지만, 나머지 하위 요소로 <floats-group>과 <response> 또는 <sub-article> 등이 있다.

JATS에서 학술지 게재 논문에 대한 메타데이터 요소를 선별하기 위해서는, 최상위 요소인 <article>에서 시작해야 하며, 이 최상위 요소에서 검토해야 할 하위 요소명은 <front>가 될 수 있다. <body>와 <back>의 경우 실제 논문의 본 내용과 참고문헌이나 부록에 해당하는

요소이므로 이 연구에서는 대상으로 삼지 않았다.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE article
PUBLIC "-//NLM//DTD JATS (Z39.96) Journal Publishing DTD v1.0 20120330//EN"
"JATS-journalpublishing1.dtd">
<article article-type="research-article" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:mml="http://www.w3.org/1998/Math/MathML">
  <front>
    <journal-meta>...</journal-meta>
    <article-meta>...</article-meta>
  </front>
  <body>
    <sec sec-type="results" id="s1">...</sec>
    ....
  </body>
  <back>
    <ack>...</ack>
    <fn-group>...</fn-group>
    <ref-list>...</ref-list>
  </back>
</article>
```

<그림 1> JATS의 논문 XML 사례(일부)

<그림 1>에서 점선으로 된 영역 안에 <front> 요소는 필수인데, 선택항목인 학술지에 대한 메타데이터(<journal-meta> 태그)와 필수항목인 논문의 메타데이터(<article-meta> 태그)를 포함하고 있다. 해당 태그의 주요 기술요소와 하위요소를 살펴보면 각각 <표 2>, <표 3>과 같다.

<표 2> JATS의 학술지(<journal-meta>) 관련 기술요소

요소명	하위 요소명	설명	필수:반복
journal-id		학술지식별자	0:M
journal-title-group	journal-title	본표제	0:M
	journal-subtitle	부표제	0:M
	trans-title, trans-subtitle	대등표제, 대등부표제	0:M
	abbrev-journal-title	축약표제	0:M
contrib-group, 또는 aff, 또는 aff-alternatives		기여 그룹(편집인) 기관명(조직명) 기관의 다른 명칭	0:M
issn		ISSN	0:M
issn-l		ISSN 링크	0:1
isbn		ISBN	0:M
publisher	publisher-name	발행자명	1*:1
	publisher-loc	발행지	0:1
notes		주기	0:M
self-uri		다른 버전 문헌용 URI	0:M

* 해당시 필수

〈표 3〉 JATS의 논문(〈article-meta〉) 관련 기술요소

요소명	하위 요소명	설명	필수:반복
article-id		논문식별자	0:M
article-categories	subj-group	주제범주	0:M
	series-title	총서표제	0:M
title-group	article-title	본표제	0:M
	subtitle	부표제	0:M
	trans-title, trans-subtitle	대등표제, 대등부표제	0:M
	alt-title	별표제	0:M
contrib-group, 또는 aff, 또는 aff-alternatives	name, degrees, address, aff (aff-alternatives), bio, email, role, uri	이름, 학위, 주소, 소속 (소속별칭), 약력, 이메일, 역할, uri	0:M
author-notes		저자 주기	0:1
pub-date		발행일	0:M
volume		권	0:M
volume-id		권 식별자	0:M
issue		호	0:M
issue-id		호 식별자	0:M
issue-title		호 제목	0:M
issue-sponsor		호 스폰서	0:M
issue-part		호 일부	0:1
volume-issue-group		권호 그룹	0:M
isbn		ISBN	0:M
supplement		추가자료	0:1
fpage 그리고, lpage 또는 page-range 또는, elocation-id		시작페이지, 끝 페이지, 페이지범위, 전자 주소 식별자	1:1 0:1 0:1 1:1
email 또는, ext-link 또는, uri 또는 product		이메일	0:M
		추가 링크	0:M
		URI	0:M
		제품 정보	0:M
history		기록정보	0:1
permissions	copyright-statement	저작권 문장	0:M
	copyright-year	저작권 년도	0:M
	copyright-holder	저작권자	0:M
	license	라이선스	0:M
self-uri		다른 버전 문헌용 URI	0:M
related-article 또는, related-object		관련 논문	0:M
		관련 객체	0:M
abstract		초록	0:M
trans-abstract		타언어 초록	0:M

kwd-group	kwd 또는, compound-kwd 또는, nested-kwd	키워드 복합 키워드 중첩 키워드	0:M 0:M 0:M
funding-group	award-group funding-statement open-access	수상 그룹 기금 설명 오픈액세스	0:M 0:M 0:M
conference	conf-date 또는, conf-name 또는, conf-num 또는, conf-loc 또는, conf-sponsor 또는, conf-theme 또는, conf-acronym 또는, string-conf	학술회의 개최일 학술회의명 학술회의 횟수 학술회의 지역 학술회의 스폰서 학술회의 주제 학술회의 약자 기타 학술회의명	0:M 0:M 0:M 0:M 0:M 0:M 0:M 0:M
counts	count fig-count table-count equation-count ref-count page-count word-count	일반적인 개수 그림 개수 표 개수 수식 개수 참고문헌 수 페이지 수 단어 수	0:M 0:1 0:1 0:1 0:1 0:1 0:1

* 해당시 필수

JATS에서 학술지 관련 주요 기술요소는 학술지식별자, 학술지 표제, 편집인, ISSN, 발행자 등이며, 논문 관련 주요 기술요소는 논문식별자, 논문 표제, 대등 표제, 저자 그룹, 논문이 수록된 권호 정보, 시작페이지, 키워드, 각종 통계 등이다. 다만, 앞서 기술하였듯이 JATS 표준은 NLM DTD를 기반으로 출발하였다. 이는 JATS의 이러한 메타데이터 기술요소가 NLM DTD과 같이 주로 과학, 기술, 의학(STM)분야 학술지와 이들에 게재되는 논문에 맞게 최적화 되어 있음을 의미한다.

3. 학술지 출판용 XML DTD

대부분의 학술지와 그에 수록된 논문은 큰 규모의 상업 출판사와 학회에 의해 발행되고 있다. 이들은 관련 콘텐츠를 생성하고 관리하기 위해 자신만의 XML(SGML) DTD를 가지고 있다. 즉 이들은 자신들의 출판용 XML DTD를 이용하여 학술지를 생산하고, 도서관 같이 자신들의 콘텐츠를 필요로 하는 기관에 배포하거나 공유한다.

출판사는 논문을 수록하는 학술지나 단행본 도서 등 자신들의 콘텐츠를 생산하기 위해 다양한 시스템과 업무 처리 절차를 가지고 있다. 몇몇 출판사는 콘텐츠 생산 초기부터 이러한 시스템이나 과정을 적용한다. 저자에게 논문의 표제나 초록 등을 따로 입력하게 하는 행위가 대표적인 사례이다. 이러한 상황에서 출판사는 출판용 XML DTD를 사용하여 콘텐츠 생산 초기 단계부터 메타데이터를 생성하고 저장한다.

출판사는 자신들의 XML DTD가 학술지 및 게재 논문과 같은 콘텐츠의 메타데이터와 데이터를 완벽하게 표현할 수 있도록 이를 지속적으로 발전시킨다. 왜냐하면 출판사가 XML DTD를 이용하여 자신의 콘텐츠를 인쇄하거나 전자출판 하고 서비스 하는데 반드시 필요하기 때문이다. 이러한 측면에서 본다면, XML DTD를 학술지와 게재 논문의 메타데이터 요소를 개발하는데 활용할 가치가 있다.

가장 많은 학술지를 생산하고 있는 대표적인 출판사는 Elsevier사와 Springer사이다. Elsevier사는 보유하고 있는 학술지가 3000종이 넘으며 Springer사는 2900종이 넘는다. 학술지와 게재 논문의 메타데이터 요소를 개발하기 위해, 이 연구에서는 이들 출판사가 보유하고 있는 XML DTD를 살펴보았다.

먼저 Elsevier사의 DTD는 1993년부터 2014년까지 지속적으로 개정하여 현재 5.4버전을 가지고 있다. 이 출판사의 DTD는 비교적 오래되었으며, 관련 정보나 문서도 많아 다른 기업이나 JATS에 영향을 미친 것으로 보인다. 이 DTD는 크게 논문, 정기간행물의 호, 단행본용으로 이루어져 있다(Elsevier 2015).

Elsevier에서 논문 관련된 DTD(Journal Article DTD 5.4.0)를 보면, 최상위 요소인 <article> 태그는 <item-info>, <head>, <body>, <tail> 등으로 구성된다. 본문에 해당하는 <body>태그를 제외하면 <item-info>는 논문에 대한 관리적 측면의 기본 정보를 담고 있으며, <head>는 논문에 대한 메타데이터를 담고 있다. <tail>의 경우 참고문헌과 용어, 개인 약력 정보 등을 담고 있다. 논문과 관련된 주된 메타데이터가 담긴 <item-info>와 <head> 태그에 대한 기술요소를 살펴보면, 각각 <표 4>, <표 5>와 같다(단, 네임스페이스는 생략).

<표 4> Elsevier XML DTD의 논문 기본 요소(<item-info>)

요소명	하위 요소명	설명	필수:반복
jid		학술지식별자	1:1
aid		논문식별자	0:1
article-number		논문번호	0:1
pii		publisher item identifier	1:1
doi		DOI	0:1
document-thread	doi 또는, pii와 doi	관련 문서 DOI, PII	0:1 0:1
copyright	copyright copyright-line	저작권	1:1 0:1
doctopics		논문 주제	0:1
preprint		프리프린트버전 링크	0:1

〈표 5〉 Elsevier XML DTD의 논문의 헤드 요소(<head>)

요소명	하위 요소명	설명	필수:반복
article-footnote		논문 표제지 각주	0:M
dochead		논문타입	0:1
title		본표제	1:1
subtitle		부표제	0:1
alt-title		별표제	0:M
alt-subtitle		별부표제	0:1
presented		학술대회 발표여부	0:1
dedication		헌정	0:1
author-group	collaboration 또는 author (initials, indexed-name, degrees, name, roles, e-address, link) affiliation correspondence footnote	공저 또는 저자명 (이니셜, 정렬된 이름 학위, 이름, 역할, 전자주소, 링크) 소속 교신여부 각주	1:1 0:M 0:M 0:M
date-received		접수일	0:1
daterevised		수정일	0:M
date-accepted		수락일	0:1
abstract	section-title abstract-sec figure	섹션 제목 초록 그림	0:1 1:M 0:1
keywords	section-title keyword	섹션 제목 키워드	0:1 1:M

Elsevier사의 XML DTD의 논문 관련 요소는 학술지와 논문 식별자, 다양한 표제 정보(본 표제, 부표제, 별표제 등), 저자그룹, 논문 리뷰 관련 정보(접수일, 수정일, 수락일) 등을 포함하고 있다.

Elsevier의 DTD에서 학술지에 대한 메타데이터는 거의 제시되지 않고 있는데, 이는 학술지 관련 메타데이터는 다른 영역에서 관리되고 있는 것으로 보인다. 또 이 DTD에서 학술지의 권호에 대한 기술 요소는 다른 DTD 파일(Serial Issue DTD 5.4.0)에서 규정하고 있다. 이 DTD에서는 최상위 요소는 <serial-issue>이다. 이 요소는 기본 관리 정보가 담긴 <issue-info>, 권호 관련 정보인 <issue-data>, 본문이 수록되는 <issue-body> 요소를 포함하고 있다. <issue-info> 요소에는 필수인 <pii>, 선택인 <doi>, 필수인 <jid>, 필수인 <issn>, 필수인 <volume-issue-number>, 선택인 <isbn> 요소가 포함되어 있다. <issue-data> 요소에는 호의 발행일인 <cover-date>(필수), 시작페이지와 끝페이지를 포함하는 <pages>(선택), 표지 이미지인 <cover-image>(선택), <issue-designation>(선택), <title-editors-group>(선택) 요소 등이 있다.

Springer사의 DTD는 2001년부터 개발되어 배포되고 있는데, 논문과 단행본 도서용의 DTD를 가지고 있으며 2011년에 2.4 버전으로 개발되었다. 웹상에서는 Springer사의 오픈 액세스 저널을 위한 DTD가 공개되어 있는데, 이 DTD는 버전(0.4)이 너무 낮고 정식 DTD 보다 요소의 수가 적다.

Springer사의 DTD에서는 발행자(<Publisher> 요소)와 학술지(<Journal> 요소), 그리고 호(<Issue> 요소)와 논문(<Article> 요소)이 순차적으로 제시되고 있어 JATS와 Elsevier사의 DTD와는 조금 다른 모습을 보인다. 즉 앞서 제시된 JATS와 Elsevier사의 DTD는 발행자나 학술지 관련 메타데이터가 상대적으로 적었으나 Springer사의 DTD는 이들에 대한 메타데이터를 다수 제시하고 있다. 이들의 학술지와 논문 관련 메타데이터를 나열하면 <표 6>, <표 7>, <표 8>과 같다.

<표 6> Springer XML DTD의 발행자 및 학술지 기본 요소

요소명	하위 요소명	설명	필수:반복
Publisher	PublisherName	발행자명	1:1
	PublisherLocation	발행지	1:1
	PublisherImprintName	발행사행 발행자명	0:1
	PublisherLogo	발행자 로고	0:1
	PublisherURL	발행자 URL	0:1
	CoPublisher	공동발행자	1:1
Journal	JournalID	학술지 식별자	1:1
	JournalPrintISSN	학술지 인쇄 ISSN	0:1
	JournalElectronicISSN	학술지 전자용 ISSN	0:1
	JournalISPIN	학술지 ISPIN	0:1
	JournalTitle	학술지 표제	1:1
	JournalSubTitle	학술지 부표제	0:1
	JournalAbbreviatedTitle	학술지 축약표제	1:1
	JournalSubjectGroup	학술지 주제그룹	0:1
	JournalSubject	학술지 주제	1:1
Volume	VolumeIDStart	권 식별자-시작	1:1
	VolumeIDEnd	권 식별자-종료	1:1
	VolumeIssueCount	권호 횟수	1:1
	TOC	목차	0:1
Issue	IssueIDStart	호 식별자-시작	1:1
	IssueIDEnd	호 식별자-종료	1:1
	IssueTitle	호 본표제	0:1
	IssueArticleCount	호 논문 수	0:1
	IssueHistory	호 기록	0:1
	OnlineDate	온라인 발행일	0:1
	PrintDate	종이 발행일	0:1
	CoverDate	표지 발행일	0:1
IssueCopyright	호 저작권	1:1	

<표 6>에서 발행자 요소인 <Publisher>는 발행자 관련 메타데이터 요소로 필수인 발행자명 <PublisherName> 요소, 필수인 발행지 <PublisherLocation> 요소, 선택의 발행자 URL <PublisherURL> 요소 등이 있다. 또한 학술지 관련 요소인 <Journal>은 학술지식별자, 인쇄용 ISSN, 전자용 ISSN, 학술지 본표제, 부표제, 축약표제, 학술지 주제범주 등을 담고 있으며 이중에서 학술지식별자, 학술지 본표제, 축약표제는 필수에 해당한다.

<Volume> 요소는 시작과 끝의 권 식별자(Volume ID)를 포함하며 권호 횡차 관련 정보, 권의 목차(<TOC> 요소) 등을 담고 있다. 또한 호에 관련 메타 정보를 제공하는 요소로 <Issue>는 시작과 끝의 호 식별자(Issue ID), 호의 본표제, 호에 수록된 논문의 수, 호의 저작권 등을 담고 있다. Springer사의 DTD는 권이나 호와 관련된 메타데이터 관련 내용을 충실히 보여주고 있음을 알 수 있다.

<표 7> Springer XML DTD의 논문 기본 요소

요소명	하위 요소명	설명	필수:반복
ArticleInfo	ArticleID	논문 식별자	1:1
	ArticleDOI	논문 DOI	0:1
	ArticleCitationID	논문 인용 식별자	0:1
	ArticleSequenceNumber	논문 일련번호	0:1
	ArticleRelatedObject	논문 관련 객체	0:M
	ArticleSuperTitle	논문 상위 표제	0:1
	ArticleTitle	논문 본표제	1:1
	ArticleSubTitle	논문 부표제	0:1
	ArticleClassification	논문 분류번호	0:1
	ArticleCategory	논문 범주	0:1
	ArticleSubCategory	논문 하위범주	0:1
	ArticleFirstPage	논문 첫페이지	0:1
	ArticleLastPage	논문 끝페이지	0:1
	ArticleHistory, Received, Revised, Accepted, OnlineDate	논문 기록, 접수일, 개정일, 수락일, 온라인 발행일	0:1, 0:1, 0:1, 0:1, 0:1
	ArticleEditorialResponsibility	논문 편집 책임	0:1
	ArticleCopyright	논문 저작권	1:1
	ArticleGrants	논문 관련기금	0:1
ArticleContext	논문 전후관계	0:1	
ArticleHeader	PageHeaders	페이지 헤더	0:1
	AuthorGroup	저자 그룹	1:1
	TocChapter	목차	0:1
	Abstract	초록	0:M
	KeywordGroup	키워드 그룹	0:M
	AbbreviationGroup	약어 그룹	0:1
	ArticleNote	논문 주기	0:M

논문 관련 요소인 <Article>은 다른 요소와 유사하게 <ArticleInfo>, <ArticleHeader>, 본문 관련 요소나 참고문헌용 요소, <ArticleBackmatter> 등을 포함한다. 다만 앞에 두 요소를 제외한 나머지는 본문이나 참고문헌용이라 제외한다. 따라서 <표 7>에 제시된 <ArticleInfo>는 논문식별자, DOI, 인용식별자, 논문번호, 논문 관련 객체 정보, 논문의 상위 표제, 논문 본표제, 부표제, 분류번호, 범주, 하위 범주, 논문 첫 페이지, 논문 끝 페이지, 논문 편집 책임자, 논문 저작권, 논문관련 기금 등을 포함한다. 여기에서 논문식별자, 본표제, 저작권만이 필수이다.

<표 8> Springer XML DTD의 저자 기본 요소

요소명	하위 요소명	설명	필수:반복
AuthorName	Prefix	경칭	0:1
	GivenName 또는 NoGivenName	이름	1:1
	Particle	파티클	0:1
	FamilyName	성	1:1
	Suffix	접미명	0:1
	Degrees	학위	0:1
	NativeName	모국어 이름	0:1
Role		역할	0:1
Contact	OrgDivision, OrgName, AddressModel	조직의 이름, 부서, 주소	0:1
	Phone, Fax, Email, URL	전화번호, 팩스, 이메일, URL	0:M
Biography		약력	0:1
Figure		사진	0:1
Affiliation	OrgDivision	조직의 부서	0:1
	OrgName	조직명	0:1
	Orgaddress	조직주소	0:1
	URL	URL	0:1

마지막으로 <표 8>에 제시된 Springer의 저자와 관련된 요소로는 저자명을 나타내는 <AuthorName>, 저자의 역할을 표현하는 <Role>, 연락처와 같은 정보를 다루는 <Contact>, 소속 정보를 포함하는 <Affiliation> 등을 포함한다. 저자의 이름과 성이 필수 요소로 지정되어 있다.

추가적으로 XML를 이용하여 학술지나 논문의 콘텐츠를 표현하는 대표적인 큰 규모의 학회 및 학술단체로 IEEE(Institute of Electrical and Electronics Engineers)와 ACM (Association for Computing Machinery)을 들 수 있다. 이들도 자신의 콘텐츠를 위한

DTD를 가지고 있고 비교적 규모가 작아 다음의 <표 9>에서 <표 12>의 출처 란에 바로 제시하였다. 이들 단체에 대해 간단히 설명하면 다음과 같다.

먼저 IEEE(2015)는 전기 및 전자분야의 이론 및 실용기술 발전을 위한 비영리 전문 학술 단체이다. 현재 규모나 권위에 있어서 세계에서 가장 우수한 학회에 속한다. 현재 전기 및 전자공학, 컴퓨터 공학, 생명공학, 통신 등 여러 분야에서 세계 논문의 30퍼센트를 발표하며, 매년 300개가 넘는 기술 컨퍼런스를 후원 또는 공동후원하고 있다. IEEE에 속한 학회에서 발간되고 저널, 잡지, 컨퍼런스 자료 등을 IEEE Xplore 서비스를 통해 제공하고 있다.

ACM(2015)는 컴퓨터 및 IT 관련 모든 분야에 대한 최신 정보를 다루는 학회로써, 44개의 특별 분야와 170여개의 지역 조직을 거느리고 저널, 잡지, 뉴스레터 등을 발간하며 정기 학회 및 컴퓨터 관련 이벤트를 개최하고 있다. ACM 관련 출판사에서 발행하는 자료에 전문(fulltext)을 제공하는 ACM Digital Library와 3,000개 이상의 컴퓨터 분야 주요 출판사에서 발간하는 서지정보를 제공하는 서비스가 주를 이룬다.

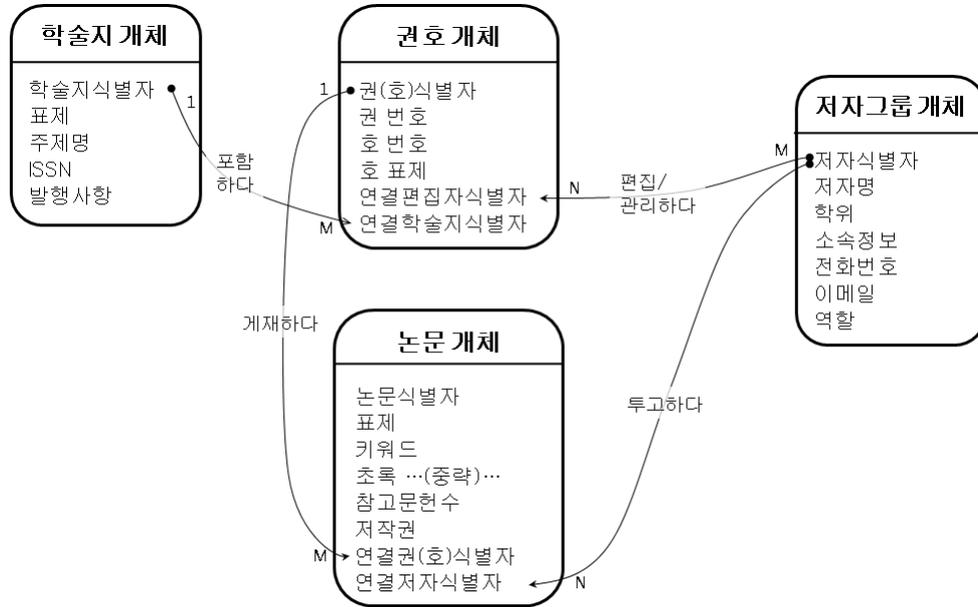
Ⅲ. 학술지와 논문을 위한 메타데이터 요소 개발

1. 개체 관계 및 요소 선별 기준

학술지와 논문을 대상으로 메타데이터를 개발하기 위해서는 학술지, 권호, 논문 그리고 저자의 관계를 규정해야 한다. 하지만, 이들 간의 관계를 어떻게 규정해야 하는지에 대해서는 다양한 의견이 있어 합의가 되지 않고 있다.

대표적으로 Shadle(2007)은 FRBR의 개체-관계 모형에서 전체/부분 관계로 표현하고 해석하고자 하였다. 이 연구에서는 학술지를 기술하기 위해 저작으로서의 특정 주제, 표현형으로서의 특정 주제를 다루는 학술지, 구현형으로서의 학술지의 각 호(issue)로 간주하였다. 또한 학술지에 실린 하나의 논문(article)은 그 자체가 특정 주제에 관련된 하나의 관련 저작으로 규정하고자 하였다. 또한 Taniguchi(2013)는 이러한 관계를 RDA 관점에서 해석하였는데, 학술지를 전체로 간주하고 논문을 개별개체(components)로 보고 학술지의 부분집합으로 규정하고자 하였다.

이와 같이 FRBR과 RDA에서 학술지와 논문 사이의 관계를 보다 정확하게 설정하기 위해, 목록 연구자들과 관련 종사자들이 연구를 진행하고 있다. 따라서 이 연구에서는 앞서 제시된 다수의 XML DTD에 근거하여 주요 4개 개체인 학술지, 권호, 게재 논문 그리고 저자의 관계(연결)를 설정하였다. 관계를 설정하기 위한 개체-관계 방법을 적용한 결과는 <그림 2>와 같다.



<그림 2> 학술지, 논문, 저자 개체의 관계

이 그림을 보면, 하나의 학술지는 다수의 권호를 포함하며, 각각의 권호에는 다수의 논문이 수록된다. 또한 학술지를 관리하는 편집자와 논문을 생산하여 투고하는 저자는 모두 저자그룹에 해당한다. 저자와 논문의 관계도 다수 대 다수 관계이며, 학술지와 편집진도 다대다 관계이다.

앞서 제시된 기술규칙이나 DTD의 모든 기술요소들을 선별하는 것은 문제가 된다. 왜냐하면 제시된 DTD의 모든 기술요소를 나열하는 것은 어떤 DTD는 특정 요소를 보다 자세히 기술하고 있거나 반대로 일부는 축소되어 선별에 도움이 되지 않기 때문이다. 학술지와 논문의 메타데이터 요소를 개발하기 위해 다음과 같은 기준을 적용하였다. 먼저 앞서 살펴본 목록규칙 중 RDA와 주요 XML DTD인 JATS, Elsevier XML DTD(줄여서 EL), Springer XML DTD(줄여서 SP), IEEE DTD(줄여서 IEEE), ACM DTD(줄여서 ACM)에서 3개 이상의 공통으로 사용된 기술요소를 선별하였다. 둘째 공통적인 요소가 아니라도 각 기술규칙 사례 중에서 주요 개체들을 유일하게 식별하기 위한 필요한 기술요소를 선별하였다. 셋째 각각의 기술규칙 사례에서 2회 이상 출현하는 필수 또는 해당시 필수인 기술요소를 우선 선별하였다. 또한 이를 이용하여 최종 개발된 메타데이터의 필수여부에 반영하였다. 마지막으로 앞서 제시된 대부분의 기술규칙이나 XML DTD가 외국에서 만들어지고 사용되어서 국내 실정에 맞지 않을 경우, 예외적으로 이를 반영하였다. 대표적인 예로 ‘대등표제’나 ‘대등키워드’가 여기에 해당한다.

XML DTD 기준으로 주요 개체의 기술 정도를 보면, 논문 메타데이터 요소는 JATS와 EL의 DTD에서 자세히 기술하였고, 학술지와 저자의 메타데이터 요소는 SP와 목록규칙에서 사

세히 기술하였음을 알 수 있다. 주요 개체의 관련 메타데이터 요소를 선별하는데 있어 이러한 측면을 고려하였다.

2. 학술지와 논문 관련 메타데이터 요소

이 연구에서는 앞서 제시된 기준에 근거하여 학술지, 권호, 논문, 그리고 저자에 대한 메타데이터 요소를 최종적으로 선별하였다. 이를 각각 표로 나타내면 <표 9>, <표 10>, <표 11>, <표 12>와 같다. 전체 선별된 메타데이터 요소는 모두 50개에 해당한다. 이들 요소에서 ‘저자식별자’와 같이 식별자 요소는 특정 개체에 대해 유일한 값을 갖도록 구성하며, ‘본표제’나 저자명 같이 세계의 각기 다른 언어로 표현될 수 있는 요소들은 언어 속성을 가져야 한다.

<표 9> 학술지 메타데이터 요소

요소	하위요소	기술 내용	필수	출처
학술지식별자		학술지를 식별하기 위한 기호	필수	JATS, SP, EL, IEEE, ACM
표제	본표제	학술지의 본표제	해필	JATS, SP, RDA, IEEE, ACM
	부표제	학술지의 부표제	재량	JATS, SP, RDA
	대등표제	다른 언어로 작성된 학술지의 대등표제	재량	JATS, RDA
	대등부표제	다른 언어로 작성된 학술지의 대등부표제	재량	JATS, RDA
	축약표제	학술지의 축약된 표제	재량	JATS, SP, RDA, IEEE
주제명		학술지의 주제명	재량	SP, RDA, ACM
ISSN		국제 표준 연속간행물 번호	해필	JATS, EL, SP, RDA, IEEE, ACM
발행사항	발행자	학술지를 발행한 기관	재량	JATS, SP, RDA, IEEE, ACM
	발행지	학술지를 발행한 지역	재량	JATS, SP, IEEE

<표 9>를 보면, 학술지 메타데이터 요소는 ‘표제’와 ‘발행사’ 등 총 10개를 선별하였다. 필수 요소는 ‘학술지 식별자’이며 해당시 필수(해필)은 ‘본표제’와 ‘ISSN’이고 나머지는 재량으로 제시하였다. ‘본표제’의 경우 EL이나 RDA에서 필수 요소로 제시되고 JATS에서는 선택으로 표시되었기에 해당시 필수로 설정하였다. 마찬가지로 해당시 필수인 ‘ISSN’도 EL에서 필수로 되어 있다. 다만, Elsevier사 DTD(EL)의 경우 학술지 관련 요소가 없기에, 출처에서 거의 빠졌음을 알 수 있다. 이 표에서 제시되지 않았지만, ‘e-issn’ 요소와 ‘발행자 URL’ 또는 유사한 ‘URI’ 요소는 2개의 DTD에서 출현하여 추가를 고려해 볼 수 있다.

‘표제’와 관련하여 다수의 하위요소가 존재하며, 특히 영어권이 아닌 다른 나라에서는 그 나라의 고유 언어와 영어를 병기해야 하기에 해당 요소들이 선택되었다. 이미 선별 기준에서 설명하였듯이, 이러한 예로 ‘대등표제’와 ‘대등부표제’를 들 수 있다. 외국의 유명 출판사는

주로 하나의 언어로 된 표제를 사용하여 대등표제나 대등부표제가 거의 나타나지 않는다. 이와 달리 우리나라에서 대등표제나 대등부표제가 달린 학술지가 많으므로 구분하여 선별하였다.

<표 10> 권호 메타데이터 요소

요소	하위요소	기술 내용	필수	출처
권		권의 번호	해필	JATS, SP, RDA, IEEE, ACM
권식별자		권 식별자	재량	JATS, IEEE
호		호의 번호	해필	JATS, SP, RDA, IEEE, ACM
호식별자		호 식별자	재량	JATS, IEEE, ACM
호 표제		호(특별호)의 표제	재량	JATS, SP, IEEE
연결편집자식별자		저자(편집자) 개체와 연결을 위한 식별자	필수	-
연결학술지식별자		학술지 개체와 연결을 위한 식별자	필수	-

<표 10>을 보면, 권호 관련한 메타데이터 요소는 ‘권’/‘호’ 번호 등 총 7개를 선별하였으며, 해당시 필수(해필)은 ‘권’ 번호와 ‘호’ 번호이고 나머지는 재량으로 제시하였다. ‘권’과 ‘호’의 번호 또는 식별자는 학술지가 어떤 행태로 발행되느냐에 따라 달라질 수 있다. 즉 학술지가 권으로만 발행되면 그 학술지에 대한 호 관련 요소(‘호’와 ‘호식별자’)는 존재하지 않게 된다. 여기에 추가적으로 2개의 DTD에서 출현한 해당 권호에 실린 총 논문의 수를 의미하는 ‘총 논문 수’와 권호의 발행 날짜와 관련된 ‘발행일’ 요소가 있다.

앞서 주요 개체들(학술지, 권호, 논문, 저자) 사이에 관계(연결)를 제시하였는데, <표 10>의 마지막 행의 ‘연결학술지식별자’는 <표 9>의 ‘학술지식별자’와 연결(링크)되는 것으로 특정 학술지의 권 또는 호를 의미한다. 또한 학술지 권호의 발행을 지속적으로 책임지는 기존의 편집진이나 특별한 편집자를 위한 메타데이터가 필요한데, 이는 저자 개체의 특정 ‘저자식별자’ 요소를 참조한 ‘연결편집자식별자’ 요소로 기술하였다.

‘호(특별호) 표제’는 학술지가 특정 주제에 관한 논문만을 게재하기 위해 특별호를 발행하는 경우 등에서 호에 표제(title)를 부여할 수 있다. 1개의 DTD나 목록규칙에서 나왔지만 고려해 볼 만한 메타데이터 요소로는 ‘첫 쪽수’, ‘끝 쪽수’, ‘총 페이지 수’ 등이 있다. 매 권호가 1쪽으로 시작하는 경우도 있지만, 한해에 나오는 권호에 대해 쪽수를 누적하는 경우도 있어 시작 쪽수에 해당하는 ‘첫 쪽수’와 마지막 쪽수에 해당하는 ‘끝 쪽수’를 나타낸다. 마찬가지로 해당 권호의 전체 쪽수에 해당하는 ‘총 페이지 수’가 있다.

<표 11> 논문 메타데이터 요소

요소	하위요소	기술 내용	필수	출처
논문식별자		논문을 식별하기 위한 기호	필수	JATS, EL, SP, IEEE, ACM
표제	본표제	논문의 본표제	필수	JATS, EL, SP, IEEE, ACM
	부표제	논문의 부표제	재량	JATS, EL, SP, ACM
	대등표제	다른 언어로 작성된 논문의 대등표제	재량	JATS, ACM
	대등부표제	다른 언어로 작성된 논문의 부대등표제	재량	JATS, ACM
	별표제	논문의 별표제	재량	JATS, EL
타입		논문의 타입	재량	SP, IEEE, ACM
주제		논문의 주제	재량	JATS, EL, SP, IEEE, ACM
키워드	키워드	본문 언어의 키워드	해필	JATS, EL, SP, IEEE, ACM
	대등키워드	다른 언어로 작성된 키워드	재량	-
초록	초록	본문 언어의 초록	재량	JATS, EL, SP, IEEE, ACM
	대등초록	다른 언어로 작성된 대등초록	재량	JATS
발행관련일자	접수일, 수정일, 수락일	논문이 편집진에게 수락된 날짜	재량	JATS, EL, SP, IEEE, ACM
DOI		DOI	재량	SP, IEEE, ACM
URI		URI	재량	JATS, IEEE, ACM
논문순서		권/호에 실린 논문의 순서	재량	SP, IEEE, ACM
쪽수	첫 쪽수	논문의 첫 쪽수	재량	JATS, SP, IEEE, ACM
	끝 쪽수	논문의 끝 쪽수	재량	JATS, SP, IEEE, ACM
참고문헌수		논문에 수록된 참고문헌의 수	재량	JATS, IEEE, ACM
저작권		논문에 표시된 저작권 정보	해필	JATS, EL, IEEE, ACM
연결권호식별자		권호 개체와 연결을 위한 식별자	필수	-
연결저자식별자		저자 개체와 연결을 위한 식별자	해필	-

<표 11>를 보면 논문 관련 메타데이터 요소는 ‘표제’, ‘키워드’, ‘초록’과 ‘DOI’ 등 총 22개를 선별하였다. 필수 요소는 ‘논문식별자’와 ‘본표제’이며, 해당시 필수은 ‘키워드’와 ‘저작권’이고 나머지는 재량으로 제시하였다.

학술지에 게재된 논문을 다른 주요 개체인 학술지의 권호와 저자 사이에 관계를 제시하기 위하여 <표 11>의 ‘권호식별자’와 <표 12>의 ‘저자식별자’를 각각 ‘연결권호식별자’와 ‘연결저자식별자’ 요소로 추가하였다. 이는 특정 논문이 학술지의 특정 권호에 게재되었으며, 한명 또는 다수의 저자에 의해 쓰여진 것을 의미한다. 다만 출판사의 DTD를 보면 경우에 따라 저자가 필수가 아닌 경우가 있어 해당시 필수로 하였다.

앞서 설명한 바와 같이 우리나라의 상황을 고려하여 ‘대등표제’나 ‘대등부표제’, ‘별표제’, ‘키워드’나 ‘대등키워드’, ‘초록’나 ‘대등초록’ 등은 구분하여 선별하였다. 특히 ‘대등키워드’ 요소의 경우 어느 DTD에서도 제시되지 않았지만, 선별기준에서도 설명되었듯이, 국내 실정의

경우 많은 사례가 이를 보여주고 있어 추가하였다.

‘논문순서’나 ‘참고문헌수’와 같은 독특한 요소가 있다. ‘저작권’의 경우 분석된 XML DTD가 상업적인 출판사나 학회에 해당하므로, 이들에게 학술지와 논문 콘텐츠는 매우 중요한 지적 재산이므로 반드시 표기하는 것으로 되어 있어 해당시 필수로 되었다.

〈표 12〉 저자 메타데이터 요소

요소	하위요소	기술 내용	필수	출처
저자식별자		저자를 식별하기 위한 기호	필수	JATS, ACM
저자명	성	저자의 성	해필	JATS, EL, SP, IEEE, ACM
	이름	저자의 이름	해필	JATS, EL, SP, IEEE, ACM
	경칭	이름 앞에 붙이는 경칭	재량	JATS, EL, SP
	접미명	이름 뒤에 붙이는 명칭	재량	JATS, SP, ACM
학위		저자가 받은 학위	재량	JATS, EL, SP
소속	기관명	저자가 소속된 기관명	재량	JATS, EL, SP
	주소	저자가 소속된 기관의 주소	재량	JATS, EL, SP
전화번호		저자의 전화번호	재량	JATS, EL, SP, IEEE
이메일		저자의 이메일	재량	JATS, EL, SP, IEEE, ACM
역할		저자가 기여한 역할	재량	JATS, EL, SP, ACM

〈표 12〉를 보면, 저자 관련 메타데이터 요소는 저자명 등 총 11개를 선별하였다. 필수 요소는 ‘저자식별자’이며, 해당시 필수는 저자의 ‘성’과 ‘이름’이고 나머지는 재량으로 제시하였다.

저자 관련 메타데이터 요소의 경우 개인 저자와 단체 저자가 모두 해당한다. 또한 이 저자 개체는 논문에 기여하는 저자도 있지만, 학술지를 편집하고 관리하는 편집자도 있다. 이들을 동명이인으로부터 전거하기 위해 저자식별자를 적용하였다. 개인이나 단체에 소속과 연락처 정보를 위한 ‘소속’ 요소나 기여정도나 종류를 구분하기 위한 ‘역할’ 요소가 있다.

IV. 결론

학술지와 관련된 정보는 지속적으로 증가하고 있어, 이를 체계적으로 관리하고 서비스해야 한다. 우리나라에서는 학술지와 학술지에 수록된 논문의 메타데이터 품질에 관한 연구는 다수 존재하나 이들의 메타데이터 스키마에 관한 연구는 소수에 불과하다. 학술지와 관련하여 보다 좋은 서비스를 제공하려면 이들로부터 어떤 메타데이터 요소를 추출해야 하는지 연구할

필요가 있다.

이 연구에서는 학술지와 관련하여 메타데이터 요소를 추출하기 위해, 우선 도서관의 목록 규칙에서 학술지에 대한 관련된 규정과 기술요소를 살펴보았다. 논문 관련 미국 표준인 JATS를 분석하여 메타데이터 구조와 기술요소를 제시하였다. 상업용 출판사(Elsevier, Springer, IEEE, ACM)의 XML DTD는 특수성으로 인해 다양한 기술요소를 포함하므로, 이들도 분석에 활용하였다.

이 연구는 학술지와 게재 논문에 대한 메타데이터 스키마로 50개의 기술요소를 제안하였으며, 메타데이터의 선별적인 선택이 가능하도록 7개의 필수요소, 9개의 해당시 필수요소, 34개의 재량요소로 구분하여 제시하였다. 이 중에서 학술지 개체 관련 기술요소는 10개, 권호는 7개, 논문은 22개, 그리고 저자는 11개에 해당한다. 이러한 학술지와 논문을 위한 메타데이터 기술요소의 연구결과를 바탕으로 학술 데이터베이스 업체나 기사색인 서비스에 활용하면 보다 높은 품질의 메타데이터를 통해 더욱 좋은 검색과 접근을 이용자에게 제공할 수 있을 것으로 생각된다.

이 연구는 국내의 학술지와 논문 관련 메타데이터 스키마가 존재하지 않아 이를 대체하기 위한 방법으로 논문관련 태깅 표준이 JATS와 상업 출판사 및 학회의 출판용 DTD를 활용하여 메타데이터 기술 요소를 선별하였다는데 의의를 둘 수 있다. 또한 우리는 학술지나 논문에 영어나 타 언어를 한국어와 병기하는 사례가 많은데 비해, 외국은 주로 하나의 언어로 된 표제나 관련 정보를 사용하여 타언어로 된 기술요소가 거의 없어 앞으로 메타데이터 관련해서 이러한 측면을 고려해야 할 필요가 있다.

이 연구의 제한점으로는 5개의 XML DTD를 바탕으로 메타데이터 기술요소를 선별하였는데, 더 많은 수의 관련 DTD를 이용하여 연구결과를 보다 일반화할 필요가 있다. 또한 관련 업체와 도서관, 더 나아가 국가도서관 차원에서 검토하여 실제로 적용가능한지 살펴볼 필요가 있다.

참고문헌

- 김수정, 김용. 2013. 무형문화유산 기록물을 위한 FRBR 기반 메타데이터 요소에 관한 연구. 『정보관리학회지』, 30(2): 119-141.
- 김관준. 2010. 연구정보를 위한 보존 메타데이터 요소 개발에 관한 연구. 『정보관리학회지』, 27(4): 169-191.
- 문헌정보학용어사전 편찬위원회 편. 2010. 『문헌정보학용어사전』. 개정판. 서울: 한국도서관협회.
- 박옥남. 2012. PREMIS 기반 보존 메타데이터 요소 개발에 관한 연구- 국립중앙도서관 디지털 자료를 중심으로. 『한국문헌정보학회지』, 46(2): 83-113.

- 박진희. 2009. RDA의 제정동향 및 내용구조에 관한 연구. 『한국도서관·정보학회지』, 40(1): 317-338.
- 윤정옥. 2003. 연속간행물 종합목록 데이터베이스의 레코드 품질 평가. 『한국문헌정보학회지』, 37(1): 27-42.
- 이은철, 이상복, 오삼균, 박옥남. 2009. 정기간행물 기사색인 서비스 현황 및 발전방향에 대한 연구. 『한국문헌정보학회지』, 43(1): 189-212.
- 조재인. 2009. RDA 구현 시나리오와 서지 제어의 장래에 대한 고찰. 『한국문헌정보학회지』, 43(3): 85-105.
- 한성국, 이현실. 2005. 학술지 메타데이터의 표준화 체계에 관한 비교 연구. 『한국도서관·정보학회지』, 36(1): 415-440.
- 한종엽. 2000. 연속간행물 기사 DB의 서지데이터요소 표준화 연구. 『한국문헌정보학회지』, 34(3): 109-132.
- Association for Computing Machinery [ACM]. 2015. <<http://www.acm.org/>> [cited 2015. 3. 4].
- Beck, Jeffrey. 2011. "NISO Z39.96 The Journal Article Tag Suite (JATS): What Happened to the NLM DTDs?." *The Journal of Electronic Publishing*, 14(1). <<http://quod.lib.umich.edu/j/jep/3336451.0014.106/--niso-z3996-the-journal-article-tag-suite-jats-what-happened?rgn=main;view=fulltext>> [cited 2015. 1. 10].
- Elsevier. 2015. Elsevier XML DTDs and Transport Schemas. <<http://www.elsevier.com/author-schemas/elsevier-xml-dtds-and-transport-schemas>> [cited 2015. 1. 5].
- Institute of Electrical and Electronics Engineers [IEEE]. <<http://www.ieee.org/index.html>> [cited 2015. 3. 5].
- Maxwell, Robert L. 2004. *Maxwell's Handbook for AACR2: Explaining and Illustration the Anglo-American Cataloguing Rules through the 2003 Update*. 4th ed. 오동근 옮김. 2005. 『영미편목규칙 제2판 핸드북』. 대구: 태일사.
- OCLC. 2015. Special Cataloging Guidelines. <<http://www.oclc.org/bibformats/en/special-cataloging.html#CHDBIIAF>> [cited 2015. 2. 4].
- Shadle, S. 2007. *FRBR and Serials: One Serialist's Analysis*. In Arlene G. Taylor (Ed.), *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools*. Westport, CT: Libraries Unlimited.
- Taniguchi, Shoichi. 2013. "Aggregate and Component Entities in RDA: Model and Description." *Cataloging & Classification Quarterly*, 51(5): 580-599.

- The Joint Steering Committee for Development of RDA. 2010. *RDA: Resource Description & Access*. Chicago: American Library Association.
- The National Information Standards Organization. 2012. JATS: Journal Article Tag Suite. Baltimore, Maryland: the National Information Standards Organization <http://www.niso.org/apps/group_public/download.php/10591/z39.96-2012.pdf> [cited 2015. 1. 15].
- The Standing committees of the IFLA section on Cataloguing and the IFLA Section on Serial Publications. 2002. *ISBD(CR) : International Standard Bibliographic Description for serials and other continuing resources*. Munchen: K. G. Saur.

국한문 참고문헌의 영문 표기

(English translation / Romanization of reference originally written in Korean)

- Cho, Jane. 2009. "A Study for RDA Implementation Scenarios and Future Bibliographic Control." *Journal of the Korean Society for Library and Information Science*, 43(3): 85-105.
- Han, Jong-Yup. 2000. "A Study on Standards for Bibliographic Data Elements of the Articles in Serial." *Journal of the Korean Society for Library and Information Science*, 34(3): 109-132.
- Han, Sung-Kook & Lee, Hyun-Sil. 2005. "A Comparative Study of Metadata Standards for Serials." *Journal of Korean Library and Information Science Society*, 36(1): 415-440.
- Kim, Pan-Jun. 2010. "A Study on Preservation Metadata Elements for Research Information." *Journal of Korea Society for Information Management*, 27(4): 169-191.
- Kim, Soojung & Kim, Yong. 2013. "A Study on the Development of Metadata Elements for Intangible Cultural Heritage Records Based on FRBR." *Journal of Korea Society for Information Management*, 30(2): 119-141.
- Korean Library Association. 2003. *Korean Cataloguing Rules*. 4th ed. Seoul: Korean Library Association.
- Lee, Eunchul, Lee, Sangbok, Oh, Sam Gyun, & Park, Ok Nam. 2009. "A Study of Ways to Improve Periodical Indexing Services in Korea." *Journal of the Korean Society for Library and Information Science*, 43(1): 189-212.

- Library Terminology Committee. 2010. *The Glossary of Library and Information Science*. Revised Ed. Seoul: Korean Library Association.
- Park, Jin-Hee. 2009. "A Study on Movements to Establish RAD and Its Contents Structure." *Journal of Korean Library and Information Science Society*, 40(1): 317-338.
- Park, Ok Nam. 2012. "A Study on Developing Preservation Metadata Based on PREMIS Focusing on Digital Data in National Library of Korea." *Journal of the Korean Society for Library and Information Science*, 46(2): 83-113.
- Yoon, Cheong-Ok. 2003. "Evaluation of the Quality of Records of the Serials Union Catalog Database." *Journal of the Korean Society for Library and Information Science*, 37(1): 27-42.