XML을 이용한 전자정보원의 메타데이터 구성에 관한 연구*

A Study on the metadata construction of electronic information by using XML

김현희, 명지대학교 문헌정보학과/정보통신교육연구센터
안태경, 대외경제정책연구원 정보자료실

Hyunhee Kim, Myongji University
Taekyoung Ahn, Korea Institute for International Economic Policy

본 연구에서는 경제, 경영학 분야의 전자 정보원을 좀 더 효율적으로 검색하기 위해서 각 정보원의 특성 및 내용을 체계적으로 기술하는 메타데이터를 설계한 후 XML과 ACCESS 데이터베이스 시스템을 이용하여 메타데이터 검색 시스템을 구현해 보았다.

1. 서론

최근 많은 정보원이 디지털화 되어 인터넷 또는 공공 통신망 등의 네트워크를 통하여 제공되고 있으나 네트워크 상에 산재하는 전자정보원이 어디에 어떻게 존재하고 있는지를 조사 하기 위해서는 많은 노력과 시간을 필요로 한다.

본 연구에서는 경제, 경영학 분야의 전자 정보원을 좀 더 효율적으로 탐색하기 위해서 각 정보원의 특성 및 내용을 체계적으로 기술하는 메타데이터를 설계한 후 검색 시스템을 구현해 보았다.

2. 연구 방법과 제한점

본 연구는 주제분야를 경제, 경영분야로 제한하여 메타데이터 요소를 설계하였다. 메타데이터 요소는 더블린 코아를 기반으로 하여 16개 요소로 설계하고 XML과 ACCESS를 이용하여 데이터베이스로 구현하였다. 시스템의 구현환경은 하드웨어는 Windows NT 서버를 사용하고 ACCESS로 저장한 메타데이터를 웹상에서 검색하기 위해 비주얼 베이직 6.0으로 작성한 ASP(Active Server Page) 프로그램을 제작하여 이용하였다.

* 이 연구는 일부 정보통신부의 정보통신 우수시범학교 지원사업에 의하여 수행된 것임.

3. 선행연구

먼저 메타데이터 기술을 위해 XML을 이용한 연구를 살펴 보면, 노트러스(Nottrott) 등은 생태학 데이터의 질적 확인 과정을 자동화하기 위해서 XML로 기술한 메타데이터를 이용하였다. 이 연구에서는 생태학 데이터의 메터데이터 구성을 위해 8개의 XML DTD를 설계하고 있다.

메타데이터 설계에 대한 연구에는 上村(Kamimura)과 齊藤(Saito) 등의 연구가 있다. 먼저 上村(Kamimura)은 웹정보원의 인터넷 디렉토리인 OReL(Online Resource Locator)를 구현하였다.

이 디렉토리는 학술 논문 등 웹상에 산재된 정보원을 찾아내어 그 정보원에 대한 메타데이터를 데이터베이스화하여 웹상에서 검색하도록 하는 서비스이다. OReL의 메타데이터는 더블린 코아의 요소중 관계(relation), 범위(Coverage), 권리처리(rights)의 세 가지 데이터 요소를 제외한 13개의 요소를 채택하고 이외에 등록번호, 공개블럭(정보원의 공개/비공개에 대한 필드), 소멸블럭(정보원이 소멸된 경우에 사용하는 필드) 등을 첨가하여 사용하고 있다.

또한 齊藤(Saito) 등은 더블린 코아를 이용하여 메타데이터 검색 시스템을 구성하였다. 이 시스템은 서로 다른 구조를 갖는 두 종류의 메타데이터를 포함하고 있다. 이 시스템의 편리한 점은 검색시 이용자가 지정한 더블린 코아의 데이터 요소를 두 메타데이터의 원래 데이터 요소로 자동 변환한 뒤 검색을 수행하기 때문에 두 메타데이터의 구조를 모르더라도 더블린 코아의 데이터 요소만 알고 있으면 검색이 가능하다는 점이다.

4. 메타데이터 요소 설계 및 기술

4.1 메타데이터 요소 설계

사이트(인터넷), 데이터베이스(예, LISA), PC 통신 (예, 천리안) 등의 전자 정보원의 메타데이터를 구축하기 위해 먼저 메타데이터에 포함시킬 요소들을 선정하였다. 선정 기준은 더블린 코아의 15개 요소들을 기반으로 하여 범위와 저작권을 제외한 13개 요소를 그대로 포함시켰고 국가, 가격 및 파일다운로딩 요소를 새로 첨가시켜 16개 요소로 구성하였다. <표 1>은 16개 메타데이터 요소에 관한 표이다.

<표 1> 메타데이터 요소 집합

요 소	필수	속 성	코 드
subject(주제)	0	type	통제어, 자연어
description (기술)	0	type	초록, 해제, 목차
title(표제)	0	type	본서명, 부서명, 총서명, 잡제
creator(저자)			
contributor (관련자)			
publisher (출판기관)			
date(날짜)			연도를 네자리로 표현 예) 1999
type(유형)	0	scheme	internet, online, file, PC
format(포맷)	0	type	html, text, pdf, xml, word, image, ftp
identifier (식별기호)	0	scheme	URL, Dialog, STN, JOIS, Kristal, 천리안
language(언어)	0	scheme	Kor, eng, jap
relation(관계)		scheme	수평, 상위, 하위
source(소스)			
nation(국가)	0	scheme	kor, usa, jap, fra
file-down (다운로딩)			
price(가격)	0	scheme	free, high medium, low

4.2 메타데이터 기술

1) XML을 이용한 메타데이터 기술

(1) DTD 설계

메타데이터를 기술하기 위해서 XML 1.0을 이용하였다. 메타데이터의 DTD를

<그림 1>과 같이 작성하였다. 메타데이터는 13개의 단일요소(single element)와 3개(creator, contributor, publisher)의 개념 모델(concept model)로 구성되어 있다. 또한 더블린 코아 처럼 각 요소가 scheme, type, language라는 속성값을 갖도록 설계하였다.

```
<!--Metadata DTD-->
<!-- Entity declarations-->
        .....
        <!--General attribute -->
<!ENTITY %  title.att
    "type          CDATA      #IMPLIED
    language       CDATA      #IMPLIED
    scheme         CDATA      #IMPLIED">
        .....

<!-- The top-level Meta element:-->
<!ELEMENT Meta (ctitle+, creator+, subject+,
   descrip*, contributor*, publisher*, date*,
   type, format*, identifier*, lang*,relation*,
   source*, nation?,file_down?, price)>
      <!-- Title-->
<!ELEMENT ctitle (%phrasal;)*>
<!ATTLIST ctitle
        %ptr.att;
        %title.att;
>
```

<그림 1> 메타데이터 DTD

(2) XML 문서 작성

앞에서 작성한 DTD 규칙에 따라 XML 문서를 작성하였다. <그림 2>는 XML 문서의 한 예로 XML 문서의 앞부분에는 문서의 버전(version)과 언어 코드, 해당 XML 문서의 DTD 파일명이 나타난다. 이 XML 문서는 화면상에 디스플레이해 주기 위해서 스타일 시트 언어로는 XSL을 사용하였다.

2) DBMS을 이용한 메타데이터 기술

메타데이터의 좀 더 효율적인 관리를 위해 ACCESS를 활용하여 데이터베이스로 구축하였다. 레코드의 필드는 기본키(문헌번호)를 포함시키고 주제 필드를 두 개의 필드(자연어, 통제어)로 구분하고 전자우편 필드를 따로 독립시켜 총 19개의 필드로 구분하였다.

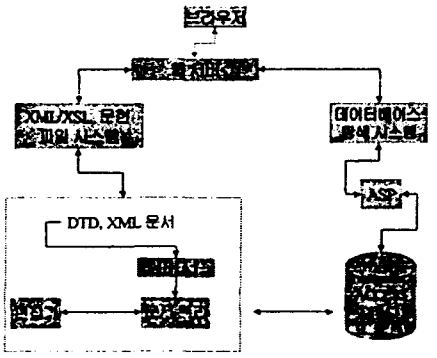
```
<?xml version="1.0" encoding="euc-kr"?>
<?xml:stylesheet type="text/xsl" href=
   "meta2.xsl"?>

<!DOCTYPE meta SYSTEM "meta.dtd">

<meta>
<ctitle>Dr. Ed Yardeni's Economics
   Network </ctitle>
<creator>
   <name>Dr. Ed Yardeni</name>
   <email href="mailto:yardeni@yardeni.com"
   >yardeni@yardeni.com</email>
</creator>
<subject scheme="자연어">경제 지표,
   economic indicators</subject>
<subject scheme="통제어">general business
   resources</subject>
<date>1999</date>
<type>internet</type>
<descrip type="초록">Dr. Ed is the Chief
   Economist at Deutsche Morgan Grenfell.
He has assembled this excellent collection
   of resources on the Web for economists
   and laypersons alike.
</descrip>
<identifier href="http://www.yardeni.com"
   >http://www.yardeni.com</identifier>
<lang>eng </lang>
<price>free</price>
</meta>
```

<그림 2> XML 문서의 예

5. 메타데이터 검색 시스템

5.1 시스템 개요

본 연구에서 구현한 검색 시스템은 메타데이터 검색과 XML 문서 브라우징을 할 수 있는 두 개의 서브 시스템이 결합된 시스템이다. 본 시스템을 이용하기 위한 웹브라우저에는 XML DTD 및 문서의 파싱 및 브라우징 기능을 갖고 있는 익스플로러(explorer) 5.0 및 게코(GECKO)가 사용되어야 한다. 시스템의 전체 구성도는 <그림 3>과 같다.



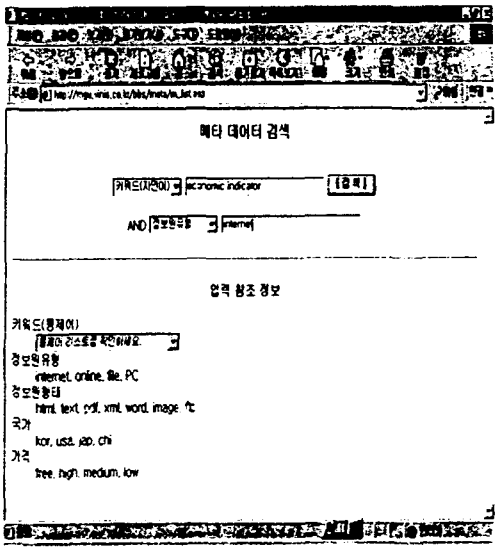
<그림 3> 시스템 구성도

5.2 시스템 구현

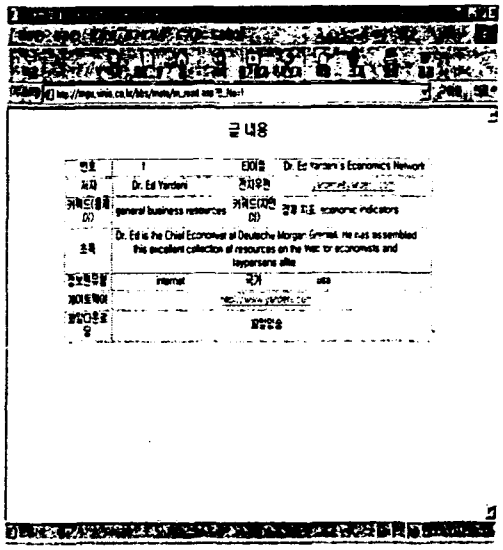
1)메타데이터 검색

메타데이터를 저자, 표제 및 주제 등으로 검색해 볼 수 있는 서브 시스템으로 데이터베이스는 ACCESS로 구성하고 검색 프로그램은 ASP로 작성하였다. <그림 4>는 시스템의 초기화면(http://mgu.vinis.co.kr/meta.htm)에서 첫 번째 메뉴(메타데이터 검색)을 선택하면 나오는 화면이다.

<그림 4>에서 주제가 "economic indicator"이면서 정보유형인 'internet"인 데이터를 검색한 결과 한 건의 레코드가 검색되었다(그림 5 참조).



<그림 4> 검색 화면

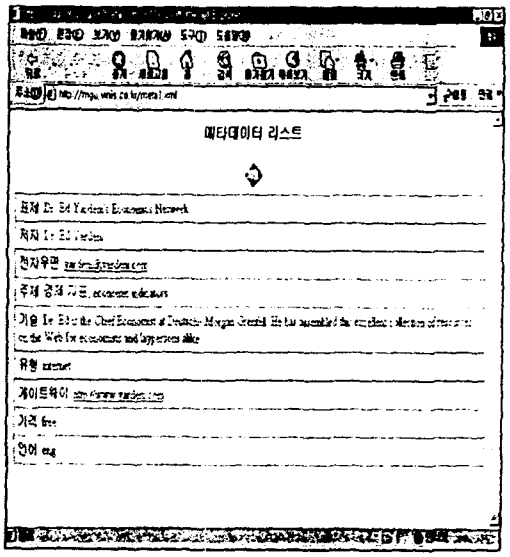


<그림 5> 검색 결과

2) 브라우징

시스템 초기화면에서 두번째 메뉴(브라우징)을 누르면 메타데이터를 주제별로 구분

해 놓은 메뉴가 나온다. <그림 6>은 "economic indicator"를 선택한 다음 제시되는 자료 중에서 한 문서를 선택한 다음 나오는 화면이다.



<그림 6> 검색 결과의 출력 화면

6. 결 론

본 연구는 주제분야를 경제, 경영분야로 제한하여 메타데이터 요소를 설계하였다. 메타데이터 요소는 더블린 코아를 기반으로 하여 16개 요소로 설계하고 XML과 ACCESS를 이용하여 데이터베이스로 구현하였다.

본시스템을 좀 더 효율적인 시스템으로 만들기 위해서는 다음과 같은 사항들이 요구된다.

첫째, 공식 또는 비공식적으로 유통되고 있는 전자정보원의 정보를 체계적으로 수집하기 위한 제도적인 장치를 마련해야 한다.

둘째, XML 문서를 효율적으로 활용할 수 있는 소프트웨어가 많이 개발되어야 한다.

셋째, XML로 작성된 문서를 기존의 검색엔진에서 검색할 수 있어야 한다.

<참고문헌>

門馬敎仁. 1999. XMLとメタデータ. 「情報の科學と技術」. 49(1): 16-22

杉本重雄 .1999. "Dublin Core Metadata Element Set について-現在の状況と利用例" ディジタル圖書館(Digital Libraries) No. 14 Mar. <http://www.dl.ulis.ac.jp/DLjournal/No_14/1-sugimoto/1-sugimoto.html>

齋藤ひとみ, 宇陀則彦, 石塚英弘. 1998. "Dublin Core Metadata Element Set による複數メタデータの檢索." ディジタル圖書館(Digital Libraries) No. 11 Mar. <http://www.dl.ulis.ac.jp/DLjournal/No_11/5-saito.html>

Carvalho Moura, Ana Maria. 1999. A Metadata Architechture to Represent Electronic Documents on the Web. <http://computer.org/conferen/proceed/meta/1999/papers/15/amoura.html>

National Archives of Australia. 1999. The Australian Government Locator Service(AGLS) Manual for Users. <http://www.naa.gov.au>

Nottrott, Rudolf W. 1999. Using XML-Structured Metadata to Automate Quality assurance Processing for Ecologicai Data. <http://computer.org/conferen/proceed/meta/1999/papers/64/ronottrott.html>