

# 메타데이터 기반 분담목록시스템 모형 구축에 관한 연구

## -국내 대학 학술지 논문을 중심으로

### A Study on the Model for Shared Cataloging System Based on Metadata

박종섭, 이응봉

충남대학교 대학원 문헌정보학과

Jong-Seop Park, Eung-Bong Lee

Graduate School of Library & Information Science, Chungnam National University

인터넷과 웹 기술의 발전은 전자도서관의 목록정보나 목차정보서비스에서 원문정보서비스로의 변화를 주도하고 있다. 그러나 대학에서 생산하는 학문적 가치가 높은 학술지들은 웹을 통해 원문으로 서비스되고 있지만 상이한 OPAC 시스템과 목록을 위한 통일된 메타데이터의 부재로 원문정보를 공유·활용하기가 매우 힘든 실정이다. 본 고에서는 각 대학에서 생산하고 구축된 원문정보를 메타데이터 기반의 목록시스템을 이용하여 모든 대학의 학술지 논문정보를 검색하여 공유할 수 있는 분담목록시스템 모형을 제시함으로써 대학 학술지 이용의 효율적 방안을 모색해보고자 한다.

## 1. 서론

### 1.1 연구목적

인터넷 기술의 빠른 성장과 웹 자원의 급속한 증가는 정보를 찾는 데 많은 노력과 시간을 소비하게 한다. 따라서 웹 자원에서 필요한 정보만을 찾는 다양한 연구와 노력들이 시도되고 있으며 메타데이터는 웹 정보자원의 표준화를 통한 자원간의 상호운용성과 검색효율 증대를 위한 자원기술방법으로 각광받고 있다. 최근, 국내 대다수의 대학도서관들이 각 대학에서 발

행되는 학술지, 학위논문, 연구보고서 등을 웹을 통하여 원문정보서비스를 제공하고 있지만, 대학마다 서로 다른 자원기술방식과 원문제공 방법을 채택하고 있기 때문에 표준화된 검색시스템이나 원문제공시스템의 개발이 시급한 실정이다.

특히 인터넷을 통하여 서비스되는 대학학술원문정보들은 각 대학의 검색시스템이나 한국교육학술정보원(KERIS) 시스템을 이용해야만 검색이 가능하다. 따라서 학술적 가치가 높은 학술지 논문의 신속한 원문 제공과 이용 확대 차원에서 접근경로를 개선하고, 인터넷 검색엔진에서도 효율적으로 탐색될 수 있도록 메타데

이더 활용방안을 모색하여야 한다.

웹 환경에서는, MARC 포맷이나 텍스트 형식의 관계형 데이터베이스를 활용한 원문검색 시스템들은 특정 검색시스템에 의존적이므로 탐색 시에는 정보제공 사이트에서만 이용이 가능하고 인터넷상에서의 상호운용성이나 호환성 등을 명확하게 보장하고 있지 못한 실정이다.

따라서 본 고에서는 대학 학술정보에 대한 공동활용 및 통합검색을 위하여 더블린코어 메타데이터를 기반으로 한 대학 학술정보의 분담 목록시스템 모형을 제안하고자 한다.

## 1.2 연구방법 및 제한점

본 연구를 위하여 국내 대학에서 발행된 인문과학, 사회과학, 자연과학 분야의 학술지와 각종 학술기관에서 발행된 인쇄형태의 학술논문기사색인을 대상으로 각 학술지에 수록된 논문의 서지정보와 기사색인목록에 수록된 서지정보들을 분석한 후, 학술지 논문에 사용된 일반화된 서지정보의 종류와 형식을 선별하여 더블린코어 메타데이터의 데이터요소들과 대응시켜 보았다. 학술논문의 목록을 위하여 데이터 요소를 입력할 수 있는 템플릿 모형을 만들고, 입력된 요소들은 XML 웹 문서로 변환, 생성되도록 하였다. 생성된 XML 웹 문서들은 웹 환경에서 키워드나 데이터요소별로 검색할 수 있도록 인터페이스를 설계하였다.

그러나 학술지 및 학술논문 기사색인의 발행기관별로 다양한 수록양식과 목록형식을 제공하고 있으므로, 학술논문의 목록을 위한 요소는 가장 일반적인 형태의 요소를 선정하였음을 미리 밝혀 둔다.

## 2. 학술지 논문을 위한 데이터 요소

더블린코어 메타데이터(Dublin Core Metadata)란, 웹 자원이 지닌 가장 중요한 특

성인 자원이 분산되어 있다는 점과 동일 자원이 다양한 형식으로 표현될 수 있다는 점이다. 이로 인해 이들 자원의 소재를 식별하고 접근하기 위해 지금까지 주제분야마다 독자적인 메타데이터 형식이 사용되어 왔으며 자원의 접근에 제약을 받아 왔다. 따라서 이들 자원에 대한 데이터의 호환성을 유지하고 신속한 검색을 목적으로 더블린에서 개최된 회의에서 자원의 기술에 필요한 일련의 핵심 데이터 요소를 규정한 메타데이터를 더블린코어 메타데이터라고 한다(정보학사전, 2001).

1995년 OCLC와 NCSA 주체로 열린 제1차 더블린코어 워크숍에서는 13개의 메타데이터 핵심요소를 제안하였으며, 1995년 제2차 UKOLN과 OCLC 주제 워크숍에서는 2개 요소를 추가하여 15개의 기본 데이터요소를 확정하여 사용하고 있다.

더블린코어 메타데이터는 15개 기본요소 외에 각 요소를 세부적으로 기술할 수 있도록 분야별로 사용되고 있는 한정어(qualifiers)를 제정함으로써 메타데이터간의 범용성과 호환성을 유지하고 있다. 한정어는 크게 두 가지 범주로 구분되는데, 요소의 의미를 좁히거나 더욱 특정화시키는 상세구분 한정어와 요소의 값에 대한 해석을 도와주는 인코딩 스킴(encoding scheme)이 있다(한국 DCMI, 2002).

본 고에서는 웹을 통하여 원문정보를 서비스하고 있는 학술지 논문에 기술된 서지정보들을 조사하고, 학술지 논문의 목록을 위한 서지항목들을 정리하여 더블린코어 데이터요소들과 대응시켜 보았다.

더블린코어 데이터요소 중에서 title, creator, subject, description, publisher, date, type, format, identifier, source, language 등 11개 요소와 대학별 소장정보를 기록하기 위한 별도의 소장기관(holdings) 요소를 추가하여 12개 요소를 선정하였다. 각 요소에는 요소를 식별

[표 1] 학술지 논문 메타데이터 요소 및 한정어

데이터요소	한정어	설명	비고
Title	DC.Title	논문제목	필수
	DC.Title.Parallel	논문대등제목	선택
	DC.Title.Other	기타제목	선택
Creator	DC.Creator.PersonalName	개인저자명	필수
	DC.Creator.PersonalName.Alternative	개인저자의 다른 이름	선택
	DC.Creator.PersonalName.Affiliation	개인저자의 소속	선택
	DC.Creator.PersonalName.Email	개인저자의 email	선택
	DC.Creator.PersonalName.Homepage	개인저자의 Homepage	선택
	DC.Creator.CorporateName	단체저자명	필수
	DC.Creator.CorporateName.Alternative	단체저자의 다른 이름	선택
	DC.Creator.CorporateName.Address	단체저자의 주소	선택
	DC.Creator.CorporateName.Email	단체저자의 email	선택
	DC.Creator.CorporateName.Homepage	단체저자의 Homepage	선택
Subject	DC.Subject	주제어(키워드)	선택
Description	DC.Description.Abstract	초록	선택
	DC.Description.Content	목차	선택
Publisher	DC.Publisher	발행자명	필수
	DC.Publisher.Alternative	발행자의 다른 이름	선택
	DC.Publisher.Address	발행자의 주소	선택
	DC.Publisher.Email	발행자의 email	선택
Date	DC.Publisher.Homepage	발행자의 Homepage	선택
	DC.Date.Created(ISO8601)	발행년	필수
Type	DC.Date.Metadata.Created(ISO8601)	메타데이터 작성일	필수
Format	DC.Type	자원의 장르	필수
Identifier	DC.Format	자원의 형태	선택
Source	DC.Identifier	자원의 위치정보(URL)	선택
	DC.Source.Name	학술지명	필수
	DC.Source.Issn	학술지 ISSN	선택
	DC.Source.Alternative	학술지의 다른 이름	선택
	DC.Source.Frequency	학술지 간기	선택
	DC.Source.Vol	학술지 권호	필수
Language	DC.Source.Pp	수록페이지	필수
Holding	DC.Language	기술언어	선택
	DC.Holding	소장기관	필수

하거나 상세하게 구분하기 위한 한정어를 두고 있으며, 학술지 논문의 목록을 위하여 필수항목과 선택항목을 구분하여 사용하였다.

특히 소장기관 요소는 각 대학에 산재해 있는 학술지 논문을 통합검색 또는 인터넷 검색 엔진을 통한 검색을 수행하였을 때 명확한 소장정보를 알려주기 위하여 별도로 추가한 요소이다.

### 3. 학술지 논문 목록용 DTD 작성

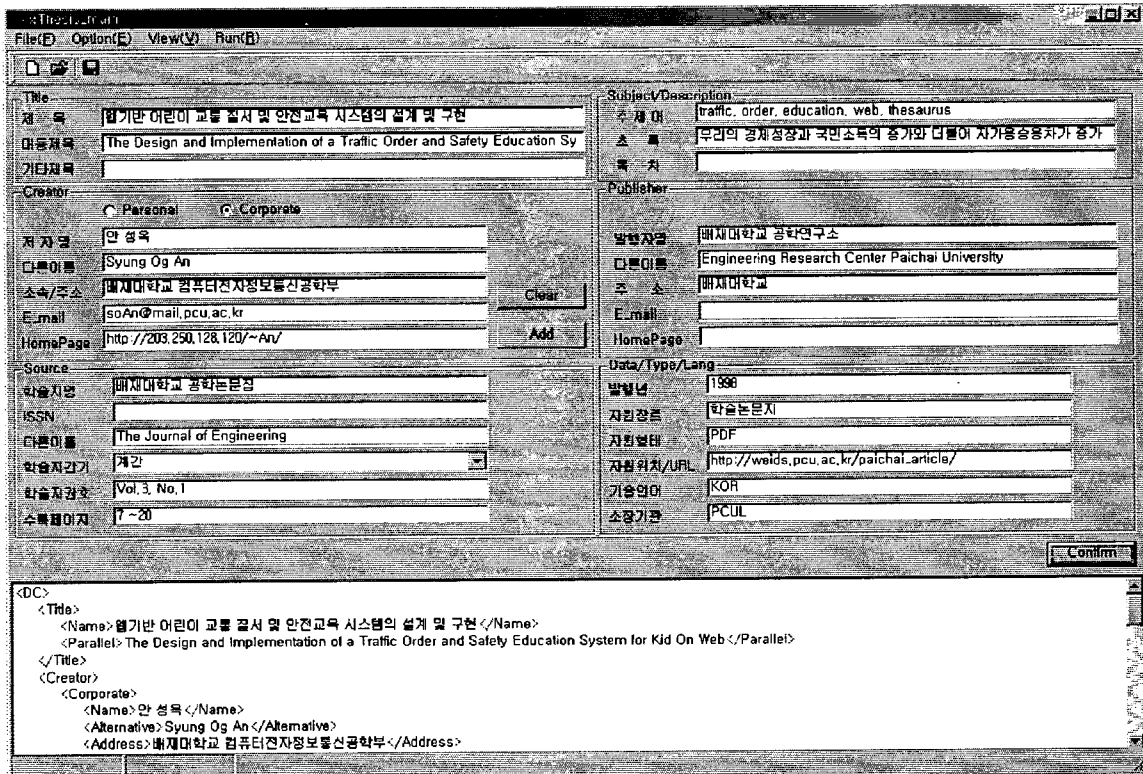
상이한 메타데이터로 기술된 웹 자원에 효과적으로 접근하기 위한 워릭구조(Warwick Framework)의 개념을 구체적으로 실현하는 수단으로 웹 컴소시엄(W3C)에서 자원기술구조(Resource Description Framework: RDF)를 제안하고 있는데, RDF 데이터 모형은 메타데이

```

<!-- 학술지 논문 목록용 XML DTD 설계 -->
<!-- ===== XML 선언문 ===== -->
<?xml version="1.0" encoding="euc-kr"?>
<!-- ===== DTD(문서형 정의) ===== -->
<!ENTITY % doctype "DC">
<!ENTITY % e.categoryItem "Title, Creator, Subject?, Description?, Publisher, Date, Type, Format, Identifier, Source, Language, Holding">
<!ENTITY % e.title "Name, Parallel?, Other?">
<!ENTITY % e.peopleScope "Personal+, Corporate+">
<!ENTITY % e.man "Name, Alternative?, Affiliation?, Email?, Homepage?, Role?">
<!ENTITY % e.source "Name, ISSN?, Alternative?, Frequency?, Vol?, Pp?">
<!--===== 범주별 Element 정의 ===== -->
<!ELEMENT %doctype; (%e.categoryItem;)><!-- 루트 DC에 대한 정의 -->
<!ELEMENT Title (%e.title;)> <!-- 범주의 Title 에 대한 정의 -->
<!ELEMENT Parallel (#PCDATA)> <!-- 범주의 Parallel 에 대한 정의 -->
<!ELEMENT Other (#PCDATA)> <!-- 범주의 Other 에 대한 정의 -->
<!ELEMENT Creator (%e.peopleScope;)> <!-- 범주의 Creator 에 대한 정의 -->
<!ELEMENT Personal (%e.man;)> <!-- 범주의 Personal 에 대한 정의 -->
<!ELEMENT Corporate (%e.man;)> <!-- 범주의 Corporate 에 대한 정의 -->
<!ELEMENT Name (#PCDATA)> <!-- 범주의 Name 에 대한 정의 -->
<!ELEMENT Alternative (#PCDATA)> <!-- 범주의 Alternative 에 대한 정의 -->
<!ELEMENT Affiliation (#PCDATA)> <!-- 범주의 Affiliation 에 대한 정의 -->
<!ELEMENT Email (#PCDATA)> <!-- 범주의 Email 에 대한 정의 -->
<!ELEMENT Homepage (#PCDATA)> <!-- 범주의 Homepage 에 대한 정의 -->
<!ELEMENT Role (#PCDATA)> <!-- 범주의 Role 에 대한 정의 -->
<!ELEMENT Subject (#PCDATA)> <!-- 범주의 Subject에 대한 정의 -->
<!ELEMENT Description (Abstract?, Content?)><!-- 범주의 Description 에 대한 정의 -->
<!ELEMENT Abstract (#PCDATA)> <!-- 범주의 Abstract 에 대한 정의 -->
<!ELEMENT Content (#PCDATA)> <!-- 범주의 Content 에 대한 정의 -->
<!ELEMENT Publisher (%e.man;)> <!-- 범주의 Publisher에 대한 정의 -->
<!ELEMENT Contributor (%e.peopleScope;)> <!-- 범주의 Contributor 에 대한 정의 -->
<!ELEMENT Date (Created_Year, MetaData_Created_Date)><!-- 범주의 Date 에 대한 정의 -->
<!ELEMENT Created_Year (#PCDATA)> <!-- 범주의 Created_Year 에 대한 정의 -->
<!ELEMENT MetaData_Created_Date (#PCDATA)><!-- 범주의 MetaData_Created_Date 에 대한 정의 -->
<!--
<!ELEMENT Type (#PCDATA)> <!-- 범주의 Type 에 대한 정의 -->
<!ELEMENT Format (#PCDATA)> <!-- 범주의 Format 에 대한 정의 -->
<!ELEMENT Source (%e.source;)> <!-- 범주의 Source 에 대한 정의 -->
<!ELEMENT ISSN (#PCDATA)> <!-- 범주의 ISSN 에 대한 정의 -->
<!ELEMENT Frequency (#PCDATA)> <!-- 범주의 Frequency 에 대한 정의 -->
<!ELEMENT Vol (#PCDATA)> <!-- 범주의 Vol 에 대한 정의 -->
<!ELEMENT Pp (#PCDATA)> <!-- 범주의 Pp 에 대한 정의 -->
<!ELEMENT Identifier (#PCDATA)> <!-- 범주의 Identifier 에 대한 정의 -->
<!ELEMENT Language (#PCDATA)> <!-- 범주의 Language 에 대한 정의 -->
<!ELEMENT Holding (#PCDATA)> <!-- 범주의 Holding 에 대한 정의 -->

```

[그림 1] 학술지 논문 목록용 XML DTD



[그림 2] 목록입력 템플릿

터를 정의하고 사용하기 위한 구체적인 구문이 필요하다. 이를 위해서 RDF에서는 XML을 사용한다.

XML은 HTML과 SGML의 장점을 받아들이고 인터넷상에서 구조화된 전자문서를 기술하기 위한 마크업 언어로 1996년 W3C의 XML 워킹그룹에서 제안한 웹 문서의 표준이다. XML의 문서구조는 논리적 구조와 물리적 구조로 구성되며 논리적 구조는 문서의 전체적인 구조를 표현하는 부분으로 엘리먼트, 속성 등의 구성요소를 이용하여 표현하며, 물리적 구조는 엔티티를 이용하여 표현한다. 이들이 혼합되어 문서의 구조를 나타내는 DTD(Document Type Definition)를 형성한다(정희경 1999).

학술지 논문 목록을 위한 데이터요소들을 구조화된 XML 웹 문서로 표현하기 해서는 먼저

DTD를 작성하여야 한다.

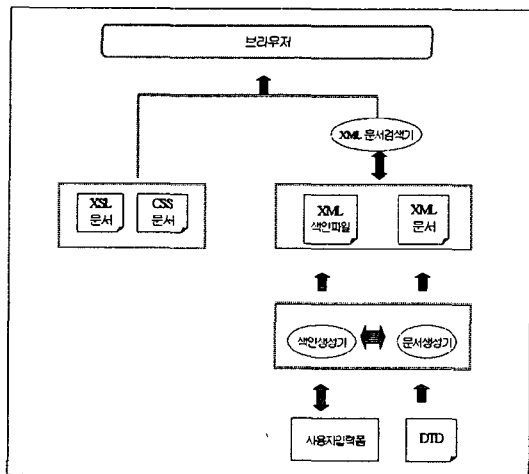
기본적으로 XML 문서구성은 ① XML 선언 ② DTD(문서형 정의) ③ XML 인스턴스로 나누어지며, XML DTD에서는 엘리먼트(element)의 출현순서는 연결자의 ',' (comma)를 사용해서 표시하고, 엘리먼트 반복회수의 경우 '?' 는 발생할 수도 있고 발생하지 않을 수도 있는 선택적 요소를 말하며, '+' 는 1회 이상 출현하는 요소, '\*' 는 출현하지 않거나 반복적으로 출현하는 엘리먼트를 의미한다.

학술지 논문목록을 위하여 사용되는 기본요소 중에서 필수요소로 title, creator, publisher, date, type, format, identifier, source, language, holding 요소를 선정하고, 선택요소는 subject, description 요소로 구분하였다. 또한 기본요소 밑에는 하위요소를 한정어로 구분하여 필수와 선택항목으로 구분해서 선택적으로 목록에 활용하였다.

목록을 위한 데이터 요소들을 쉽고 간편하게 입력할 수 있는 템플릿을 작성하여 반복적으로 데이터 요소들을 입력할 수 있도록 설계하였으며 입력된 데이터 요소들은 선언된 DTD를 기본으로 구문과 형식 또는 인스턴스가 일치하는지 점검을 받는데 그 역할은 파서(parser)가 맡는다. 입력 데이터 요소가 정상적인 경우에는 XML 웹 문서형태로 변환, 생성되어 구조화된 목록파일로 작성되게 된다.

#### 4. 검색모형 설계

XML 문서는 MARC 포맷으로의 변환을 통하여 대학도서관의 자체 서지 데이터베이스와 통합 운영이 가능하도록 하고 통합 검색기능도 갖추는 것이 바람직하다. 특히 대학별로 학술 논문 웹 데이터베이스를 구축함으로써 인터넷 검색엔진들의 정보수집 경로를 차단하지 않도록 하여야 한다. 이는 웹 환경에서 기존의 관계형 데이터베이스들이 많은 서지정보를 보유하고 있으면서도 외부이용자나 검색엔진들이 쉽게 접근할 수 없는 시스템으로 구성되었음을 알 수 있다.



[그림 3] 목록시스템 구성도  
인터넷을 이용하는 일반 이용자들이 대학 학

술정보에 쉽게 접근할 수 있는 방법으로는 이미 구축된 XML 웹 문서형태의 목록파일이 다양한 인터넷 검색엔진에게 체계적이고 구조화된 목록정보를 제공하도록 하이퍼텍스트 링크를 유지해야 한다.

이를 위한 방법으로 XML 문서를 생성시 지정된 주제어들을 키워드 색인파일로 작성하고 이용자들이 키워드를 열람하거나 키워드 검색 방법을 통해 목록정보에 접근하도록 하였으며, 키워드 검색방법은 목록정보로 입력된 모든 데이터 요소들을 대상으로 XML 문서에서 정보를 검색하도록 설계하였다.

인터넷 검색엔진을 이용한 학술지 논문의 목록정보에 접근하는 방법은 일반검색엔진들이 목록정보를 수록하고 있는 XML 웹 문서를 탐색하는 방법과 대학 학술지 논문만을 전문으로 탐색할 수 있는 전문검색엔진의 개발을 통한 방법이 있다.

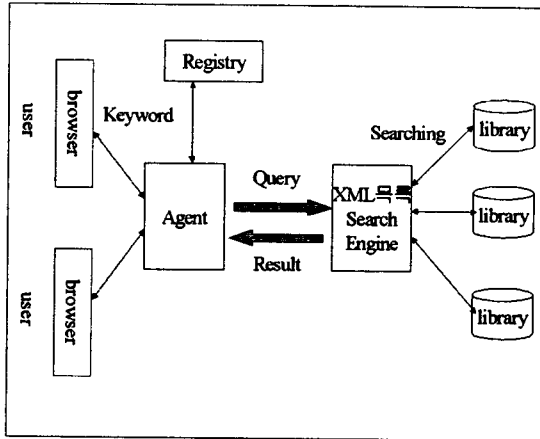
일반검색엔진을 이용한 방법은 아직 체계화되지 않은 정보수집 방법과 검색엔진마다 고유한 수집방법을 채택하고 있기 때문에 전문검색엔진 개발이 더 효과적일 것이다.

전문검색엔진은 더블링크어 메타데이터 요소를 기반으로 XML 웹 문서로 표현된 학술지 논문 목록정보를 체계적이고 효과적으로 수집 또는 탐색할 수 있는 방법이다.

결국, 대학별로 구축된 XML 웹 형태의 분담 목록시스템은 인터넷 검색엔진 및 전문검색엔진을 이용한 다중검색 방법 등을 통해 모든 대학의 학술논문의 목록정보를 검색할 수 있는 구조를 갖게 된다.

또한, 이용자 측면에서는 학술지 논문정보의 최신성 유지와 동일 학술분야의 중복연구 방지 등의 효과를 기대할 수 있으며, 도서관 측면에서는 해당 대학에서 발행된 학술지 논문만을 목록하여도 다른 대학의 학술논문정보를 쉽게 공유할 수 있는 효과를 기대할 수 있다.

[그림 4] 검색엔진을 이용한 검색구조



인터넷상에서 공개가 가능하고, 활용가치가 높은 각 대학의 학술지 논문의 원문정보를 다양한 검색시스템으로부터 독립적이고 개방된 구조의 목록시스템을 이용하여 인터넷 검색엔진과 학술지 전문검색엔진을 개발, 활용할 수 있는 검색모형을 구상하여 학술지 논문 이용의 활성화를 기대해 볼 수 있다.

### 5. 결론

각 대학에서 발행되는 학술지 논문들은 절대 중복될 수 없는 학술자원으로, 현재 각 대학도서관에서 디지털도서관의 일환으로 학술지에 대한 수집, 목록, 검색기능을 수행하고 있기 때문에 표준화된 메타데이터를 기반으로 한 분담 목록시스템을 구성하는 것은 어려운 일이 아니다.

분담목록을 위한 메타데이터로, 다른 메타데이터와의 호환성 및 상호운용성을 충분히 확보하고 있는 XML은 목록정보뿐만 아니라 목록정보에서 본문정보까지 확장하여 서비스할 수 있는 논리적 구조를 갖추고 있는 웹 문서의 표준이고, 인터넷 검색엔진들의 탐색 대상인 색인 요소들의 구조화가 가능한 문서로, 모든 학술지 논문은 검색엔진 등을 이용하여 논리적인 검색구문으로 정확한 검색결과를 얻을 수 있는

장점을 지니고 있다.

따라서 대학의 유용한 정보자원이라고 할 수 있는 학술논문을 대학도서관의 정책과 상관 없이 현재의 폐쇄된 형태의 서비스 방식에서 탈피하여 정보자원의 개방과 공유를 목적으로 메타데이터 기반의 학술지 논문 목록시스템을 활성화하는 방안이 절실히 요구된다.

### 참고문헌

- 백현기. “XML을 이용한 논문 관리 시스템 설계 및 구현”, 우석대학교 교육대학원 석사논문, 2002.
- 이명규, 김성준. “우리 나라 대학도서관에서 웹을 통한 원문정보서비스 현황 연구.” 『한국도서관정보학회지』, 32(1)(2001): 285-307.
- 정동열, 김성희. “대학학술정보 유통 촉진을 위한 제도 개선에 관한 연구.” 『한국문헌정보학회지』, 35(1)(2001):47-75.
- 정희경. “차세대 웹 문서 표준 XML.” 『정보처리』, 6(3)(1999):25-35.
- 최한석. “메타데이터 기반 학술지 논문종합 목록시스템 설계 및 구현.” 『정보관리학회지』, 18(2)(2001):57-76.
- 김태수. 2000. “더블린코어(Dublin Core).” <<http://dewey.yonsei.ac.kr/metadata/dc.htm>>
- DCMI : Dublin Core Metadata Initiative. <<http://dublincore.org/documents/dces>>