

# 웹 학술정보자원 DB구축을 위한 메타데이터 요소 분석에 관한 연구

A study on a Metadata Element Analysis for DB  
constructing of the scientific resources on the web

김규훈, 신성기 (성균관대학교 대학원 문헌정보학과)

Kim kyu hoon, Shin sung ki

Dept. of Library & Information Science, graduate school of Sung Kyun Kwan Univ.

웹 학술정보자원의 급속한 증가 및 이용도의 증가로 인하여 이들 자원의 원활한 활용을 위한 체계 마련이 시급하다. 본 연구는 인터넷에 산재해 있는 웹 학술정보자원에 대한 효율적인 DB구축을 위한 기초작업으로서 웹 학술정보자원에 한정된 메타데이터 세트 추출과정과 모형제시 및 각 메타데이터 요소에 대한 분석을 실시하였다. 메타데이터 요소 추출은 세계적으로 통용되는 메타데이터 세트의 분석과 매핑 작업 및 웹 학술정보자원의 구축을 위해 필요하다고 판단되는 요소를 추가하였으며 메타데이터 요소의 분석은 더블린 코어의 기본 15개 요소에 기저하여 실시하였다.

## 1. 서론

### 1.1 연구 목적 및 필요성

정보의 효과적인 유통을 위한 메타데이터의 연구와 개발은 이미 많은 노력과 시도가 이루어지고 있는 실정이다. 특히 인터넷의 보급이 확산되면서 웹 환경에서 제공되어지는 정보의 양과 수가 기하급수적으로 늘어나고 있으며 그 형태와 유형 또한 다양하여 효율적인 제공과 이용을 위한 체계가 시급한 실정이다. 더구나 최근 연구자들의 연구 행태 또한 인쇄매체 형태의 자료이용 중심에서 홈페이지나 자료실, 포털사이트등 웹상에서 제공하는 정보에 대한 이용이 증가하는 점을 감안한다면 웹 학술정보 자원의 체계적인 분류와 정리를 위한 메타데이

터 요소의 분석 및 추출은 가장 선행되어야 할 과정이라 할 수 있다.

이에 본 연구에서는 인터넷에서 제공되는 학술정보자원의 효율적인 데이터베이스 구축을 위한 기초 작업으로서 메타데이터 요소 추출을 위한 모형을 제시하고 추출된 메타데이터 요소를 분석하여 웹 학술정보자원에 한정된 효과적인 메타데이터 셋의 모형을 제시하고자 한다.

### 1.2 내용 및 범위

- 1) 본 연구는 웹 학술정보자원의 데이터베이스 구축을 위한 과정의 한 부분으로서 메타데이터 요소 추출 및 분석에 관한 내용으로 한정한다.
- 2) 본 연구는 미국, 영국 및 전세계적으로 사용되는 관련 메타데이터 스키마들을 조사하고 각각의 요소를 정리하여 각기 다른 메타데이터

셋의 요소들을 상호 매핑시켜 고유한 의미 요소들을 추출해 낸다.

3) 2)에서 추출해낸 메타데이터 요소를 더블린 코어의 15개 기본요소에 기저하여 분석한다.

4) 메타데이터 요소 추출 및 분석을 위한 대상 학문은 문헌정보학으로 한다.

## 2. 웹 학술정보자원 메타데이터 요소 추출 작업

웹 학술정보자원의 DB 구축을 위한 메타데이터 요소 추출작업은 다음과 같은 작업과정을 통해 이루어진다.

### 2.1 타 메타데이터들의 요소 조사 및 정리

일반적인 웹 자원의 기술을 위해 사용되는 요소들은 무엇인지를 조사한 후 공통적인 요소들이 무엇인지, 여러 메타데이터 셋에서 공통적으로 고유한 의미로 사용되는 요소들은 무엇인지를 조사하는 것은 매우 중요하다. 이를 위해서는 타 메타데이터 셋에 대한 요소들을 나열하고, 요소들을 조사 및 정리하는 것이 필수적이라고 할 수 있다. 웹 학술정보자원에 대한 DB 구축에서 이러한 과정은 향후 다른 웹 학술정보자원에 대한 DB들과의 연계성 및 상호운용성을 높여주는 역할을 한다.

본 연구에서 참고한 메타데이터의 스키마들은 다음의 표와 같다.

<표1> 참고한 메타데이터 리스트

스키마 이름	국가	분야
Dublin Core	세계적	일반 웹 자원
Dublin Core Education(2000-10-5)	세계적	교육자료
IEEE Learning Object Metadata(3.6)	세계적	교육자료
GEM Element Set and Profile(s)	미국	교육자료
IMS Learning Resource Metadata Core(1.1)	영국	교육자료
FAILTE metadata(2000-11)	영국	기술분야 자료
FIU DL Metadata	미국	일반 웹 자원

### 2.2 요소표 작성

2.1에서 조사·분석하여 추출된 요소들을 상호 매핑하여 각 메타데이터 셋에서 동일한 의미를 가진 요소들을 추출하고, 이를 그룹화하여 요소표를 작성한다. 공통요소 이외의 특수한 요소들은 그 개념을 최대한 살려 요소들을 그룹화 하였으며, 이렇게 추출된 요소들은 Dublin Core의 틀에 맞추어 상위요소와 하위요소로 구분하여 나열하였고, Dublin Core에 없는 특수한 요소들은 상위요소 또는 하위요소로 추가하는 방식으로 요소표를 작성하였다.

### 2.3 기본 메타데이터 셋 작성

2.1과 2.2 과정을 통해서 나온 요소표를 기초로 하여 체크된 요소들을 통합하고, 다시 해당 웹 학술정보 기술에 필요한 요소들을 추출하여 기본 메타데이터 셋을 작성하였다.

작성한 기본 메타데이터 셋은 웹 자원의 기술을 위해 가장 널리 쓰이고 있는 Dublin Core의 몇 가지 상위요소와 하위요소를 추가시킨 스키마를 가진 형태가 되었다. 즉, Dublin Core 스키마를 확장시킨 형태의 스키마를 가지게 되는 것이다.

본 연구에서 웹 학술정보자원에 대한 DB 구축을 위해 추출한 요소들은 기존의 웹 자원의 기술에서 가장 많이 사용되고 있는 Dublin Core의 기본 요소들을 골자로 하고 있으며, 특수한 요소들은 추가하는 방법을 사용하고 있는데 이는 기존의 타 웹 학술정보자원 DB와의 연계성 및 상호운용성을 고려한 결과라고 생각할 수 있다.

요소표를 통해 그룹화되고, 정리된 요소들을

<표2> 기본 메타데이터 셋

번호	요소명	설명
1	자료유형	자료의 성격, 장르
2	자료명	자원에 부여된 제목
2.1	대체 자료명	자원의 공식적인 이름을 대신하는 대체 자료명 (ex. 다른 언어로 번역된 제목, 간략제목 등)
3	작성자명	자료를 현 상태로 작성한 사람 또는 기관의 이름
3.1	작성자 역할	작성자의 역할
3.2	작성자 유형	작성자의 유형
3.3	작성자 상세정보	작성자에 대한 기타 상세정보
3.4	작성자 이메일	작성자의 이메일 주소
3.5	작성자의 홈페이지	작성자의 홈페이지 주소
4	주제명	자원의 내용이 지닌 주제(topic)에 관한 사항(협의된 주제리스트 중에서 값을 취함)
5	키워드	자원의 text로부터 선정한 용어나 어구로서, 자원에 의해서 다루어지고 있는 주요 개념들(비통제어 사용)
6	Description	자원의 내용과 관련된 설명 또는 요약
6.1	저자초록	자료의 원저자가 기술한 초록
7	목차	자원의 내용을 나타내는 하부 단위들의 리스트
8	발행자명	자원의 발행자 및 온라인 상의 자료 배포자 이름
8.1	발행자유형	발행자 유형
8.2	발행자 상세정보	발행자에 대한 기타 상세정보
8.3	발행자 이메일	발행자의 이메일 주소
8.4	발행자 홈페이지	발행자의 홈페이지
9	기여자명	작성자 요소에 명시된 개체 이 외에 자원의 내용에 대해 2차적인 책임이 있는 기타 개체의 이름
9.1	기여자 역할	기여자의 역할
9.2	기여자 유형	기여자의 유형
9.3	기여자 상세정보	기여자에 대한 기타 상세정보
9.4	기여자 이메일	기여자의 이메일 주소
9.5	기여자 홈페이지	기여자의 홈페이지 주소
10	날짜의 종류	자원이 존재하는 기간 동안 발생한 날짜의 유형 (ex. 갱신날짜, 최초자료 생성날짜 등)
10.1	날짜	날짜 종류에 대한 실제 값
11	자료형식	자원의 표현형식, 기술상(technical)의 데이터 유형

번호	요소명	실명
12	원정보제목	현재 자원의 출처가 되는 원정보자원 또는 현재의 디지털 자원을 생성하는 데에 사용되었던 출처 자원의 제목
12.1	원정보저자	원 자원의 내용에 주된 책임을 가진 개체
13	언어	자원의 지적인 내용을 기술하고 있는 언어의 이름
14	자원의 URL	자원의 온라인 상의 현 위치(URL)
14.1	기타 주소	이차적인 대체 주소들
15	관련자원 제목	
15.1	관계 종류	기술되고 있는 자원에 대한 관련자원의 관계유형
15.2	관련자원의 위치	관련자원의 위치
16	이용 조건 설명	자원 이용의 조건을 설명
17	평가 정보	리뷰나 평가, 또는 수상 등 자원의 질에 대한 정보
18	메타데이터 작성자명	메타데이터 작성자 및 작성기관의 이름
18.1	메타데이터 최종수정일	메타데이터의 최종수정일
18.2	작성자의 주석	메타데이터의 생성 및 유지 관리, 또는 자원의 관리를 맡고 있는 사람들을 위해서 목록자가 제공하는 내부적 관리를 위한 주석문(이용자에게는 보이지 않으며, 목록의 상태를 후임 목록자가 이해할 수 있도록 하거나 여러 명의 목록자가 있을 때 유용)

기초로 하여 작성한 기본 메타데이터 셋은 위의 <표2>와 같이 나타낼 수 있다.

### 3. 메타데이터 요소 분석

#### 3.1 자료유형

이 요소는 더블린 코어의 Resource Type과 대응되는 개념으로서 자료 유형은 강의자료, 논문, 사전 및 인명록, 2차정보원, 멀티미디어 자료, 각종 실무자료와 간행물, 그리고 웹 학술정보자원의 고유한 유형인 홈페이지와 자료실, 커뮤니티등을 포괄하는 개념이다. 웹 학술자원이라는 특성상 홈페이지와 커뮤니티, 자료실과 같은 다양한 유형을 세분화해야 하는 특성을 지니며 또한 내부적으로 다양한 유형의 학술자원을 가질 수 있으므로 단일유형이 아닌 다양한 유형이 반영될 수 있도록 하였다.

#### 3.2 자료명

자료명은 표제에 해당되는 개념으로서 메타데

이터 작성자가 육안으로 쉽게 인식할 수 있다. 이 요소는 더블린 코어의 Title과 대응되며 특히 홈페이지의 경우 공식명칭 이외에 약자, 영문명칭 등의 다양한 자료명을 사용하여 대표성을 지닌 자료명 식별에 어려움을 겪는 경우가 다분하다. 이 경우 메인 페이지에 쓰이는 자료명을 사용하고 나머지 자료명은 웹 학술정보자원에 대한 다양한 접근점을 제공하는 차원에서 대체 자료명으로 사용함이 바람직하다.

#### 3.2.1 대체 자료명

웹 학술정보자원이 지닌 다양한 자료명을 갖는 특성으로 인하여 대표성을 지닌 자료명 이외에 기타 자료명은 대체 자료명으로 기술한다. 더블린 코어에서는 본표제 이외의 표제에 대해서는 alternative를 사용하도록 하고 있는데 부표제, 대등 표제 등과 직접 대응되는 DC 필드가 없는 실정이다. 이에 김태수는 DC.Title의 하위 요소로서 Subtitle과 Parallel, Series와 같은 필드를 고려할 수 있음을 밝히고 있다.

### 3.3 작성자명

더블린 코어의 Creator에 해당되는 요소로서 전통적인 목록에서의 표목으로 중요하게 다루어지듯 웹 학술자원에서의 작성자 또한 자원에 대한 책임을 가진 요소로서 역할, 유형, 작성자와 같은 하위요소로써 상세히 구분하여 중요한 접근점으로서 다루었다. 문헌정보학에 한정된 웹 학술자원이 가진 특징을 살리기 위하여 역할 요소는 저자, 번역자, 자료가공자, 사진/삽화가의 하위 요소를 두었으며, 유형 요소는 사서/사서교사, 학생, 교수, 강사/연구원/기타, 국립기관, 교육/연구기관, 기업/민간기관의 하위 요소를 두었다. 작성자 요소는 개인/기관명, E-Mail, Homepage, 간략설명의 하위 요소를 두어 이용자의 열람이 가능하도록 하였다. 웹 학술자원의 특성상 개인명 이외에 다양한 단체에서 운영중인 홈페이지나 게시판, 자료실 등의 유형이 많으므로 단체의 성격을 규정할 수 있도록 상세한 기술이 용이하도록 하였다.

### 3.4 주제명

더블린 코어의 Subject에 해당되는 요소로서 자료유형과 더불어 얼마나 주제의 세부사항을 잘 제어하는가가 본 항목의 유용성에 대한 관건이라 할 수 있다.

### 3.5 키워드

주제명과 더불어 더블린 코어의 Subject와 대응하는 요소이다. 특히 키워드 요소는 검색을 위한 최적의 접근점을 제공하는 요소로서 기술 요소가 아니므로 시소러스의 유용성에 의해 효율성이 결정된다고 할 수 있다.

### 3.6 Description

웹정보자원의 내용을 문장의 형식을 빌어 간략하게 기술한 요소로서 자원에 대한 간략한 설명이라 할 수 있다. 초록과는 달리 자원의 기술자의 전문지식을 통해 요약된 고도의 지적

작업이 요구되는 부분이라 할 수 있다. 따라서 이 요소는 해당 주제에 대한 기술자의 전문지식이 요구되며 논문이외에 기타 자원은 초록의 요소가 없으므로 이용자는 자신의 정보요구를 충족시킬 수 있는 자원을 선택하는데 있어 중요한 판단 요소가 될 수 있다.

#### 3.6.1 저자초록

웹 학술자원에서는 초록을 담고 있는 논문 및 보고서가 차지하는 비율이 높으며 초록은 저자가 자신의 저작에 대하여 직접 작성한 요약된 내용이므로 Description과 마찬가지로 이용자의 자원선택에 있어 중요한 판단기준을 제공해 준다. 더블린 코어에서는 Description요소의 하위 요소로서 abstract요소를 사용한다.

### 3.7 목차

목차에 해당되는 더블린 코어의 요소는 Description의 하위 요소인 tableOfContents이다. 목차는 Description 및 초록과 더불어 이용자로 하여금 자원 선택의 판단요소가 되는 자원의 내용을 짐작케 하는 요소로서 자원에 기술된 목차 그대로 수정 없이 사용이 가능하다. 홈페이지, 게시판, 자료실 등 또한 메인 페이지에 나타나는 각 메뉴의 기술을 통해 자원의 내용을 짐작할 수 있도록 표현해야 함은 물론이다.

### 3.8 발행자명

발행기관이나 혹은 배포처로서 더블린 코어의 Publisher에 해당한다. 인쇄형태의 자원과는 달리 웹 학술자원은 개인에 의한 저작물뿐만 아니라 단체나 기관에 의해 운영되는 자료실 등의 개개의 자원이 통합된 이른바 묶음자원 또한 상당수라는 특징을 가지고 있다. 발행자 요소 또한 작성자 요소와 마찬가지로 역할, 유형, 발행자의 제 요소가 적용되어 발행자의 성격을 규정할 수 있도록 하나 역할 요소의 하위에는

발행자요소만이 가능하다는 점이 다르다고 할 수 있다.

### 3.9 기여자명

전통적인 목록에서의 부출표목에 해당하는 요소로서 더블린 코어의 Contributor와 대응되는 개념으로 작성자 및 발행자 요소와 마찬가지로 기여자의 역할, 유형, 기여자/기관의 제 하위 요소가 적용되나 역할 요소에서 번역자, 논문 수여기관, 사진/삽화가, 기타의 하위 요소를 두었다는 점이 다르다고 할 수 있다.

### 3.10 날짜의 종류

자원의 제작 및 이용에 관련된 요소로서 더블린 코어의 Date요소에 해당된다. 날짜요소의 기술형식은 ANSI X3.30-1985에서 규정한 YYYYMMDD형식을 사용하며 날짜의 종류에는 작성일, 발행일, 최종 수정일로 세분하여 작성한다. 더블린 코어에서는 작성일을 created, 발행일을 available, 최종 수정일을 modified로써 사용하고 있다. 작성일은 자원이 최초 작성된 날짜를 의미하고 발행일은 이용자로 하여금 인터넷을 통하여 이용 가능해진 시점을 의미하며 최종 수정일은 관리자로서 하여금 가장 최근에 수정 보완된 날짜를 의미한다. 날짜 요소가 웹 학술자원에 있어 특히 더 중요한 이유는 웹 학술자원의 특성상 인쇄매체의 자원에 비하여 수정, 보완 및 갱신되는 주기가 빠르다는 점을 생각해 볼 때 최신의 정보자원을 요구하는 이용자의 판단기준이 될 수 있으며 특히 갱신주기는 웹 학술자원의 질을 판단하는 중요한 요소가 될 수 있기 때문이다.

### 3.11 자료 형식

더블린 코어의 Format에 해당하는 자료형식은 웹에서 제공할 수 있는 모든 형태의 학술자원을 포괄한다. text기반의 txt, hwp, doc를 비롯하여 html, ASCII, PDF, ppt형태는 물론 멀티

미디어 자원을 표현할 수 있는 mpeg, jpg, wav, mp3등의 형태가 가능하다. 자료실과 같은 이른바 묶음자원은 다양한 형식을 지닌 하위자원을 내포할 수 있으므로 자료유형요소와 마찬가지로 단일 형식이 아닌 다양한 형식을 반영할 수 있도록 해야 한다.

### 3.12 원자원 정보

현재 자원의 출처가 되는 원자원 및 인터넷을 통한 자원의 제공을 위해 디지털 자원을 생성하기 위하여 사용되었던, 출처가 되는 정보로서 개인 홈페이지에서 제공하는 강의록과 같이 웹 학술자원에서 가공된 자원의 비율이 높은 현실을 반영한 요소이다. 더블린 코어의 Source요소와 대응되며 원정보의 제목, 원정보 저자, 작성일등과 같이 세부적인 기술이 요구된다.

### 3.13 언어

자원의 내용을 표현하고 있는 언어로서 더블린 코어의 Language요소에 해당한다. 여기서는 주로 문헌정보학 분야의 웹학술자원의 상당수를 차지하는 한국어, 영어, 일본어, 중국어, 한문, 기타의 하위 요소를 이용하였으며 C, Java와 같은 컴퓨터 언어는 반영하지 않았다.

### 3.14 자원의 URL

더블린 코어에서는 Identifier요소에 해당되는 URL요소는 웹 학술자원의 소재 위치를 알려주는 요소가 됨은 물론 자원을 식별해 주는 요소이다. 자원과의 직접적인 연결 고리가 되는 요소이므로 자원의 이동에 따른 dead-link상태를 주기적으로 파악해야 하며 필요시 대체 URL의 기술이 가능하도록 하였다.

### 3.15 관계정보

더블린 코어의 Relation요소에 해당하는 관계정보 요소란 해당자원과 또 다른 자원과의 상,

하위 혹은 관련 자원과의 관계 등을 명시해 주는 요소로서 웹 학술자원의 성격을 고려하여 하위자원, 강의계획서, 강의 관련자료, 참조자원의 하위요소를 두었다. 웹 학술자원에서는 자료실과 자료실이 담고 있는 개개의 웹 학술자원, 홈페이지와 홈페이지 내부의 강의록과 같이 자원과 자원의 내부에 존재하는 상, 하위자원과의 관계 혹은 참고자원과 같은 관련자원의 명시를 통해 이용자는 관련자원의 더 많은 검색을 용이하게 할 수 있다.

### 3.16 이용조건

이용조건은 더블린 코어의 Right Management에 대응되는 요소로서 자원의 관권은 물론 유료자원의 경우 이용료 및 자원을 열람하기 위하여 요구되는 프로그램의 설치 등 해당자원을 이용하기 위해 요구되는 모든 사항을 포함한다.

### 3.17 평가정보

자원에 대한 리뷰나 평가 혹은 특히 홈페이지의 경우 권장, 추천 사이트나 수상과 같은 웹 학술자원의 질을 판단하기 위한 정보로서 이용자의 자원선택을 위한 편의 중심의 요소이다.

### 3.18 메타데이터 작성자 정보

메타데이터 작성자 정보는 실제 개개의 자원에 대해 메타데이터를 작성한 작성자 및 작성기관에 대한 정보로서 이용자를 위한 자원의 상세 정보항목에서는 보이지 않으나 작성자의 주석과 같이 메타데이터 작성자간에 내부적 관리를 위해 제공되는 작성자 중심적인 요소이다.

## 4. 결론

더블린 코어의 15개 메타데이터 요소가 자원의 기술이 아닌 자원의 효율적인 탐색을 위해서 개발되었듯이, 본 연구에서 제안한 웹 학술정보자원의 DB구축을 위한 메타데이터 요소 역

시 자원의 기술보다는 이용자의 자원 탐색의 편의를 제공하기 위한 요소가 중심이 되었다. 기존의 Yahoo, Altavista, Google과 같은 웹 탐색엔진은 특히 자원의 질에 대한 판단 및 여과 없이 단순히 끊어오는 수준의 문제를 갖고 있었다. 이를 극복할 수 있는 대안으로써 인간의 지적 작업을 통하여 인터넷에 산재해 있는 웹 학술정보자원의 메타데이터 작업을 통한 자원과의 직접적인 연결을 통해 효율적인 자원의 검색수단이 됨과 동시에 '상세보기 구현'과 같은 자원에 대한 메타데이터요소의 열람이 가능한 목록의 수단이 되었다. 하지만 전통적 목록의 대체 수단으로서 자원을 기술하기 위한 수단으로서의 기존의 연구에서 나타나듯 아직 많은 한계성을 내포하고 있다. 특히 기존의 도서 관목록레코드에서 나타나는 완벽한 전거제어를 통한 목록의 기술이 아닌 부분적인 전거제어의 수준에 머무르고 있기 때문이다. 그 중에서도 자료유형, 주제명, 자료형식, 언어 요소는 전거제어의 틀을 이용하여 일관된 자원기술이 이루어지고 있지만 키워드 요소는 시소러스나 주제명표목표와 같은 일관된 전거제어의 틀이 요구되는 점에서 앞으로의 보완사항으로 지적된다. 또한 IFL(<http://www.ifl.or.kr>)의 시범 서비스는 메타데이터의 입력 도구로써 단일 폼을 사용하고 있는데 다양한 웹 학술자원의 메타데이터 요소를 구현하기 위해서는 각각의 자원 유형에 맞는 다양한 입력 폼을 사용해야 할 것이다.

## 참고문헌

- 김태수. 1999. "메타데이터 변환과 자원기술구조의 연구" 정보관리학회지 15(3)  
IFL(Internet Faculty Library). 2002  
<<http://www.ifl.or.kr>>  
Dublin Core Metadata Initiative Education working group. 2001

<<http://www.dublincore.org/groups/education>>

Dublin Core Metadata Initiative. 2001

<<http://dublincore.org>>.

Gateway to Educational Materials. 2001.

<<http://www.geminfo.org>>.

IEEE P1484.12 Learning Object Metadata  
Working Group. 2001

<<http://ltsc.ieee.org/wg12/index.html>>.

FIU DL Metadata Creation & Maintenance  
Manual.

<<http://www.fiu.edu/~diglib/metadata/index.html>>.