

碩 士 學 位 論 文

고서용 KORMARC의 XML변환모듈
구현에 관한 연구

2006年

漢城大學校大學院

文獻情報學科

情報學專攻

申 翊 澈

碩 士 學 位 論 文

指導教授 姜 順 愛

고서용 KORMARC의 XML 변환 모듈
구현에 관한 연구

2006년 6월 日

漢城大學校大學院

文獻情報學科

情報學專攻

申 翊 澈

碩 士 學 位 論 文

指導教授 姜 順 愛

고서용 KORMARC의 XML변환모듈

구현에 관한 연구

An Implementation of XML Transformation Module of Old
Books on KORMARC Format

2006年年 6月月 일 月 日

漢城大學校大學院

文獻情報學科

情報學專攻

申 翊 澈

申翊澈의圖書館學 碩士學位論文을 認定함

2006年 6月 日

審査委員長 (인)

審査委員 (인)

審査委員 (인)

목 차

1 서론	1
1.1 연구의 필요성 및 목적	1
1.2 연구범위 및 방법	2
1.3 선행연구	3
2 이론적 배경	7
2.1 고서용 KORMARC	7
2.1.1 고서용 KORMARC 목록형식과 기술규칙 제정	7
2.1.2 고서용 KORMARC 기술규칙 및 형식	8
2.1.3 고서용 KORMARC의 문제점	13
2.2 고문헌 기술을 위한 메타언어 및 메타데이터	14
2.2.1 XML	14
2.2.2 DC 및 RDF Syntax	16
3 XML 변환모듈의 설계 및 구현	18
3.1 고서용 KORMARC의 XML 변환요소 설계	18
3.1.2 고서용 KORMARC 구조분석	19
3.1.3 고서용 XML 변환요소 매핑	23
3.1.4 고서용 XML의 DTD 구조분석	25
3.2 고서용 KORMARC의 XML변환모듈 개발	29
3.2.1 변환모듈 개발목적	29
3.2.2 변환모듈 개발환경	30
3.3 XML 변환모듈 구현	31
3.3.1 구현 방안	31

3.3.2 구현 화면	32
3.4 XML 변환모듈에서 사용된 스타일시트 의의	37
4 고서용 KORMARC의 XML변환모듈과	
다른 변환 프로그램의 비교	42
4.1 비교 프로그램 선정기준	42
4.2 SOLARS	43
4.2.1 도입시기 및 활용	43
4.2.2 고서용 KORMARC의 입력 및 변환화면	44
4.2.3 SOLARS 변환 프로그램의 특징	46
4.3 XMLAS	48
4.3.1 도입시기 및 활용	48
4.3.2 고서용 KORMARC의 입력 및 변환화면	48
4.3.3 XMLAS 변환 프로그램의 특징	51
4.4 고서용 XML 변환모듈과 SOLARS 및 XMLAS의 비교평가 ..	52
5 결론	55
참고문헌	58
부록자료	62
Abstract	95

표 목 차

<표 1> 고서용 KORMARC의 구조 분석	19
<표 2> DC 요소 및 MODS 요소비교	23
<표 3> 고서용 변환요소 비율	29
<표 4> DC요소별 스키마 및 비교자료	39
<표 5> 고서용KORMARC의XML변환모듈과 SOLARS 및 XMLAS의비교	52

그림 목차

<그림 1> 고서용 변환요소명 구조구문	26
<그림 2> 주역언해 MARC 편집 화면	33
<그림 3> 표법연화경 MARC 편집화면	33
<그림 4> MARC 데이터를 XML로 변환 한 소스보기 화면	34
<그림 5> 형태기술필드 태그 편집화면	36
<그림 6> 고서데이터 입력화면	45
<그림 7> 변환화면	45
<그림 8> XML로 표출하는 화면	46
<그림 9> XMLAS 고서용 KORMARC 데이터 입력화면	49
<그림10> MODS 보기 변환화면	50

부록 목차

<부록 1> 고서용 KORMARC XML변환 매핑구조구문	62
<부록 2> 고서용 XML 변환모듈 스타일시트	81
<부록 3> 표법연화경 MODS변환자료	92

1. 서론

1.1 연구의 필요성 및 목적

세계 각국의 도서관들은 20세기초 도서관 목록의 주류를 이루었던 기존의 카드목록형식의 한계를 극복할 수 있는 기계가독형목록(Machine Readable Cataloging: MARC) 형식으로의 가능성을 타진해왔다. 이러한 가능성은 1950년대 미국의회도서관에서 시작되어 1965년 LC MARC I이 개발되었고 1968년 수정·보완한 MARC II가 제정되었다. 이 MARC II포맷(USMARC)은 ISO 2709, ANSI Z39.2 표준 규격을 수용하고 있으며, 미국의 국가적 표준인 USMARC Format for Bibliographic Data 뿐만 아니라 세계 각국의 MARC포맷 개발의 기초가 되었다. 또한 1973년에는 서지정보교환형식이 ISO로 제정됨에 따라 국제적으로 널리 MARC 형식의 개발이 진행되었다. 여러 나라들은 국제표준이 정해졌지만 나라마다 목록작성규칙이나 관습이 달라 자국의 서지기술법과 언어특성에 알맞은 독자적인 MARC를 개발하게 되었다.

우리나라에서는 국립중앙도서관이 1983년 한국문헌자동화목록법(KORMARC)운영위원회가 구성되어 LCMARC II 형식과 US MARC 및 ISO 2709를 기반으로 한국문헌자동화목록법(KORMARC) 단행본 형식의 개발에 착수한 이래 1993년 단행본용 한국문헌자동화목록형식 및 기술규칙이 KS X 6006-2로 제정되었고 이후 연속간행물(KS X 6006-1, 1994), 비도서자료용(KS X 6006-3, 1996), 전거통제용(KS X 6006-4, 1997), 소장정보용(KS X 6006-5, 1999), 고서용(KS X 6006-6, 2000), 통합서지용(KS X 6006-0, 2005)이 차례로 KS 제정되었다. 이러한 일련의 KS의 제정은 목록작업의 표준화를 가져오고 나아가 도서관간 상호협력에 의한 집중목록으로서의 종합목록작성을 촉진하는데 기여하였다.

그러나 컴퓨터기술의 발전으로 서지기술의 대상물로서 새로 출현한 다양한 웹정보자원의 기술과 인터넷상에서 이러한 정보자원을 공유하려는 움직임은 기존의 표준화된 목록규칙인 MARC의 한계를 드러내게 되었다. MARC는 자료처리를 위한 목록에 근거를 둔 메타데이터로서 기종이 다른 컴퓨터간의 호환성이 보장되지 않고 새로운 레코드 자원의 필드

정보를 확장하는데 어려움이 있으며, 레코드간의 연결기능이 부족한 것이 단점으로 지적되어 왔다¹⁾. 디지털화된 정보자원을 공유하기 위해 웹 자원과 통합한 새로운 정보시스템환경은 기존에 구축된 MARC서지관리 프로그램과 호환성을 유지하고 확장성을 가진 메타데이터 활용이 요구되었다. 메타데이터에 적합한 표준기술로 DC, TEI, EAD, CIMI Access Point, CDWA(Categories for the Description of Works of Arts)등이 있다.

이들 메타데이터 중 MARC와 DC와의 매핑이 많이 시도되었는데 이는 DC가 확장성이 우수하며 데이터 정보를 구조적으로 표현할 수 있고, 데이터 통합을 위한 정보기술의 응용에도 적합한 것으로 평가 받고 있는 장점을 지니고 있기 때문이다. 최근 미 의회도서관(LC)은 DC요소에 가장 표준화된 메타언어인 XML을 응용 개발한 MARCXML을 제시하고 있다.

KORMARC의 단점을 보완하기 위해서 있어서 단행본용 KORMARC 형식과 통합된 메타데이터로 변환하는 방법론은 많이 발표되었지만 고서용 KORMARC 데이터를 메타데이터로 변환하는 것은 몇 편의 논문이 발표되는데 그쳤을 뿐이다.

이에 본 연구는 고서용 KORMARC 데이터를 XML데이터로 변환하여 변환시키는데 가장 적합하고 이용가능한 방법을 제시하여 고서용 KORMARC의 단점을 보완하고 데이터 통합을 위한 정보기술의 응용성을 점검하는데 그 목적이 있다.

1.2 연구의 범위 및 방법

연구범위는 고서용 KORMARC의 구조를 분석하며, 데이터의 변환에 있어 고서용 KORMARC 필드 데이터를 변환하는 것이다.

고서용 KORMARC 포맷의 XML 변환모듈 구현을 수행하기 위해 다음과 같은 방법으로 연구를 진행하였다.

첫째, 고서용 KORMARC 목록형식 및 기술규칙, 고문헌 기술을 위한 메타언어로 XML 및 메타데이터인 DublinCore(DC)와 RDF Syntax에 대

¹⁾ 강순애, 1998 “한국고문헌목록프로그램과 원문디지털화의 구현에 관한 연구” 제 10회 한국학 국제학술회의논문집 -21세기 정보화 시대의 한국학. pp 566- 620.

한 이론적 배경을 고찰하였다.

둘째, 고서용 KORMARC의 XML변환모듈이 웹 상에서 변환이 가능하도록 고서용 KORMARC의 구조분석을 통해 DC type의 XML요소를 설계하였고, 국제 표준규격을 바탕으로 고서용 XML변환모듈을 개발하였다.

셋째, 고서용 KORMARC XML변환모듈과 국내에서 사용 중인 목록프로그램 중 고서용 형식을 지원하고 있으며 XML데이터로 변환기능이 있는 SOLARS 및 XMLAS를 비교하였다.

본 연구의 제한점은 다음과 같다.

첫째, XML 변환모듈 개발하는 단계에서 고서용 KORMARC 레코드를 DC변환요소로 일관되게 변환하기 위해 XLink의 XPointer 및 XPath의 적용을 시도하였으나 MS Explore 6.0에서 지원이 되지 않아 일부만 적용하였다.

둘째, 고서용 KORMARC 데이터와 고서용 XML데이터의 변환모듈은 MARCXML을 응용 개발하되, XML데이터 레코드를 고서용 KORMARC 레코드로 역 변환하는 작업은 제외하였다.

1.3 선행연구

본 연구와 관련된 선행연구에는 고서용 KORMARC형식에 관련된 목록연구, 고서의 원문이미지 전산화 및 고문서시스템 구축, 메타데이터 형식에 대한 연구 등이 있다.

첫째 고서용 한국문헌자동화목록에 대한 글은 1990년대 초반에 활발하게 이루어지기 시작했다. 이를 살펴보면, 윤구호 등²⁾은 고서가 지닌 희귀성, 귀중성, 유통의 폐쇄성, 문헌의 산재성 등이 이용자의 접근에 장애가 되고 있으며 이를 효과적으로 정리하기 위한 문헌자동화목록을 제정하는데 있어 서지포맷의 표준화가 선행되어야 함을 주장했다.

박문열(1999)³⁾은 고서목록 기술규칙(안)에 대해 고서용 형식의 기술규

2) 윤구호, 천혜봉, 김태수, 김이검. 1993. “한국문헌자동화목록법(KORMARC)에 관한 연구-고서용포맷 및 기술규칙” 정보관리학회지 제10권 1호 pp.3-12.

3) 박문열. 1999. “韓國文獻自動化目錄記述規則의 통합성에 관한 연구” 서지학연구 제 18집 pp 7-26.

칙, 저자사항, 판사항 등이 '한국 문헌자동화 목록형식'으로 나아가는데 한계가 있다는 점을 지적하고 기술규칙의 통합을 위해 수정되어야 할 사항을 다음과 같이 구체적으로 제시하였다.

고서용 한국문헌자동화목록 형식 및 기술규칙(안)에 대해 표준화의 제정과 목록기술의 방향을 기술한 강순애(1999)⁴⁾는 고문헌의 개념, 고서용 한국문헌자동화목록 형식의 연구 및 시안, 국내 고문헌 정보시스템 구축 및 사례, 메타데이터 제정과 고서용 한국문헌자동화목록의 방향설정을 언급하고 있다. 류부현(1999)⁵⁾은 고서용 한국문헌자동화목록 기술규칙(안)에 대해 오류를 지적하고, 보완 개정 수정되어야 할 사항을 자세하게 고찰하였다. 박재혁(1999)⁶⁾은 고서용 한국문헌자동화목록 형식에 대해 서지기술상의 특성 및 활용을 중심으로 고찰하고 표준화안 제정에 대한 몇 가지 제안을 첨부하였다.

둘째, 고문헌의 전산화 및 시스템구축과 관련된 연구로는 김은혜(1986), 이지영(1998)의 논문이 있다.

김은혜(1986)⁷⁾는 기존 고서목록기술양식이 다양하여 기술양식의 표준화가 선행되어야 할 것을 지적하고, 제반 MARC 구조와 상호 호환성을 위해 형식은 ISO 2709, 기술규칙 ISBD를 준용할 것을 제안하고 있다. 이지영(1998)⁸⁾은 기록된 문자와 우리나라 표준문자 코드의 특징을 살펴본 다음 고서의 디지털화의 필요성과 유용성을 제시 하였다. 그녀는 고서의 디지털화를 위해 사용될 수 있는 아스키데이터의 입력과정과 문자의 이미지데이터를 입력한 색인으로 검색하는 방안을 제시 했지만, 실제로는 적용하기 어려운 이론적인 한계를 드러냈다.

고문헌의 EAD DTD구현 및 메타데이터에 관한 연구를 보면, 김정미(2000)⁹⁾는 EAD를 채택하여 그 개념 및 구조를 분석했으며, 한국형

4) 강순애. 1999. "KORMARC : 한국문헌자동화목록형식 및 기술규칙(안)- 고서용-의 제정과 목록기술의 방향 설정에 관한 연구" 서지학연구 제 18집 pp 27-58

5) 류부현. 1999. "국립중앙도서관 제정 고서용 기술규칙(안)에 대한 연구" 서지학연구 제 18집 pp 59-76.

6) 박재혁. 1999. "고서자동화목록형식에 관한 연구" 서지학연구 제 18집 pp 77-101.

7) 김은혜. 1986. 「東洋古書目錄의 電算化를 위한 標準化 研究」 숙명여자대학교 대학원 석사학위논문

8) 이지영. 1998. 「고서의 디지털화 색인에 관한 연구」 이화여자대학교 대학원 석사학위논문

9) 김정미. 2000. 「국내 고문헌 처리를 위한 한국형 EAD개발에 관한 연구」 전주대학교 대학원 석

EAD설계하여 고문헌 정보시스템을 구축 제시하였다.

리상용(2003)¹⁰⁾은 XML을 활용한 원문 디지털화 방안으로 웹상에서 고문헌의 원문을 이용자들이 손쉽게 검색할 수 방안을 강구하였다. 고문헌을 디지털화 하기위해 XML DTD를 설계하였고, 그 과정에서 발생하는 문제점과 기술적인 한계를 서술 했으며, 향후 XML 문서를 활용한 국가적인 통합구축모형을 제시하였다.

외국의 사례는 고서의 페이지 이미지와 서지사항을 링크하여 네트워크를 통해 서비스하는 예로써 바티칸 디지털도서관 프로젝트를 언급 할 수 있지만, 원문 이미지의 보존에 중점을 두어 고서용 서지데이터를 구축한 데에 비교평가를 둘 수 있다. 더블린코어형식을 휘귀본에 적용한 DCRB(Dublin Core Rare Books)¹¹⁾는 RBMS(Rare Books and Manuscripts Section)에서 주창하는 오래된 서지와 악보 및 고문서에 대한 서술하는 방안을 제시 하고자 DC형식의 메타데이터를 개발 적용하고 있다.

고서의 정보조직체계를 서술한 김정하(2003)¹²⁾는 고서의 형태와 서지분석을 통해 기본 특징을 추출 서지정보뿐 아니라 원문정보까지 XML로 구현하여, 효율적인 고서관리에 대한 접근방안을 제시 했다.

일반이용자들이 인터넷검색에 있어 이용자인터페이스 및 이용자의 요구에 알맞은 해제 및 원문정보에 보다 주의의 기울여야 한다는 김달옥(2004)¹³⁾은 웹상에서 이용자에게 서비스되는 XML데이터의 레코드가 더 상세한 정보를 제공해야 되어야함을 지적하고 있다. 고도서의 메타데이터 안에 대한 글로는 서울대학교 중앙도서관¹⁴⁾은 2002년 디지털도서관을 구축하기 위하여 XML형태의 메타데이터 구현에 관한 연구를 수행 하였는데, 결론적으로 KORMARC은 전자문서 형식의 자원을 관리하기에는 미약하다고 보고 새로운 정보관리 모델로 IFLA(International Federation of

사학위논문.

10)리상용. 2003. 「XML을 활용한 고문헌의 원문디지털화 방안에 관한 연구 : 고문헌을 위한 DTD 개발을 중심으로」 한국문헌정보학회지, 37(1), pp 171-201.

11)<http://www.folger.edu/bsc/dcrb/dcrbrev.htm> (2006-06-27).

12)김정하. 2003. 「고서의 효율적인 정보체계」 숙명여자대학교대학원 석사학위논문

13)김달옥. 2004. 「고서용OPAC시스템의 기능분석 ; 국립중앙도서관, 규장각, 존경각, 장서각을 중심으로」 부산대학교 대학원 석사학위논문

14)서울대학교 중앙도서관. 2002. 『디지털도서관 구축을 위한 XML 스키마 메타데이터 구현에 관한 연구-최종보고서』 (서울 : 서울대학교 중앙도서관)

Libraries Associations and Institutions)에서 제안한 FRBR(Functional Requirements for Bibliographic Record)체계를 따라가야 한다고 제안하였다. 김성은¹⁵⁾은 KORMARC 데이터를 SGML DTD로 변환하는데 있어 DTD를 구조화한 개발방안을 제시하였다. XML변환에 참조는 조윤희¹⁶⁾는 통합형 XML구축을 위해 고서를 제외한 KORMARC에 대한 통합된 DTD를 설계하고 이를 XML 변환하는 방안을 제시하여 서지정보와 웹 자원 메타데이터변환을 시도하였다. XML을 이용한 서지정보시스템의 변환 설계 및 구현에 대한 김종진¹⁷⁾의 글은 다른 학문에서의 문헌정보학을 관점으로 서술한 서지정보에 대한 설계안과 XML 변환에 따른 서지정보의 효율성에 대해 알 수 있다.

15)김성은. 1997. 「KORMARC의 DTD 및 변환프로그램 개발연구」 연세대학교대학원 석사학위논문

16)조윤희. 2001. 「XML/KORMARC 통합시스템 구현에 관한 연구」 중앙대학교 박사학위논문

17)김종진. 2000. 「XML을 이용한 디지털도서관의 서지정보시스템 설계 및 구현」 중앙대학교대학원 석사학위논문

2. 이론적 배경

본 장에서는 고서용 서지기술을 위한 고서용 KORMARC 형식 및 기술규칙의 제정, 기술규칙 및 목록형식, 고서용 KORMARC의 문제점에 대해 살펴보았다. 또한 고문헌기술에 적합한 메타언어인 XML의 개요와 XML의 표현문서에 대해 살펴보고, MARC데이터를 XML로 변환하는데 있어서 도서관의 표준형식으로 채택된 메타데이터인 DC와 RDF Syntax를 중심으로 고찰하였다.

2.1 고서용 KORMARC

KORMARC(한국자동화목록법)의 설계원칙은 모든 유형의 자료에 대한 서지정보를 교환할 수 있도록 하며, 도서관 전산시스템 간의 레코드 교환에 필요한 명세를 제공하는데 그 목적이 있다. 그중에서도 고서용 KORMARC은 단행본형식이 완성된 후 문서형태 중 문헌자동화목록형식으로는 마지막단계에 제정되었고 고서의 특징 및 형태를 반영하여 가장 많은 표현형식을 지니고 있다.

2.1.1 고서용 KORMARC의 기술규칙 및 형식 제정

국립중앙도서관은 문헌자료 처리기법의 표준화를 위해 1993년 KSX6006-2 단행본용 한국문헌자동화목록 형식 및 기술규칙이 제정되었다. 이어 KSX6006-1의 연속간행물용, KSX6006-3의 비도서 자료용, KSX6006-4 전거통제용, KSX6006-5 소장정보용, KSX6006-6 고서용, KSX6006-0 통합서지용이 차례로 제정되었다.

국립중앙도서관이 제정한 KORMARC의 목록기술은 ISBD에 준거하는데 형식의 기본구조는 ISO 2709를 따르고 있다. 또한 LC MARC와의 준용과 호환성에 중점을 두었고, 세부적인 편성방침은 LC MARC을 준용하고, 식별자 규정에 있어서는 UK MARC 와 UNI MARC를 부분적으로 수용했다.

고서용 한국문헌자동화목록형식 및 기술규칙은 국립중앙도서관이 1992년

고서용 한국문헌자동화목록형식 및 기술규칙에 관한 연구를 '한국정보관리학회'에 위탁하면서 시작되었다. 이를 바탕으로 1993년 당시 도서관계 158개 기관의 의견이 수렴되었고, 도서관계 제시의견에 대한 한국정보관리학회의 검토요청이 있었다. 동년 3월에 1999년 3월에서 6월 사이에는 도서관계 225기관에 배포의견을 수렴하였고, 7월에서 9월 사이에는 4개 기관의 제시의견을 분석하여 반영 안을 작성하였으며, 1999년 10월28일 공청회를 개최하였다. 1999년 11월27일에서 12월 4일 사이에 공청회가 제시한 의견에 대해 도서관계 86개 기관의 의견을 수렴하여 1999년 12월17일에 제2차 공청회를 개최하였다. 동년 12월21일에서 29일 사이 KS(안)이 국립중앙도서관 정보화위원회 위원들의 검토의견 및 서면결의 요청을 반영하여 2000년 KS X6006-6 KS로 제정되었다.¹⁸⁾

2.1.2 고서용 KORMARC 기술규칙 및 형식

1) 기술규칙

고서용 한국문헌자동화목록 기술규칙은 고서의 기술사항을 규정한 것이다. 고서용 한국문헌자동화목록 형식은 국제표준서지기술법(ISBD-A)의 기술순서와 구두점을 준수한다.

고서의 기술적 특징을 간략히 살펴보면 서명저자사항, 판, 간사형태, 주기사항 등에서 차이점을 살펴 볼 수 있다. 고서의 서지 기술적 특징을 간략히 살펴보면,

첫째, 서명저자사항은 서명, 권차, 저자사항을 포함하는 것인데 현재 일반단행본과는 달리 표제지가 없는 고서는 대신 권수제 등을 참조하여 고서의 서명을 채기 한다. 다 권 다책본의 수량이 많은 고서의 경우는 첫째 권 또는 해당 도서 첫 번째의 권수제나 표제면을 기준으로 정하지만 다른 경우에는 이를 주기한다. 서명은 축약하지 않으며, 고서는 이명동서가 많아서 이명동서집람표 혹은 전거파일이 필요하다. 고서의 권차 사항은 책의 서지적인 단위이지만 저작자, 저작내용, 저작시대, 편목의 분합 존일(存佚)과 진위여부를 가리기 위하여 동양서는 서명다음에 반드시 기록한

18) 국립중앙도서관. 2000. 「한국문헌자동화목록형식-고서용」 (서울:당도서관)

다. 저작역할 표시가 다양하게 나타나고 저자의 역조를 원 괄호로 표기해 준다. 저작자의 호나 저자의 기타 사항은 주기사항에 기술한다.

둘째, 판 사항은 판차와 판의 명칭표시이다. 고서의 판 사항은 판차와 판종을 적어놓는다. 간행순서와 이판(異版)을 구별하는데 중요한 역할을 한다. 근간 순서는 1910년 이후 간행된 고서의 경우 초판, 재판 증보판으로 구분되지만, 1910년 이전 간인된 고서는 판차 대신 판종을 구체적으로 기술한다. 원각(간), 중각(중간), 또는 후각(후간), 증보, 번각 등으로 구분된다. 판의 명칭은 목판본, 금속활자본, 필사본, 영인본 또는 복제본 등으로 구분한다. 이외에도 사경은 인쇄용지 외 재료를 표기 해주어야 형태적인 특징을 파악할 수 있다.

셋째, 간사사항은 간사지, 간사자, 간사년으로 구분한다. 고서의 대부분은 간행사항에 기록이 기재 안 된 것이 대부분이어서 목기, 간기, 사기, 서문, 발문 등을 근거로 기재한다.

같은 이름의 지명이나 혹은 옛날 지명은 같은 지명은 해당하는 행정구역이나 국명을 원 괄호로, 옛 지명은 각 괄호로 표시 해준다. 간사자의 기록은 문헌에 있는 그대로의 채기 한다. 간행년도의 표기는 연호에 따른 기년, 즉위기년, 간기 기년, 간지의 교감에 의한 기년, 간지의 특수기년 등 여러 가지 특수한 상황을 고려하여 기재해야 한다.

넷째, 형태사항은 장정, 권 책수, 삽도류 및 판식, 책의 크기, 딸림자료 등을 기재대상으로 한다. 현재의 단행본은 페이지 수, 책 권수, 삽도, 책 크기만을 기술하는데 비해 고서의 외형적인 형태를 정확히 표시 하여 한 저작의 여러 판 종이 달라지는 모양을 살펴보고, 책의 외형적인 형태를 기록하여 훗날 제본시에 원형을 알아볼 수 있게 기록을 남기는 것으로 서술에 있어 상세한 기술이 요구된다. 19)

2) 목록형식

고서용 KORMARC의 레코드구조는 리더 및 디렉터리, 가변길이필드로 구성된다. 가변길이필드는 제어필드와 데이터필드로 구성된다.

가)리더

19) 강순애. 1989. 『고문헌 정보조직과 운용』 (서울: 아세아문화사, pp. 117-141)

리더(leader)는 각 레코드의 앞에 나타나는 24자의 고정길이 필드이다. 레코드처리에 필요한 기초 정보만을 수록하여 레코드 전체를 이용하기에 앞서 예비적인 정보를 식별하는 데 필요한 정보를 수록하고 있다. 리더부분은 24개의 문자로 이루어진 고정길이필드로 자수의 위치에 따라 레코드의 성격이 구분되어 있다. 레코드길이 00-04, 레코드 상태 05, 레코드 유형 06, 서지수준 07, 빈칸 08-09, 지시기호 자리는 10, 식별기호자리수 11, 데이터기본 번지 12-16, 입력수준 17, 목록기술형식 18, 연관레코드 조건 19, 엔트리 맵 20-23 등 숫자나 매개변수를 정의하는 부호 값으로 구성된다.

나)디렉터리

디렉터리는 책의 목차와 같은 것으로 필드의 위치 및 길이 정보를 나타내는데 표시기호 00-02(태그), 필드길이 03-06(Length), 필드시작위치 07-11(Starting character position)의 세 부분으로 구성된다. 디렉터리 항목은 12자리를 한 단위로 하며, 한 레코드에서 디렉터리 항목의 수는 입력된 표시기호 수와 동일하다. 디렉터리는 시스템이 자동 생성하는데, 자신의 표시기호, 지시기호, 식별기호를 갖지 않으며 끝에는 필드종단기호(ASCII 1E₁₆)를 갖는다.

다)가변길이필드

가변길이필드는 각각의 필드를 제어필드와 데이터필드로 구성한다. 제어필드는 0XX표시기호가 부여되며, 지시기호와 식별기호 없이 데이터와 필드종단기호로만 구성된다.²⁰⁾

제어필드는 제어번호 태그 001, 최종처리일시 태그 005, 형태기술필드 태그 007, 부호화정보필드 태그 008로 구성된다. 그중 태그 007은 형태기술필드로서 자료의 형태사항에 대한 상세한 정보를 부호로서 알려주기 위한 필드이다. 자료의 형태에 따라 02자리로부터 13자리까지의 각기 다른 고정길이 필드로 구성되는데, 그 길이는 007의 첫 번째 자리(자료범

²⁰⁾ 국립중앙도서관, 2000, 앞의 책, pp4-11.

정진식, 2005, 앞의 책, 56-113.

강순애, 2006, 『고문헌의 조직과 정보활용』 (서울:아세아문화사), p.183.

강순애, 2003, 「한국고문서의 정리현황과 목록의 표준화에 관한 연구」 서지학연구 26, pp.67-70.

주)에 의해 정해진다. 이중 고서와 관련된 것을 보면, '007/00'은 자료범주표시이며, 고서는 영문 'o'로 나타낸다. '007/01'은 특정자료표시로서 장정의 종류를 나타낸다. 권축장은 a, 절첩장은 b, 호접장은 c, 포배장은 d, 선장은 e, 족자는 f, 접포는 g, 모장은 h, 낱장은 I, 기타는 z로 나타낸다. 태그 008은 부호화정보필드인데 40자리의 고정된 필드로 레코드 전반의 서지적인 특성에 대한 정보를 데이터 관리 및 검색에 편리하도록 부호화하여 나타낸다. 이 필드에서는 지시기호나 식별기호를 사용하지 않으며, 자수위치별 데이터요소는 다음과 같다. 00-05 입력일자, 06 간사년 유형, 07-10 간사년 1, 11-14 간사년 2, 15-17 간사국명, 18 삽도표시, 19 광곽, 20 사경 및 사본의 계선, 21 판구, 22 어미, 23 자료형식, 24-25 내용형식, 26-27 빈칸, 28 수정레코드, 29-31 판중, 32 목록전거, 33 문학형식, 34 전기, 35-37 언어, 38-39 빈칸으로 구분하여 기술한다.

데이터필드는 실제의 데이터를 수록하는 부분이며, 그 구조는 제1지시기호 및 제2지시기호가 있고 표시기호가 표현하는 정보 이외의 정보를 추가하여 표현하고자 할 때 사용하는 부호로서 2자리로 사용된다, 식별기호는 가변길이필드내의 각 데이터요소를 식별하기 위해 사용하는 부호로서 2자리를 사용한다. 데이터요소는 필드에 포함된 서지내용이다.

01X-09X 필드는 숫자적 필드로서 국립중앙도서관 제어번호 태그 012, 국제표준도서번호(ISBN) 태그 020, 국제표준연속간행물번호 태그 022, 사건의 일시와 장소 태그 033, 협력기관제어번호 태그 035, 입수정보원 태그 037, 목록작성기관 태그 040, 언어부호 태그 041, 검증부호 태그 042, 지역부호 태그 043, 연대부호 태그 045, 기원전 간사년(BC년도) 태그 046, 소장사항 태그 049, 국립중앙도서관 청구기호 태그 052, 한국십진분류기호 태그 056, 사용문자세트 태그 066, 국제십진분류기호 태그 080, 듀이십진분류기호 태그 082, 기타분류기호 태그 085, 자관청구기호 태그 090으로 정의된다.

1XX 필드는 기본표목으로 채택된 태그 100 개인명, 태그 110 단체명, 태그 130 통일서명을 기술한다.

20X-24X 필드는 해당자료의 서명 및 그와 관련된 사항을 기술하는데, 통일서명 태그 240과 서명저자사항 태그 245를 포함한다. 이중 태그 240은 한 자료의 서명이 여러 가지로 다르게 간행되는 경우에 이를 한자리

에 모으기 위하여 특정서명을 통일서명으로 삼아 기술한다. 통일표목을 기본표목으로 사용한 경우는 태그 130을 적용하고, 기본표목이 아닌 기술사항으로 채택된 경우에는 태그 240에 기술한다. 태그 245는 서명저자사항을 기술하며 서명저자사항은 서명, 저자명, 권차표시로 구성된다.

250-29X 필드는 해당자료의 판차와 판중, 간사등과 관련된 사항을 기술하는데 판사항 태그 250과 간사사항 태그 260을 포함한다. 이중 태그 250은 판차나 판중 등의 판사항을 기술하고, 태그 260은 간사 또는 인쇄와 관련된 내용을 기술한다.

300 필드는 해당자료의 형태사항을 기술한다. 태그 300은 장정형태, 면장수 또는 권책수표시와 삽도 및 판식표시, 크기, 딸린자료 등을 기술한다.

4XX 필드는 해당자료에 나타나 있는 총서사항을 기술하며, 총서명 /부출표목 태그 440과 총서사항-부출되지 않거나 변형되어 부출되는 총서명 태그 490을 포함한다.

5XX 필드는 주기사항을 기술하며, 일반주기 태그 500, 합철본주기 태그 501, 서지 등 주기 태그 504, 내용주기 태그 505, 접근제한주기 태그 506, 원저자, 원서명에 대한 주기 태그 507, 인용주기 태그 510, 사건의 일시와 장소 주기 태그 518, 해제 등 주기 태그 520, 인용한 자료에 관한 주기 태그 524, 이용가능한 다른 형태자료 주기 태그 530, 복제주기 태그 533, 원본/복제본의 소장처 주기 태그 535, 이용과 복제 제한에 관한 주기 태그 540, 입수처 주기 태그 541, 동일자료의 위치주기 태그 544, 전기적 또는 역사적 주기 태그 545, 언어주기 태그 546, 자료내력에 관한 주기 태그 561, 연관저록 설명주기 태그 580, 참조 정보원 주기 태그 581, 소장본 주기 태그 590을 포함한다.

6XX 필드는 통제된 주제명부출표목이나 비통제색인어를 기술하며, 주제명 부출표목-개인명 태그 600, 주제명부출표목-단체명 태그 610, 주제명부출표목-회의명 태그 611, 주제명부출표목-통일서명 태그 630, 주제명부출표목-일반 주제명 태그 650, 주제명부출표목-지명 태그 651, 비통제색인어 태그 653을 포함한다.

7XX 필드는 서지레코드에 대한 추가의 접근점을 제시해주는 부출표목을 기술하며, 부출표목-개인명 태그 700, 부출표목-단체명 태그 710, 부출

표목-통일서명 태그 730, 부출표목-서명 태그 740을 포함한다.

76X-78X 필드는 관련 서지자료간의 관계를 나타내며, 원저저록 태그 765, 번역저록 태그 767, 부록저록 태그 770, 모체레코드저록 태그 772, 기본자료저록 태그 773, 이관저록 태그 775, 기타형태저록 태그 776, 선행저록 태그 780, 후속저록 태그 785, 비특정적 관계저록 태그 787을 포함한다.

830 필드는 490필드에서 다르게 부출되는 서명인 총서부출표목-통일총서명을 기술한다.

850-890 필드는 소장기관명부호를 기술하는 태그 850과 변형문자표시를 기술하는 태그 890필드를 포함한다.

9XX필드는 0XX-8XX의 레코드내에서 표목으로 채택한 것과 다르게 변형하여 부출하는 표목을 포함하며 로컬표목-개인명 태그 900, 로컬표목-단체명 태그 910, 로컬표목-통일서명 태그 930, 로컬표목-서명 태그 940, 로컬표목-총서명 태그 949, 로컬정보-가격 태그 950, 로컬정보-소장표시 태그 980을 포함한다.

2.1.3 고서용 KORMARC의 문제점

고서용 KORMARC는 단행본용 KORMARC 구조를 참조하여 고서의 기술적인 요소를 기술하도록 설계되어 있다. 단행본용 KORMARC 포맷은 LCMARC II 형식과 US MARC 및 ISO 2709를 기반으로 하여 한국 단행본의 서지사항을 기술할 목적으로 만들어졌는데 여러 학자들에 의해 레코드 구조의 복잡성, 확장성의 결여, 연결정보의 미비 등의 문제가 지적되어 왔다.²¹⁾

고서용 KORMARC의 문제점은 다음과 같다.

첫째, 고서용 KORMARC은 LC MARC나 USMARC의 디렉토리 구조와 기계가독언어 기반으로 작성된 단행본 KORMARC구조를 바탕으로 하였고 고서자체의 특성을 정확하게 기술하기 위해 매우 복잡하게 구조화되어 있다.

둘째, 고서용 KORMARC는 고서의 특징을 기술하기 위해 구조상의 포

21) 조윤희. 2002. 앞의 책 pp15-20.

활성이나 상세성은 지니고 있지만, 네트워크 자원의 통합적 기술이나 새로운 데이터 요소의 추가 및 기존 포맷을 수정하는데 있어서는 확장성이 결여되어 있다.

셋째, 고서용 KORMARC에서는 정보자원의 소재에 접근에 대한 정보를 생성하는 태그 856이 생성되어 있지 않다. 이 태그 856은 웹상에 인터페이스 정보를 가진 연결구조인데 이용자는 고서용 KORMARC 레코드에 입력된 대상 정보자료를 검색 열람 할 수 있지만, 웹 브라우저를 통한 인터페이스로 자료의 형태에 원문을 보기 위해서는 다음 단계 혹은 별도 필드에 마련된 원문보기형태로 이동해야 한다.

넷째, 태그 300에 기술한 고서의 자세한 형태사항을 008 부호화정보필드에서 요소별로 추출할 수는 있지만, 태그 008을 자동생성 할 수 있는 상세 규정등이 없어 태그 008 및 태그 300에서 오류가 발생하더라도 오류를 검증할 수 있는 방안이 없다.

2 고문헌 기술을 위한 메타언어 및 메타데이터

이 장에서는 웹환경에서 구조화 된 고문헌을 전송 가능하도록 설계된 XML과 이러한 문서를 기술하는 표준화된 메타데이터 DC를 중점적으로 살펴보았다.

2.2.1 XML

1) XML의 개요

XML은 웹상에서 구조화된 문서를 전송 가능 한 형식으로 설계 표준화된 마크업 언어이다 . 이는 지금까지 인터넷 표준으로 사용 하고 있는 HTML의 한계를 극복하고, SGML의 복잡성을 단순화 한 것으로써 상호운용을 기반으로 한다.

XML은 1996년 W3C (World Wide Web Consortium)에서 제안됐다. 당시 XML 워킹그룹은 Sun Microsystem 사의 Joe Bosak이 주도 하였다. 이 XML의 설계목표는 SGML과 호환되고 인터넷상에서 이용이 가능한 문서를 만들며, 다양한 어플리케이션을 추구하는 것이다. 즉 XML은 단순

히 HTML과 같이 프리젠테이션만을 위한 마크업 언어가 아니고 데이터의 내용과 구조를 담을 수 있는 의도으로써 웹 브라우저만 있으면 어떤 플랫폼과 어떤 기종에서건 똑 같은 출력물을 상호 호환 할 수 있는 데이터의 상호 호환성을 가진 문서언어이다.

단, 정보를 표현하고, 연결하고, 표시하는 것은 가능하지만 그 자체가 문헌과 자원을 조직하고 기술 하는데에는 한계가 있어 이를 해결하기 위해 여러방법을 이용하는데 본 연구에서는 RDF와 DC연계하는 방법을 차용하여 구조표현에 있어 RDF Syntax를 응용한 스키마구조 변환 요소형태로 설정한다. XML1.0 표준을 이용하여 XML를 표현한다.

2) XML의 표현문서

XML 표현문서로 변환하는 방법은 크게 3가지²²⁾가 있다. 논리적으로는 유사하지만 서로 다른 XML문서를 상호 변환하기 위해서 쓰이는 방안은 첫째, XML스키마 기반의 XML문서 목적에 맞게 XSL로 생성하여 처리하는 방법이다. 확장성에 가장 적합하지만, 이용자가 변환정보 내용을 전부 입력해야 하던지 XML구조 정보의 변환 없이 XML로 문서를 생성해야하는 단점이 있다. 둘째, DTD기반의 XML문서를 목적 스키마에 맞게 문서구조를 변환 할 수 있도록 변환기를 이용 XML문서를 상호 변환하는 방식이다. 이는 변환기를 이용 변환이 쉽고, 시스템구조가 구조화될 수 있음이 장점인데 반해, 사용자가 변환정보 내용 전부를 일일이 수정해야 하는 단점이 존재 한다. 셋째, 메타데이터 레지스트리를 이용한 XML문서를 상호 변환하는 방식은 DTD기반의 XML문서를 레지스트리 구조의 상이성을 이용하여 목적스키마에 알맞도록 변환하는 구조이다. 이는 구조적으로 다른 부분을 분석하여 자동화 할 수 있다는 장점이 있지만 구조적인 차이에 따른 분류 문제나, 데이터 레지스트리를 사용하는데 XML표준에서의 제약조건이 따르는 단점이 있다.

2.2.2 Dublin Core(DC)와 RDF Syntax

²²⁾ 조윤기, 2002. 「다양한 XML문서에 대한 통합관리 시스템」 충북대학교 대학원 박사학위논문 p14.

1) Dublin Core

더블린코어란 데이터의 호환성을 유지하고 네트워크 자원의 기술에 필요한 일련의 데이터 요소를 규정하여 이들 자원의 신속한 검색을 목적으로 1995년 OCLC와 NCSA(National Center for Supercomputer Application)두 기관이 더블린에서 개최된 워크숍 당시에 합의된 메타데이터이다. 당시 회의의 주목적은 MARC를 생성하는데 있어 구조적인 경직성으로 인해 많은 비용과 시간이 소모되어, 이것을 대체하는 단순한 구조의 형식을 모색하기 위한 것이다. 자원의 본질적인 요소를 기술하고, 본질적인 요소 외에 부차적인 내용을 확장기술이 가능하게 하며, 구문의 독립성, 각 요소가 강제적으로 수록되어야 한다는 규정을 없애고, 모든 기술요소는 반복사용 할 수 있으며, 한정어를 사용 세부사항을 조정하여 표현이 가능하게 함으로써 서지관리의 어떤 데이터요소가 상호운용가능하다는 장점을 지니고 있게 하였다. 단점은 자원의 소재위치와 이 자원을 획득하는데 필요한 데이터만으로 요소를 구성하되, 안전, 검증, 이용과 관련된 요소는 제외되었다는 것을 들 수 있다. 기본 데이터요소는 현재 13가지²³⁾로 한정어를 통한 특정표현을 나타낼 수 있다. DC포맷은 이와 같은 여러 가지 장점으로 인하여 국내에서는 단행본용 KORMARC 형식을 변환하는 메타데이터의 스키마에 가장 많이 적용되었다.

2) RDF Syntax

정보자원의 목적이 이전에는 도서관이라는 한정된 공간에 카드목록과 서지관리시스템으로 정보자원을 이용하게 했지만, 현재로는 인터넷을 통한 정보자원의 이용이 가장 중요한 문제로 표출되었다. RDF(Resource Description Framework)는 메타데이터 처리를 위한 기본 원칙으로 기계가독형에서 해독 가능한 정보를 교환하는 상호연계성을 제공해 준다. W3C의 RDF프로젝트는 서로 다른 플랫폼에서 정보의 이해를 기반으로 정보자원의 이용과 상호응용이 그 목적을 가지고 수행된 프로젝트의 한 예이다. RDF구조는 XML구문을 사용하며, 자원기술에 대한 메카니즘이 도서관학에서 출발하여, 자원을 기술하도록 엔코딩된 DC의 요소에 이르

²³⁾title, subject, description, language, relation, coverage, creator, publisher, contributor, right, date, type, form, identifier 등이 있다.

기까지 호환이 가능하기 때문에 이를 적용하여 고서용 메타데이터 변환에 필요한 구조적 자원을 설계할 수 있다.

3. XML변환모듈의 설계 및 구현

고서용 KORMARC의 고서용 XML 변환모듈을 설계하여 구현하는 작업은 고서용 KORMARC데이터를 XML데이터로 변환하여 KORMARC의 단점들을 보완하고 데이터 통합을 위한 정보기술의 응용성을 점검하는데 그 목적이 있다. 본 장에서는 고서용 KORMARC의 XML 변환요소를 설계하고 변환모듈을 개발하였으며 XML변환모듈을 구현하였다.

3.1. 고서용 KORMARC의 XML 변환요소 설계

고서용 KORMARC 데이터를 고서용 XML 데이터로 변환하기 위해 XML 변환요소를 먼저 선 추출하는 작업을 수행 하였다. 변환요소는 고서용 KORMARC 데이터의 완전한 변환을 고려하여 각 필드의 자세한 구조까지 표현하고자 변환요소 부분까지 매핑했으나, 모듈의 설계단계에서 상위 및 하위필드를 연관 지워주는 기능을 구현하는데 문제점이 발견되어 DC를 참조한 RDF구문을 응용하여 변환모듈을 개발하였다.

3.1.1 XML 변환요소의 설계

고서용 KORMARC의 형식을 근간으로 한 XML변환요소 설계원칙은 다음과 같다.

첫째, 고서용 KORMARC의 구조를 분석하여 XML DTD설계의 기본 데이터로 활용한다.

둘째, 고서용 KORMARC데이터를 XML데이터로 변환 시 상호호환을 고려하여 고서용 KORMARC에서 나타나는 데이터필드의 순서를 따랐다.

셋째, 변환모듈설계 기본 모델은 서울대학교중앙도서관에서 제시한 DC 요소를 이용한 고도서용 XML 스키마를 참조하였다.

넷째, 형태기술필드 및 부호화정보필드도 가능한 부분까지 XML데이터로 변환되도록 하였다.

다섯째, 일반 식별기호는 고서용 KORMARC에서 사용하는 표시기호의 숫자 및 자수위치를 나타내는 문자, 하위필드에 대한 식별기호에 사용된 문자 등을 그대로 표현하였다.

3.1.2 고서용 KORMARC의 구조분석

고서용 KORMARC은 한국 고문헌의 특성을 고려하여 설계된 표준으로 고문헌을 기술하는데 있어서 다양한 내용과 형태를 수용할 수 있도록 설계되었다. 이러한 고서용 KORMARC 구조는 다음과 같다.

〈표1〉 고서용 KORMARC 구조분석

서지기술구분		고서용 KORMARC 형식		비고
서명 저자 사항	서명	태그245	\$a 본서명 \$b잡제 \$x대등서명	
	권차		\$n 권차	
	저자		\$d 첫번째 저작자, \$e두번째이하 저작자	
판사항	판차	태그250	\$a 판표시	
	판중			
간사 사항	간사지	태그260	\$a 간사지	
	간사자		\$b 간사자	
	간사년		\$c 간사년	
형태 사항	장정	태그300	\$a 장정형태, 면장수, 권책수 함수, 갑수, 접지수, 축수 등	
	권책수			
	삽도류			
	형식	태그300	\$b 삽도 및 판식	
	책의크기		\$c 크기	
	딸린자료		\$e 딸린자료	
	재료			지질

서지기술구분		고서용 KORMARC 형식		비고
주기 사항	일반주기	태그500	\$a 일반주기	
	합철본주기	태그501	\$a 형식화된 합철본주기	
	서지 등 주기	태그504	\$a 서지주기	
	내용주기	태그505	\$a 형식화된 내용주기	
	접근제한주기	태그506	\$a 접근제한사항	
	원저자 및 원서명	태그507	\$a 원저자명 \$t 원서명	
	인용주기	태그510	\$a 해제지, 초록지, 색인지명	
	사건의 일시와 장소 주기	태그518	\$a 사건의 일시와 장소	
	이용대상주기	태그521	\$a 이용대상자	
	인용한자료에 관한 주기	태그524	\$a 인용주기	
	이용가능한 다른 형태자료 주기	태그530	\$a 이용가능한 다른 형태 자료	
	복제주기	태그533	\$a 복제형식	
	원본주기	태그534	\$p 원본관련 표출어	
	원본/복제본의 소장 처주기	태그535	\$a 소장기관명	
	이용과 복제에 관한 제한에 관한 주기	태그540	\$a 이용과 복제에 관한 제한 주기	
	입수처 주기	태그541	\$a 입수처	
	동일자료의 위치주기	태그544	\$a 보관자	
	전기적 또는 역사적 주기	태그545	\$a 전기 또는 역사 관련 주기	
	자료내력에 관한 주기	태그561	\$a 내력주기	
	연관저록 설명주기	태그580	\$a 연관저록 설명 주기	
	참조정보원 주기	태그581	\$a 참조정보원 주기	
	소장본 주기	태그590	\$a 낙장, 파손, 배접, 보사, 포갑 \$b 인문 \$c 장서기, 수증기, 수령 기, 수권기 \$d 지어 \$w 소장원본, 복제본의 청구기호, \$x 소장원본, 복제본의 형태사항 \$y 열람용 소 장본 \$z 소장관련 관리부호	
총서	총서사항/부출표목	태그440	\$a 총서명 또는 하위총서명	
	총서사항/ 부출되지 않거나 다르게부출되 는 총서명	태그490	\$a 총서명 또는 하위총서명	
소장처	소장사항	태그049	\$a 소장기관 부호	
	소장기관	태그850	\$a 소장기관명 부호	
이미지		태그856		원문 정보 연계
	목차정보			T O C 입 력

서지기술구분		고서용 KORMARC 형식		비고
요약 초록 해제		태그520	\$a요약주기	
			\$b초록주기	
			\$c해제주기	
언어	언어정보	태그041	\$a 본문언어 \$b요약문 \$h원저자언어	
	언어주기	태그546	\$a 언어주기	
주제명 부출 표목	개인명	태그600	\$a 개인명, \$c 이름과 관련된 칭호 및 기타명칭	
	단체명	태그610	\$a 기본요소 \$b 하위기관	
	통일서명	태그630	\$a 통일서명	
	일반주제명	태그650	\$a 일반주제명	
	지명	태그651	\$a 지명 \$x 일반제목	
	비통제색인	태그653	\$a 색인어	
부출 표목	개인명	태그700	\$a 개인명	
	단체명	태그710	\$a 기본요소	
	통일서명	태그730	\$a 통일서명	
	본서명과 다르게 부출되는 경우	태그740	\$a 부출서명	
연관 저록	원저저록	태그765	\$a 기본표목	
	번역저록	태그767	\$a 기본표목	
	부록저록	태그770	\$t 서명	
	모체레코드저록	태그772	\$a 기본표목	
	기본자료 저록	태그773	\$a 기본표목	
	이판저록	태그775	\$a 기본표목	
	기타형태저록	태그776	\$a 기본표목	
	선행저록	태그780	\$a 해당자료의 선행저록	
	후속저록	태그785	\$a 해당자료의 후속저록	
	비특정관계저록	태그787	\$a 연관저록 설명 주기	

<표1>에서 제시된 바와 같이, 고서용 KORMARC의 구조는 서명, 저자사항, 판사항, 간사사항, 형태사항, 주기사항, 총서사항, 소장처, 이미지, 요약·초록·해제, 언어, 주제명부출표목, 부출표목, 연관저록 등으로 나눌 수 있다. 이들의 기술형식은 다음과 같다.

첫번째, 서명 및 저자에 대한 사항은 태그 245에 서명, 권차, 저자로 세분화하여 기술한다.

두번째, 판사항은 태그 250에, 판차와 판종을 구별하여 기술하고 태그 008부호화정보필드와 연계되도록 한다.

세번째, 고서의 간사사항은 대부분 구체적으로 기재 안된 경우가 많아 本記, 刊印, 印記, 寫記, 序, 跋文 등을 근거로 기재하여 태그 260에 기술한다.

네번째, 형태사항은 태그 300에 고서의 장정 및 권 책 수, 삽도, 판식, 크기와 딸림자료를 구체적으로 기술한다.

다섯번째, 고서에 있어서 주기사항은 서명저자사항, 판사항, 간사사항, 형태사항, 총서사항 등에서 기록하지 못한 부가적인 기입이나 구체적 정보를 태그 5XX에 수록한다.

여섯번째, 총서사항은 태그 440와 태그 490에 , 소장사항은 태그 049와 태그 850에, 요약, 초록 및 해제 등은 태그 520에, 언어정보와 주기사항은 태그 041와 태그 546에 기술한다.

일곱번째, 주제명부출표목은 개인명 태그 600, 단체명 태그 610, 주제명 회의명 태그 611, 통일서명 태그 630, 일반주제명 태그 650, 지명 태그 651, 비통제색인어 태그 653을 포함한다.

여덟번째, 부출표목은 개인명 태그 700, 단체명 태그 710, 통일서명 태그 730, 서명 태그 740을 포함한다.

아홉번째, 연관저록은 원저저록 태그 765, 번역저록 태그 767, 부록저록 태그 770, 모체레코드저록 태그 772, 기본자료저록 태그 773, 이관저록 태그 775, 기타형태저록 태그 776, 선행저록 태그 780, 후속저록 태그 785, 비특정적 관계저록 태그 787을 포함한다.

3.1.3 고서용 XML 변환요소 매핑

고서용 XML 데이터로 변환하기 위해 변환 요소는 고서용 KORMARC 데이터의 완전한 변환을 고려하여 DC를 기반으로 하고 RDF구문을 응용하여 고서용 KORMARC 필드의 자세한 구조까지 매핑하고자 하였다.

XML 변환요소의 매핑에 DC를 기반으로 하고 RDF 구문을 응용한 것은 DC는 유사한 그룹형태의 다양한 서지정보를 표현하는데 가장 적합한 메타데이터이고, RDF는 DC구조를 사용하는 것이 가능한 문서자원이기 때문이었다.

고서용 KORMARC 데이터를 고서용 XML 데이터로 변환하는 과정에서 DC의 변환요소가 다른 메타데이터 요소와 상호관련이 있는지를 구체적으로 살펴보기 위해 MODS(Metadata Objected Description Scheme) ver3.0과의 비교를 <표2>과 같이 시도하였다.

<표2> DC요소 및 MODS 요소 비교

DublinCore요소	MODS ver 3.0 요소
Title	<titleInfo><title>
Creator Contributor	<name><namePart>
Subject	<subject> <topic> <name> <occupation> <classification>
Description	<abstract> <note> <tableOfContent>
Publisher	<originInfo> <publisher>
Date	<originInfo><dateIssued> <originInfo><dateCreated> <originInfo><dateCaptured> <originInfo><dateOther>
Type	<typeOfResource> <genre>
Format	<physicalDescription> <internetMediatype> <extent> <form>
Identifier	<identifier> <location><URL>
Language	<language>
Relation	<relatedItem>

DublinCore요소	MODS ver 3.0 요소
Coverage	<subject> <geographic> <temporal> <hierachicalGeographic> <cartographic>
Right	<accessCondition> <recordInfo>

<표2>는 DC 14개 요소와 MODS 23개 요소를 비교 한 것이다²⁴⁾. 이 표를 살펴보면, title은 서명의 역할어를 지칭한다. creator과 contributor은 저작자와 저작에 관계된 역할을 지칭한다. subject는 주제와 이름, 그리고 직업 등을 지칭한다. description은 초록 및 주기사항, 목차 등을 지칭한다. publisher는 발행 및 간사사항을 나타낸다. date는 간사년도 혹은 출간일 등을 나타낸다. type은 고서의 판형, 내용형식 등에 관한 정보를 나타낸다. format은 고서의 형태, 인터넷정보, 외형 등을 나타낸다. identifier는 소장위치 혹은 고서의 원문을 볼 수 있는 정보나 인식자 등을 나타낸다. language는 고서의 내용을 표기하고 있는 언어를 나타낸다. relation은 고서와 관련 된 정보를 나타낸다. coverage는 시대적인 구분 혹은 지명 등을 나타낸다. right는 접근정보를 나타낸다.

고서용 KORMARC의 변환모듈에서 적용하기 어려운 요소는 Language, Right 및 Identifier 등이었다. 이는 언어의 경우 고서의 내용이 모두 한자어로 표기되었다고 해서 중국어로 판단할 수는 없었고, 또한 접근정보와 소장위치를 식별하는 데 있어 URL로 표기하는 것은 기술적인 부분과 관련있는 사항에 해당되기 때문이다.

3.1.4 고서용XML의 DTD²⁵⁾구조 분석

²⁴⁾ <http://www.loc.gov/standard/mods/mods-dcsimple.html>

²⁵⁾ DTD(Document Type Definition)은 일반 XML에서 사용되는 속성값을 지닌 것이 아닌 형태만을 지정한 것으로 일반적인 DTD형태와는 차이가 있음을 밝혀둔다.

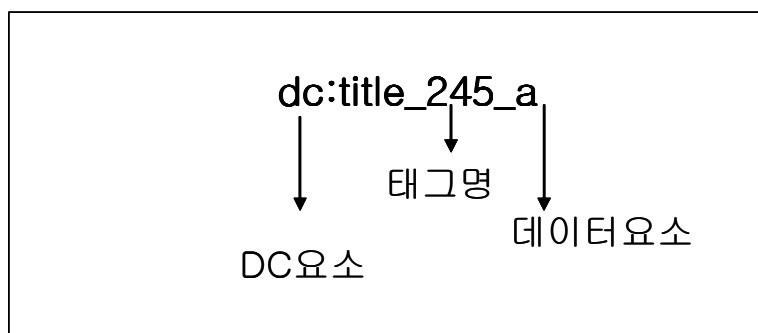
고서용 XML의 DTD 구조는 <부록1>에 제시된 Oldbooks.XSL 매핑요소를 참조하여 변환요소 구조구문과 KORMARC 변환요소로 나누어 분석하면 다음과 같다.

1) 변환요소 구조구문

첫째, 변환요소 구조구문의 처음 위치에는 DC의 13개 요소와 적용필드를 혼합하여 dc:format, dc:identifier, dc:type, dc:creator, dc:description, dc:title, dc:subject, dc:publisher, dc:contributor, dc:language, dc:relation, dc:date, dc:coverage, vcard:familyname 및 vcard:honorofname으로 표기하였다.

그 예로써 dc:title의 본서명인 경우 'dc:title_245_a'의 형식으로 기술하였다. 첫 번째 요소그룹은 DC요소를 지정하는 것이고, 뒤의 요소는 언더바를 사용하였는데 이는 데이터 변환주소 및 변환요소를 지칭하는 것이다.

변환요소의 구조구문은 <그림1>과 같다



<그림1> 고서용 변환요소 구조구문

<그림1>의 고서용 변환요소명 구조구문을 살펴보면 다음과 같다. 첫째, 변환 요소명 구조구문은 변환요소에서 DC요소, 연결자(-), 태그명, 연결자(-), 데이터 요소로 기호화되어 있다. 그중 DC요소는 변환요소의 명칭이 해당 주제를 알 수 있도록 구분한 것이다. 태그명은 고서용 KORMARC의 3자리 숫자를, 데이터요소는 고서용 KORMARC의 영문 지시자를 그대로 차용하였다.

둘째, 종단기호의 마크표기는 고려하지 않았다. 다른 DTD설계에 비해

제1지시기호와 제2지시기호에서 지시하는 내용을 삭제함으로서 데이터 변환 시 바로 참조될 수 있도록 하였다.

셋째, 요소명칭은 고서용 KORMARC의 부호화정보필드, 형태기술필드 및 가변길이필드에 나타난 모든 명칭을 부여하였다.

넷째, 태그 007 형태기술필드 및 태그 008 부호화정보필드를 적용하는데 있어서는 고서용 KORMARC의 판종 유형부호와 판종 내용부호에 제시된 대문자 캐릭터와 숫자기호를 따랐다.

다섯째, 저자사항에 나타나는 가족명, 자, 호, 범명, 범휘 등의 표기에는 vcard:family를 이용하였다. 또한 언어코드 태그 041 표기기호에 있어 dc:language를 표기했지만 고서에 표기된 언어보다는 출판당시의 국가, 국명, 왕조, 시기 등의 정보가 필요하다고 생각되어 dc:coverage 요소를 적용하였다²⁶⁾.

2) 고서용 KORMARC의 변환요소

첫째, 서명사항은 dc:title로 관련 요소는 115개이다. 태그 240의 통일서명과 태그 245의 서명저자사항, 태그 440의 총서사항/부출표목 및 태그 490의 총서사항/부출되지 않거나 다르게 부출되는 총서명, 태그 630의 주제명부출표목-통일서명, 태그 730 부출표목-통일서명, 태그 830 부출총서표목 및 태그 930의 로컬표목 통일서명 및 태그 949의 로컬표목 총서명이 매핑되었다.

둘째, 저자사항은 dc:contributor, dc:creator, vcard:family 등으로 변환 요소는 모두 75개이다. 그중 dc:creator는 관련요소가 60개이며 태그 008 전기 자서전, 태그 100의 개인명과 태그 110의 단체명이 매핑되었다. dc:contributor는 9개, vcard:family는 6개가 매핑되었다.

셋째, 판사항은 dc:format 195개 및 형태기술필드의 일부분, dc:type의 부호화정보필드 102개에 매핑되었다.

넷째, 간사사항은 dc:publisher에 12개가 매핑되었으며, 태그 008의 간사국명, 태그 100의 하위기관과 회의장소, 태그 260의 간사지 간사자, 태그 590의 소장본주기사항, 태그 787의 간사사항 등이 이 범주에 속한다.

²⁶⁾ 예를들면 태그 100의 지시기호3과 4에는 가족명과 개인의 자와 호 등이 표시되는데 이를 namecard 처리한다. 태그 041 언어부호는 고서의 경우 거의 한자로 되어있어 대부분 “chi”로 표기한다. 그러나 고서는 당시 한국, 중국, 일본 등에서 생산된 자료를 포함하고 있어서 언어표기보다는 자료가 생산된 국가를 표기하는 방안이 고려되어야 한다고 생각되어 이를 제기한 것이다. 통합목록에 있어서는 주기사항이나 부출표목에서 표기하는 것으로 대체 할 수도 있다.

간사년도는 dc:date의 26개 요소에 매핑되었다. 태그 008의 입력일자, 간사년 유형, 수록년도와 태그 033의 사건의 일시, 태그 260의 간사년과 인쇄년, 간사사항에 대한 정보, 태그 630의 발행년도, 태그 773, 태그 775, 태그 776, 태그 780, 태그 785, 태그 910의 회의 및 조약의 서명일자 등이다.

다섯째, 형태사항은 dc:format으로 관련요소는 197개이다. 태그 008의 판종과 태그 300의 장정 형태, 면 장수, 권 책수, 삽도, 판식, 딸린자료 태그 037의 자료형태, 태그 533의 복제주기, 태그 534의 원본형태사항에 매핑되었다.

여섯째, 주기사항은 dc:relation과 dc:description으로 관련요소는 125개이다. relation의 원명칭은 relation resouce인데 약칭을 사용하여 dc:relation으로 표기하였으며 관련요소는 53개이고, dc:description은 72개이다. 이들 요소는 태그 040의 목록작성기관정보, 태그 500 태그 501의 일반주기, 태그 504의 참고인용주기, 태그 504 태그 505 태그 510의 서지 인용 해제주기, 태그 520의 요약, 초록, 해제, 태그 521 인용 대상주기, 태그 524 태그 534 인용 및 원본주기, 태그 534 원본관련표출어, 태그 540 인용 및 복제, 태그 541 입수처주기, 태그 546 언어주기, 시대제목 및 부록주기, 태그 890의 변형문자표시, 기타 잡정보 및 태그 950의 로컬정보에 매핑되었다.

일곱째, 언어는 dc:language이며 관련요소는 15개이다. 태그 041의 언어와 원저작 언어, 태그 546의 언어주기, 태그 600 태그 610 태그 630 태그 700 태그 710 태그 730 태그 900 태그 910 태그 930의 언어, 태그 775의 언어부호 등에 매핑되었다.

여덟 번째, 주제요소는 dc:subject로 관련요소는 48개이다. dc:subject에는 주제별 분류기호를 포함하며, 주제명부출표목에 있어 태그 600, 태그 610, 태그 650의 일반 주제, 태그 653의 색인어에 매핑되었다.

아홉 번째, 식별기호는 dc:identifier로 관련요소는 92개이다. 제어번호 및 소장, 지역 등록번호, ISBN, 분류기호, 청구기호, 태그 535의 소장처주기, 태그 772의 모체레코드저록, 태그 940의 자료식별 표시등이 여기에 포함된다.

열 번째, dc:coverage는 관련요소가 13여개이다. 태그 033의 지역검증

부호, 태그 045자료의 연대범위, 태그 651의 지명 및 지리세목, 태그980의 소장 권 책 사항 등에 매핑되었다.

위에서 살펴본 내용에 의거하여 고서용 KORMARC의 변환 요소 비율을 <표3>으로 제시하면 다음과 같다

〈표3〉 고서용 변환요소 비율

변환요소 DC	요소개수	비율
title	115	14.11%
creator	60	7.36%
vcard	6	0.74%
contributor	9	1.10%
publisher	12	1.47%
format	195	23.93%
type	102	12.52%
relation	53	6.50%
description	72	8.8%
language	15	1.84%
subject	48	5.89%
identifier	92	11.29%
coverage	11	1.35%
date	25	3.07%
총 합 계	815	100%

위의 〈표3〉에서 나타나는 바와 같이, DC요소로 분류하여 고서가 지닌 특성 중 많은 비중을 차지하는 요소를 살펴보면, format 23.93%, title 14.11%, type 12.52%순으로 나타났다. 이 세 가지 요소들이 상위 51%이상을 차지하고 있는데 형태사항과 관련된 요소가 가장 많고, 다음은 서명 및 형식에 관한 요소가 비중이 높은 것을 알 수 있다.

3.2 고서용 KORMARC의 XML 변환모듈 개발

고서용 KORMARC의 변환요소를 기반으로 하여 KORMARC 레코드가 고서용 XML데이터로 변환될 수 있는 변환모듈을 개발한다. 변환모듈의 개발목적과 개발환경은 다음과 같다.

3.2.1 변환모듈 개발 목적

본 연구의 변환모듈은 세 가지 사항을 고려하여 개발하고자한다.

첫째, 변환모듈은 기본적으로 고서용 XML로 변환하는 과정에서 고서용 KORMARC 데이터의 정보자원이 손실이나 오류가 없도록 하였다.

둘째, 데이터의 변환은 파일 불러 오기 기능을 두어, KORMARC 레코드 데이터를 XML데이터로 변환시키는 과정을 표출하도록 하였다.

셋째, 형태기술필드 부분에 대한 편집기능을 두어, 고서용 KORMARC 데이터의 변환레코드의 특징 및 편의성을 돕고자 하였다.

3.2.2 변환모듈의 개발 환경

고서용 KORAMRC 데이터를 고서용 XML 데이터로 전환하고 XML데이터 형태로 보여주는 변환모듈은 다음과 같이 개발 하였다. 변환모듈은 설계 시 고서용 KORMARC의 완전한 변환을 고려하여 각 필드의 자세한 구조까지 개발하고자 했으나, 개발단계에서 상위 및 하위필드를 연관 지워주는 기능을 구현하는데 한계가 있어 DC를 참조한 변환모듈 구조를 개발하였다

이렇게 개발된 변환모듈이 운용될 수 있는 개발환경은 언어, 실행파일, 개발알고리즘 등으로 살펴보면 다음과 같다.

기본운용환경은 Windows 98SE(Standard Edition) 버전이상의 환경에서 실행되도록 하였으며, 개발언어는 Visual Basic 6.0이고 Visual Basic 관련 DLL을 사용하였다. 웹 브라우저는 MS Internet Explorer6.X이상을 지원하도록 하였다. MARCXML²⁷⁾ conversion 및 디렉터리를 참조하였다. 변환모듈의 실행 파일명은 MARC2XML.exe이다. 변환모듈 개발알고리즘은 고서용 KORMARC으로 입력된 데이터가 변환모듈 XML변환데이터로 처리되도록 하였다.

변환모듈에서 제외된 것은 데이터변환과정에서 오류사항에 대한 원 데이터의 원인을 밝혀주는 것과 마크업으로 교체해 주는 관리기능 등이다.

이를 토대로 기본적인 레지스트리에서 변환되는 결과를 보여주는 <부록2>는 XML2MARC를 차용한 스타일시트와 MARC handler 소스 일부이다. 이 스타일시트는 고서용 KORMARC 데이터를 고서용 XML 메타데이터로 변환하는데 필요한 레코드필드의 변환 위치를 규정하고 있다.

²⁷⁾ <http://www.loc.gov/MARC>

이를 응용하여 변환모듈을 구현하였다.

3.3 XML변환모듈의 구현

이 절에서는 <부록2>에서 제시한 고서용 XML 변환모듈 요소를 참조하여 변환이 가능하도록 구현방안을 제시하고, 구현된 화면의 내용을 자세히 살펴보고자 한다.

3.3.1 구현방안

변환모듈의 구현 알고리즘은 다음과 같이 이루어진다.

첫째, 리더 ,디렉터리, 및 가변길이 데이터필드 부분을 분리 후 고서용 KORMARC에서 규정한 최상위 논리적 데이터그룹인 00X 제어필드, 01X-09X 숫자필드 태그 010-090, 1XX 기본표목 태그 100-130, 20X-20X 서명 및 서명관련사항 240~245, 250 판사항, 260 간사사항, 300 형태사항, 4XX 총서사항 태그 440, 태그 490, 5XX 주기사항 태그 500-590, 6XX 주제명부출표목 태그 600-653, 태그 700-740부출표목, 76X-78X 연관저록, 830 총서부출표목, 9XX 로컬필드²⁸⁾의 요소를 추출한다.

둘째, 추출 된 데이터 그룹에 해당되는 변환 프로시저를 호출하여 고서용 KORMARC의 세 자리 표시기호를 이용하여 해당 데이터 요소에 상당하는 가변길이필드 데이터를 불러온다.

셋째, 레코드필드에서 출현한 모든 데이터 그룹이 XML 요소별로 인식되고 일대일처리방식으로 다음 단계에 레코드내용을 보여주게 된다.

넷째, 각 가변길이 필드내에 포함된 하위필드들은 식별기호를 이용하여 분리해 낸 다음 변환요소에서 정의한 규칙에 따라 마크업한 후 데이터변환이 이루어지도록 한다.

3.3.2 구현화면

변환모듈의 구현화면은 파일, 변환(MARC->XML변환), MARC편집의

²⁸⁾ 강순애 2006 (앞의책 p.199)재인용

3가지 메뉴로 구성되어 있다.

첫째, 파일메뉴는 KORMARC데이터 및 일반 텍스트용 파일을 불러오는 기능에서 메타데이터의 변환 결과를 저장하는 것이다.

둘째, 변환메뉴는 MARC데이터를 XML로 변환하는 기능을 가진 메뉴이다. 변환메뉴에서 MARC->XML변환을 클릭하면 고서용 KORMARC 데이터에서 XML데이터로 변환이 이루어지며, 변환된 원 소스데이터를 화면에 나타내는 것과 이를 웹 뷰어에서 볼 수 있는 브라우저 보기 기능이 있다.

셋째, MARC편집메뉴는 007 태그의 내용을 추가 편집할 수 있는 기능으로 구성되어 있다. 그중 자료범주메뉴에는 고서 유형을 먼저 선택하고 특정자료 표시는 권축장, 절첩장, 호접장, 표배장, 선장, 족자, 접포, 모장, 낱장, 기타 순으로 선택된다.

각 메뉴의 구현화면은 파일메뉴는 <그림2>와 <그림3>, 변환메뉴는 <그림4>, MARC편집메뉴는 <그림5>로 제시하여 설명하고자 한다.

、

고서MARC 변환기

파일(F) 변환(C) MARC편집(M)

파일 및 XML 관련 [MARC 편집]

상 태 n	레코드형 w	서지수준 m	입력수준	기술형식 k	연 관	간사년형 n
간사년도		간사국명 ko	삽도표시	광 락 a	사 경	판 구
매 미 e	자료형식	내용형식 t	수정여부	판 중 A24	목록전거	문학형식
전 기	언 어 chi					

```

>> [001][ ] OUB2005006222
>> [005][ ] 20050801190022
>> [007][ ] oe
>> [008][ ] 050801n ko a g t A24 chi
>> [012][ ] aKOL000012960
>> [035][ ] a(011001)KOL000012960
>> [040][ ] a011001 c011001
>> [052][00] a1231 b68
>> [085][0 ] a1231 2KDCP
>> [245][00] a周易諺解/ d胡廣(明)...[等]奉勅撰
>> [250][ ] a金屬活字本(戊申字)
>> [260][ ] a[刊寫地未詳]: b[刊寫者未詳], c[刊寫年未詳]
>> [300][ ] a9卷5冊: b四周單邊 半郭 25.8 x 16.6 cm, 界線, 10行18字 註雙行, 內向2葉花紋魚尾; c33.6 x 22.1 cm
>> [590][0 ] c印記: 昌寧世家, 曹氏命協, 李春熙印, 蒼史
>> [653][ ] a경서 a주역 a주역언해
>> [700][1 ] a호광 f(명)
>> [730][0 ] a경서.

```

<그림 2> 주역언해 MARC 편집 화면

고서MARC 변환기

파일(F) 변환(C) MARC편집(M)

파일 및 XML 관련 [MARC 편집]

상 태 n	레코드형 w	서지수준 m	입력수준	기술형식 k	연 관	간사년형 n
간사년도		간사국명 ko	삽도표시 a	광 락 a	사 경	판 구 a
매 미 e	자료형식	내용형식 m	수정여부	판 중 C00	목록전거	문학형식
전 기	언 어 chi					

```

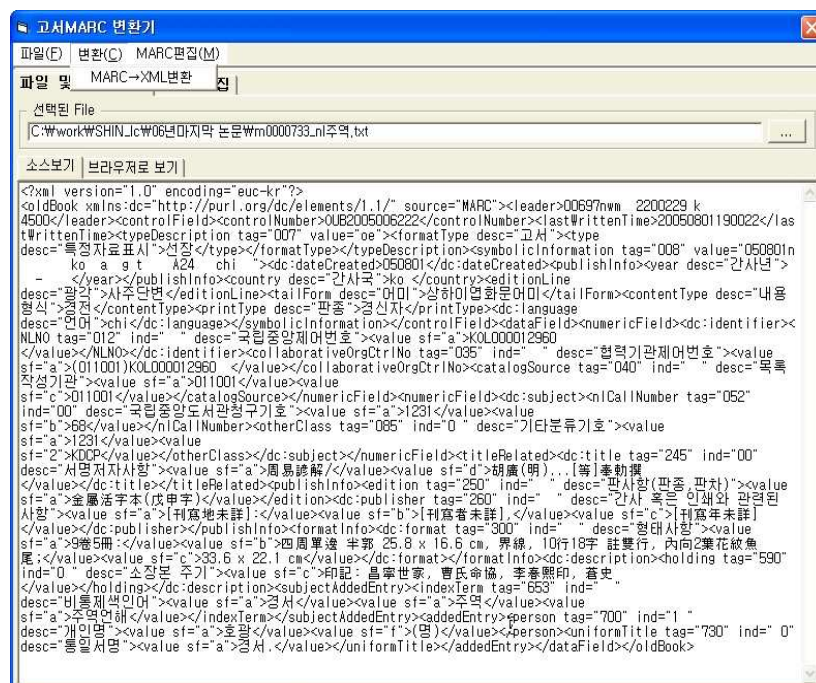
>> [001][ ] OUB2005148552
>> [005][ ] 20051018163742
>> [007][ ] oe
>> [008][ ] 050923n ko aa ae m C00 chi
>> [012][ ] aKOL000028736
>> [035][ ] a(011001)KOL000028736
>> [040][ ] a011001 c011001
>> [052][00] a0236 b3
>> [085][0 ] a0236 2KDCP
>> [245][00] a妙法蓮華經. n卷1-6/ d戒環(宋) 解; e一如 集註; e尹師路...等奉敎雕
>> [250][ ] a木板本
>> [260][ ] a[刊寫地未詳]: b[刊寫者未詳], c[刊寫年未詳]
>> [300][ ] a6冊: b樺圖, 四周單邊 半郭 20.9 x 17.9 cm, 9行17字, 註雙行, 黑口, 內向黑魚尾
>> [500][00] p刊記: 天順七年(1463)...尹師路等奉敎雕造
>> [653][ ] a불전 a법화경 a묘법연화경
>> [700][1 ] a계관 f(송)
>> [700][1 ] a윤사로
>> [700][1 ] a일여
>> [730][0 ] a불전.

```

<그림3> 묘법연화경 MARC 편집 화면

<그림2>와 <그림3>은 MARC포맷 파일을 불러와 정해진 규칙에 따라 MARC 및 XML스타일에 대응하는 태그에 저장된 데이터를 웹으로 표현

하는 것이다. <그림3>은 주역언해, <그림4>는 표법연화경의 서지데이터를 국립중앙도서관의 KOLIS-Net에서 마크테이터를 다운받아 변환모듈에서 반입시킨 화면이다. 변환모듈에서 고서용 KORMARC 데이터를 반입하여 저장한 결과 데이터의 손상이 없이 구현되었다.



<그림4> MARC 데이터를 XML로 변환 한 소스보기 화면

<그림4>는 고서용 KORMARC를 고서용 XML데이터로 변환한 원 소스테이터를 보여주는 화면이다. 변환모듈에서는 선택된 MARC데이터를 XML으로 변환하는 기능은 물론 변환한 파일의 저장 및 보관이 가능하도록 하였다. <그림4>에 제시된 원 소스테이터를 태그별로 정리하여 제시하면 다음과 같이 표시된다.

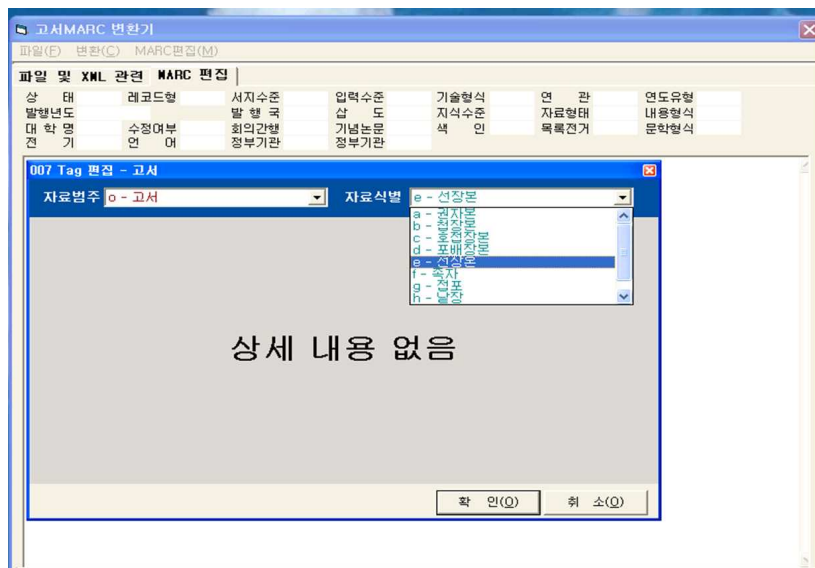
```
<?xml version="1.0" encoding="euc-kr"?>
<oldBook xmlns:dc="http://purl.org/dc/elements/1.1/" source="MARC">
<leader>00720nwm 2200217 k 4500</leader>
<controlField><controlNumber>0000024460</controlNumber>
<lastWrittenTime>20000528063020</lastWrittenTime>
<symbolicInformation tag="008" value="990331q18011834ulk a kor ">
<dc:dateCreated>990331</dc:dateCreated>
<publishInfo><year desc="간사년">1801-1834</year></publishInfo>
<country desc="간사국">ulk</country>
<authorityList desc="목록전거">국립중앙도서관</authorityList>
<dc:language desc="언어">kor</dc:language>
</symbolicInformation></controlField>
```

```

<dataField><numericField><catalogSource tag="040" ind=" " desc="목록작성기관">
<value sf="a">341054</value>
</catalogSource>
<localCallNumber tag="090" ind=" " desc="자관 청구기호">
<value sf="a">K1-208</value></localCallNumber>
</numericField><numericField>
<dc:subject><otherClass tag="085" ind=" " desc="기타분류기호">
<value sf="a">M16-1404</value>
</otherClass>
</dc:subject>
</numericField>
<mainEntry>
<dc:creator><person tag="100" ind="1" desc="개인명"><value sf="a">선조,
</value>
</person>
</dc:creator>
</mainEntry>
<titleRelated>
<dc:title tag="245" ind="10" desc="서명 저자사항">
<value sf="a">주역언해 /</value>
<value sf="d">선조(조선왕) 명찬</value>
</dc:title>
</titleRelated>
<publishInfo>
<edition tag="250" ind=" " desc="판사항(판종,판차)">
<value sf="a">정유자복각판.</value>
</edition>
<dc:publisher tag="260" ind=" " desc="간사 혹은 인쇄와 관련된 사항">
<value sf="c">순조연간(1801-1834).
</value>
</dc:publisher>
</publishInfo>
<formatInfo>
<dc:format tag="300" ind=" " desc="형태사항">
<value sf="a">7권 5책(권5, 6절) :</value>
<value sf="b">사주쌍변, 반곽 23.4 x 16.8 cm, 유계, 반엽 10행19자, 주쌍행, 내향 3엽화문어미. ;
</value><value sf="c">33.6 x 22.3 cm. 선장.
</value></dc:format>
</formatInfo>
<dc:description>
<general tag="500" ind="00" desc="일반주기">
<value sf="a">표제:주역언해.
</value>
</general>
<general tag="500" ind="00" desc="일반주기">
<value sf="a">지질:저지.
</value>
</general>
<additions tag="530" ind="00" desc="이용가능한 다른형태 자료 주기">
<value sf="c">인 : 사관장, 무주적상산사고소장본, 이왕가도서지장.
</value>
</additions>
</dc:description>
</dataField>
</oldBook>

```

Tag 편집화면



<그림5> 형태기술필드 태그 편집화면

<그림5>는 태그 007의 형태기술필드 편집을 콤보 박스 화면으로 구현한 화면이다. 이는 다음 장에서 언급될 고서용 KORMARC을 지원하는 목록 프로그램군을 보면 형태화 기술 필드 및 부호화 정보 필드부분에 대한 입력을 고서용 KORMARC에 정의된 문자와 숫자 기호를 입력하여 자동생성 시켜서 변환 레코드에 대한 설정값을 변경하는 불편함을 보완하고자 한 것이다.

3.4 XML변환모듈에서 사용된 스타일시트의 의의

고서용 XML 변환모듈에서 사용된 스타일시트(Oldboos XSL mapping 요소)는 고서용 KORMARC 데이터를 XML변환모듈로 개발하기 위한 것이며, 변환에 대한 표준규칙을 적용하기 위한 것이다. 이 스타일시트가 갖는 특징을 파악하기 위해 DC 요소(A), 단행본용 KORMARC(B), 고서용 KORMARC(C), Oldbooks XSL 요소(D), SNU 고도서 스키마(E), DLS-SOLARS DC(F)를 <표4>으로 비교하였다.

첫째, A의 dc:title 요소는 B에서 단행본 MARC의 본표제, C의 본서명에 대응하는 것으로, D에서는 dc:title_245_a로 구조화 하였고, E에서는 snuterm:s maintitle, F에서는 dls:maintitle로 제시되었다.

둘째, 형태사항인 dc:format의 요소는 B의 태그300에, D, E에서는 dc:format, F에서는 dls:page로 표현되었다.

셋째, dc:publisher 요소는 간사지의 의미를 나타내고자 발행지 dc:publisher_260_a에 배치하였다. dc:coverage, dc:description, dc:relation source 등 각 해당요소에 배치되는 요소명이 일치하는가를 살펴보고 고서용 KORMARC의 Oldbook.XSL 매핑이 다른 MARC데이터의 XML변환과 유사성을 지니고 매핑 되었는지를 판별할 수 있다.

넷째, dc:subject는 주제별에 따른 분류필드를 나열했다. dc:type은 단행본용 KORMARC와 DLS용에서는 고서용이 아닌 일반사항을 나타내므로 요소를 일치시키지 못했다.

다섯째, dc:language, dc:right, dc:identifier 및 기타에 있어서는 해당요소에 알맞게 배치되었는지를 비교하였다.

〈표4〉 DC 요소별 스키마 및 비교 자료

DC요소(A)	KORMARC필드(B)	KORMARC(교서용), C	Oldbooks,XSL요소(D)	SNU고도스키마(E)	DLS_SOLAPS DC(F)
dc:Title	245\$a(본표제)	245\$a 본서명	dc:title_245_a	snutrms:mainTitle	dis:mainTitle
	130(통일표제)	130\$a 총서명	dc:title_130_a	dcutrms:alternative	xxx
	210(축약표제)	xxx	xxx		
	222(등록표제)	xxx	xxx		
	245\$x(대등표제)	245\$x	dc:title_245_x	dcutrms:alternative	dis:parallelTitle
	245\$b(부표제)	245\$b 잔제	dc:title_245_b	snutrms:otherTitle	dis:sub Title
	xxx	500\$a 일반주기	dc:description_500_a	snutrms:TranslatedTitle	dis:generalNote
	500\$e(부서명)	xxx	xxx		
	500\$f(판권기서명)	xxx			
	500\$g(표지서명)	xxx		snutrms:headingTitle	
	500\$h<책등서명>	xxx			
	500\$i(번역서명)	xxx			
	730(부출통일표제)	730: 판심제	dc:title_730	snutrms:placeCentralTitle	xxx
	740(부출표제)	740	dc:relation_740	snutrms:otherTitle	dis:addedTitle
dc:Creator	100(기본표목-개인명)	100/245\$d ,	dc:creator_100_a	snutrm:creatorName	dc:creator
dc:Contributor	110(기본표목-단체명)	110	dc:contributor_110_a	xxx	dis:Organization
	111(기본표목-회의명)		xxx	xxx	dis:conferenceName
	245\$d(저자)	245\$d 첫번째 저자	dc:creator_245_d	snutrms:creatorname	dis:Person, vCard:FN
	245\$e(두번째 이하 저자)	245\$d 첫번째 저자	dc:creator_245_e	snutrms:editor Name	vCard:FN
	250\$b(해당판저자)	250\$b 당해판 저자	dc:format_250_b	snutrms:creatorOriginalNa	xxx
	500\$d(저자주기-연속)	500\$d 저자에관한주기	dc:description_501_d	snutrms:creatorOriginalNa	xxx
	508(제자주기-연속)	xxx	xxx		
	511(연주자/배역진-비도서)	xxx	xxx		
	700(부출표목-개인명)	700	dc:creator_700		
	710(부출표목-단체명)	710	dc:contributor_710	xxx	dis:Organization
	711(부출표목-회의명)	xxx	xxx	xxx	dis:conferenceName
		900\$d			
		007/00=o 고서			

dc:Format	300 (원태기술사항)	300	dc:format_300	dc:format	dis:page
dc:Publisher	260\$b (발행자)		dc:publisher_260_a		
		500\$n 판사항	dc:description_500_n	xxx	xxx
dc:Coverage	043 (지역부호)	043 지역	dc:coverage_043	dcterms:spatial	dis:place
	045 (연대부호)	045 연대부호	dc:coverage_045	dcterms:temporal	dcq:issued
	651\$y (시대세목)	651\$y 주제명부호-시대	dc:coverage_651_y	dcterms:temporal	xxx
	651\$a (지역)	651\$a 주제명부호-지역	dc:coverage_651_a	dcterms:spatial	xxx
dc:Description	520(요약조록해제 주기)	520	dc:description_520	dc:description ,dcterms: abq	dcq:abstract
	500\$a(일반주기)	500\$a	dc:relational_500_a	xxx	dis:generalNote
dc:Relation Source	247(이전/변경표제)		xxx		
	440, 490 (출서사항)	440, 490 출서 및 부호	dc:title_440, dc:title_490	dcterms:isFormatOf	xxx
	501(할월본주기)	501	dc:relation_501	xxx	xxx
	505(내용주기)	505	dc:relation_505	dcterms:tableOfContent	dis:contentNote
		506 접근제 한주기	dc:relation_506	xxx	xxx
	507(원서주기)	507 원저자, 원서명	dc:creator_507	snul:metaAccess	snul:metaAccess
	510(인용주기)	510 인용주기	dc:relation_510	xxx	xxx
	530(이용가능한한다른형태)	530 이용가능한한다른형태	dc:relation_530	xxx	xxx
	533(복제 주기)	533 복제 주기	dc:relation_533	xxx	xxx
	534(원본주기)	534 원본주기	dc:relation_534	xxx	xxx
	581(참조정보원주기)	581 참조정보원	dc:relation_581	xxx	xxx
	765(원저지록)	765 원저지록	dc:relational_765	snul:metaAccess	snul:metaAccess
	767(번역지록)	767 번역지록	dc:relational_767	xxx	dis:translatedTitle
	775(이판지록)	775 이판지록	dc:relational_775	xxx	xxx
	776(기타형태지록)	776 기타형태지록	dc:relational_776	xxx	xxx
	780(선행지록)	780 선행지록	dc:relational_780	xxx	xxx
	785(후속지록)	785 후속지록	dc:relational_785	xxx	xxx
dc:Subject	056\$a(KDC)	056 한국십진분류기호	dc:subject_056	xxx	dis:kDC
	080(UDC)	080 국제십진분류기호	dc:subject_080	xxx	xxx
	082(DDC)	082 듀이십진기호	dc:subject_082	xxx	xxx
	085(기타분류기호)	085 기타분류기호	dc:subject_085	xxx	xxx

	600(주제 명부출-개인명)	600(주제 명부출-개인명)	dc:subject_600		
	610(주제 명부출-단체명)	610(주제 명부출-단체명)	dc:subject_610		
	611(주제 명부출-회의명)	xxx	xxx	xxx	xxx
	630(주제 명부출-동일 표제)	630(주제 명부출-통일서명)	dc:title_630	xxx	xxx
	650(주제 명부출-일반주제)	650(주제 명부출-일반주제)	dc:subject_650	xxx	xxx
	651(주제 명부출-일반지명)	651(주제 명부출-일반지명)	dc:coverag_651	xxx	xxx
	653(비통제 주제명)	653 비통제 색인어	dc:subject_653	xxx	dis:keyword
dc:Type	Leader/06(레코드 형태)	w 고서	oldbooks		
	007/00 (자료범주 표시)	0=고서	oldbooks		
	256(컴퓨터화일 특성)	xxx			
dc:Date	260\$c(발행년)	xxx	dc:date_260_c		
		500\$p 간사사항	dc:date_501_p		
dc:Language	008/ 35-73(언어부호)				
	041\$(언어부호)	041 언어부호	dc:language_041	dc:language	dcq:ISO639-2
	546 (언어주기)	546 언어주기	dc:language_546	dc:language	dcq:ISO639-2
dc:Right	020\$ (	020 ISBN	dc:identifier_020	xxx	rdf:li id='ISBN'
dc:Identifier	010(LCCN)	xxx			
	015(NBN)국가서지번호	xxx			
	020(ISBN)	020 bb\$a	dc:identifier_020_a	xxx	rdf:li id='ISBN'
	022(ISSN)	022 bb\$a	dc:identifier_022_a	xxx	rdf:li id='ISSN'
	856(전자접근위치)-'USM#	xxx			
기타	042(검증기관부호)	042 검증부호	dc:coverage_042	xxx	xxx
	주기정보(500\$a)	500\$a 일반주기	dc:description_500_a	xxx	xxx
	주기정보(502)	xxx			
	주기정보(510)	510 인용주기	dc:description_510_a	xxx	xxx
	주기정보(533)	533 복제주기	dc:format_533	xxx	xxx
	주기정보(534)	534 원본주기	dc:description_534	xxx	xxx
	주기정보(546)	546 언어주기	dc:relation_544	dc:language	dcq:ISO639-2
	권호표시(362)	xxx			

이 변환모듈 스타일시트의 특징은 다음과 같다.

첫째, 본 스타일시트는 고서용 KORMARC의 변환을 고려하여 형태기술필드, 부호화정보필드 및 가변길이필드까지 변환개발을 시도하였다.

둘째, XML의 문서 변환에는 DTD를 사용하는 방법, 스타일시트를 이용하는 방법, 레지스트리를 이용하는 방법 등이 있다. 고서용 XML변환모듈은 국내에서 스타일시트 및 변환 MARC handler를 이용하여 개발되었다.

셋째, 고서용 XML 변환모듈은 고서용 KORMARC의 형식을 고려하여 개발되었는데 메타데이터 변환과정에 있어 고서용 데이터의 XML 데이터 변환에 국제 표준화안을 도입하여, 이에 대한 변환결과를 <그림2>로 확인하였다. 이는 서지 표준과 국제적인 표준안을 수용한 개발이 데이터 변환과 스타일시트에도 영향을 주며, 이로서 표준화가 필요하며, 표준화에 따른 개발모듈이 국내의 고서데이터를 변환하는 데 주요한 지침을 적용하는데 그 의의를 둘 수 있다.

4. 고서용 KORMARC XML변환모듈과 다른 변환 프로그램과의 비교

이 장에서는 우선 비교 프로그램을 선정하는 기준을 세우고, 평가기준에 적합한 SOLARS와 XMLAS프로그램을 중심으로 도입 시기 및 활용, 마크입력 및 변환화면, 특징에 치중하여 고찰하였으며, 또한 고서용 KORMARC XML변환모듈과 다른 변환프로그램을 비교하였다

4.1 비교 프로그램 선정기준

우선 비교 프로그램을 선정하는 평가기준은 ①메타데이터 입력의 호환성 ② 파일정보 출력 및 리포트 기능 ③ 관리기구로서 에디팅 기능을 고려하였다.

또한 비교 프로그램이 보유하고 있는 기능에 대해서는 다음과 같은 사항을 고려하였다.

첫째, 고서용 KORMARC레코드가 고서용XML 데이터로 목적에 맞게 변환되는 과정을 조사하였다. 데이터의 변환은 텍스트 형식의 입력을 통한 변환과 파일형식을 불러오기 기능을 통한 다수 데이터의 변환이 가능한지를 살폈다.

둘째, 데이터를 변환할 때 적합한 레코드 내 변환 요소들을 선택 할 수 있도록 유저 인터페이스를 고려하였다.

셋째, 변환프로그램의 기능과 함께 편집기능의 유무를 조사하여 필요할 때 언제든지 변환 결과를 수정하여 저장이 가능하도록 개발된 프로그램인지를 살폈다.

넷째, MS Explorer 6X이상 웹 브라우저에서 지원 가능한 프로그램을 가지고 있는지를 살폈다.

위에서 제시한 평가기준과 기능을 보유한 고서용 목록 프로그램 중 KORMARC를 XML로 변환시키는 XSL기능을 지닌 것은 SOLARS 및 XMLAS 이다. 이들 프로그램을 중심으로 도입 시기 및 활용, 마크입력 및 변환화면, 프로그램의 특징에 치중하여 분석하고자 한다

4.2 SOLARS²⁹⁾

4.2.1 도입시기 및 활용

이 프로그램은 미 텔파이사의 T-50과 UTLAS 서지관리시스템을 모태로 하여 1989년 C언어를 중심으로 개발되었으며, 1994년부터 ‘학술정보시스템’으로 명칭 지어졌다. 2000년도 이후 업그레이드 버전을 준비하여 2002년 9월 SOLARS II가 가동되었다³⁰⁾.

메타데이터에 대한 참조는 DC 중심의 고지도, 고문서를 참조한 스키마구조를 설계하여 발표한 서울대 권고안을 일부 이용했다.

학술정보시스템은 중앙도서관 및 각 단위 대학 내에 50여개가 넘는 단위별 전문도서관의 자료를 통합하여 운영하며, 고도서데이터 외에도 전거데이터 같은 데이터를 관리하고 있는 것이 특징이다.

운영환경은 서지관리 서버환경이 UNIX이고 RDBMS는 Oracle8i기반으로 하고 있고 서지검색엔진은 BRS검색엔진에서 Search검색엔진으로 버전이 세분화되어 있다. 초기개발사는 한국컴퓨터(주)이나 1998년도 아이네크(주)로 회사를 분사하여 오늘에 이르고 있다. 초기에는 서울대중앙도서관과 산학협동으로 개발하여 국공립대학 및 사립대학으로 구분하여 관하였다. 현재는 서울대학교의 다른 대학도서관을 중심으로 한 서지관리종합시스템을 구축하여 운영하고 있다.

메타데이터에 대한 연구는 초기에는 목록프로그램을 개발한 벤더사보다는 공동 소유 및 개발 권한을 가진 서울대학교 중앙도서관을 중심으로 이루어졌다. 현재는 DC기반의 메타데이터를 기반으로 한 스키마 및 설계를 기초로 하여 DLS 및 SOLARS Version 4 (DLi)등이 운용하고 있고 MARC21의 기능을 수용한 목록시스템을 개발 운용중이다.

4.2.2 고서용 KORMARC의 입력 및 변환 화면

²⁹⁾ SOLARS(SeOuL Library Automation System)은 1992년 서울대학교중앙도서관과 아이네크(당시 한국컴퓨터)와 공동으로 개발을 진행한 프로그램으로 학술정보시스템이라 통칭된다.

처음 명칭은 특정대학의 종합목록시스템을 지칭하는 것으로 인식되었지만, 점차 고유한 기능을 가진 SOLARS가 되었고, 특히 서울대학교 중앙도서관과 아이네크에서 운용하는 SOLARS는 버전상의 차이가 있다. SOLARS 2명칭은 서울대학교중앙도서관에서 서지의 SOLARS와 당시 티지인포넷(퓨처인포넷)에서 개발한 DL시스템을 지칭한다.

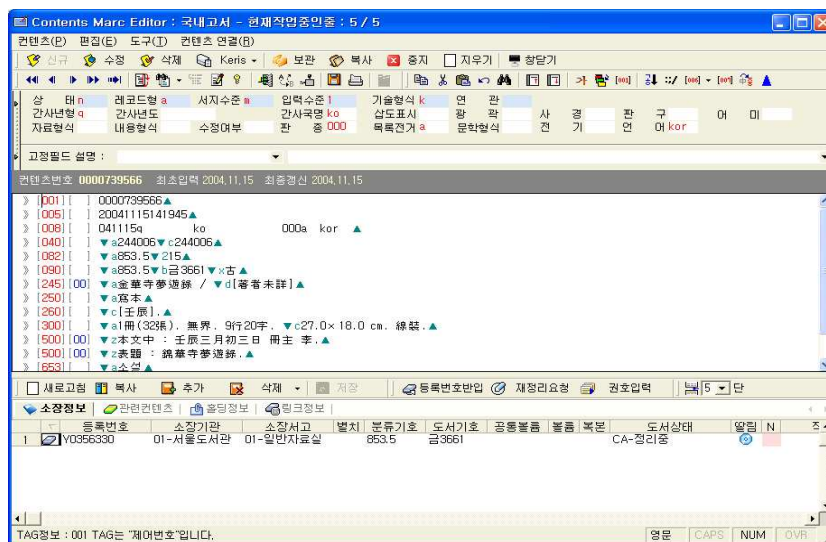
SOLARS는 SOLARS SE(Standard Edition) 버전과 SOLARS 4 & DLi 2등의 버전을 운용중이고 경상대의 SOLARS는 DLi 2버전이다. 본 글에서는 SOLARS로 통칭하기로 한다.

³⁰⁾ 서울대학교 중앙도서관 2005 서울대학교학술정보화 장기발전계획 p.4

SOLARS에서 고서 데이터를 처리 할 때는 고서용 KORMARC에 의해 신규로 입력하는 방법과 고서용 KORMARC에 의해 입력되지 않은 데이터를 수정하여 업데이트 하는 두 가지 방법이 있다. 초기에는 고서 데이터를 단행본용 KORMARC에 의해 입력을 하였는데 현재는 고서용 KORMARC로 수정하여 처리하고 있다.

XML 변환화면은 DB생성 스키마를 XML format으로 저장하여 서지관리시스템에서 전자 도서관 데이터와 연동되며 XML뷰어 보기는 XSL (eXtension Style Language) 디렉터리를 응용하는 방법이다.

이러한 SOLARS 목록시스템의 실제 과정을 보여주기 위해 <그림6>은 고서데이터 입력화면, <그림7>은 변환화면, <그림8>은 XML로 표출하는 화면을 제시하였다.



<그림6> 고서데이터 입력화면

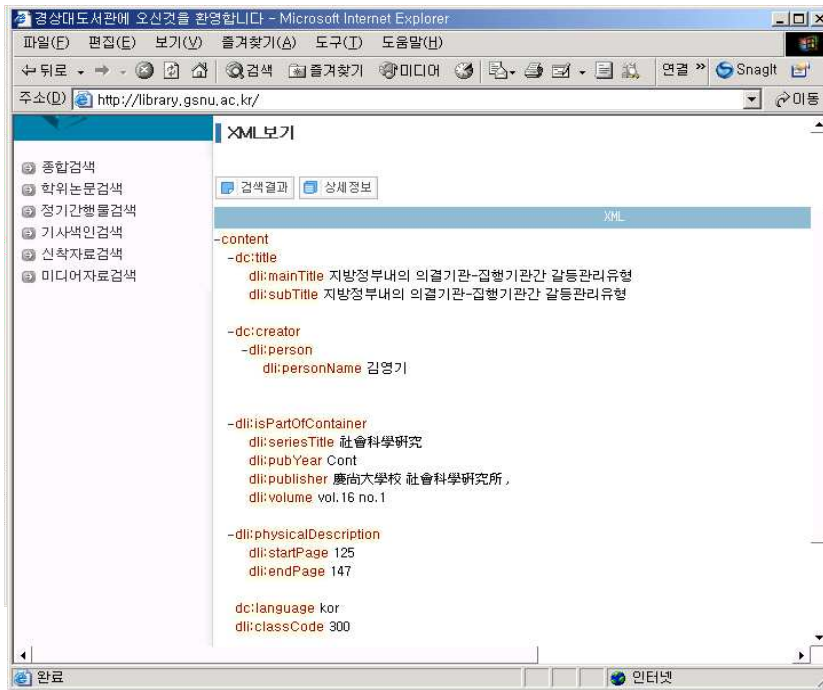
<그림7> 변환 화면

<그림7>³¹⁾은 데이터 표출에 있어 트리구조를 이용한 화면을 보여주고 있다.

<그림 8> XML로 표출하는 화면

웹 브라우저에서 XML변환으로 표현이 가능 한 것은 DC기반의 메타데이터 요소를 선정하고 이를 관리함으로써 고서용 KORMARC 레코드 데

31) 경상대학교 문천각의 고서변환 예시화면



이터의 변환을 가능하게 하는 것이다. <그림8>은 데이터를 XML 메타데이터로 변환한 화면이다. SOLARS에서는 XML을 DC에 의해 구조화하고 메타데이터 형태로 물리적인 DB 테이블을 설계하여 XML로 변환이 가능하도록 하였다.

4.2.3 SOLARS 변환 프로그램의 특징

SOLARS는 목록프로그램에서 고서용 KORMARC 포맷을 지원한다. 고서용 KORMARC 포맷에 의해 입력된 데이터는 XML데이터 형태로 저장되고 뷰어화면상에 제시되는 XML메타데이터 구조는 DC기반으로 한다.

SOLARS가 이러한 특징을 갖게 된 것은 2002년 SOLARS를 사용하는 서울대학교 중앙도서관이 디지털도서관 구축을 위한 XML스키마 메타데이터 구현에 관한 연구를 발표한 것이 그 시발점이다. 서울대학교 중앙도서관은 MARC 레코드가 갖는 구조적 경직성, 새로운 정보자원의 다양성을 지원하는 어려움, 정보자원간의 연결 관계의 표현을 지원하지 못하는 한계점을 지적하고, 이를 극복하기 위해 IFLA FRBR(Functional Requirements for Bibliographic Records)을 기반으로 한 메타데이터 모델과 XML 표현체계를 제시하였다.³²⁾ 이는 고도서 메타데이터의 설계에 있어

³²⁾ 서울대학교 중앙도서관, 2002, 전게서, pp.2-85.

서 더블린코어를 기본 메타데이터 요소로 적용함으로써 의미적 메타데이터의 상호운용성을 얻기 위한 결과라고 볼 수 있다.

하지만, 서울대학교 메타데이터 모델은 고도서 네임스페이스관리를 위해 DC의 15개 기본 요소 중 11개 요소만을 채용하여 고도서 메타데이터 스키마를 설계하였다.

기존의 SOLARS 2는 서울대가 제시하는 목록프로그램의 특징을 그대로 수용했지만 아이네크가 개발한 SOLARS 4(DLi)에서는 다음과 같은 특징을 보이고 있다.

첫째, 화면구조가 CMS(Content Management System) 개념으로 이루어져 있어서 고정길이필드는 따로 콤보 박스로 조정하여 필드자체를 이용하고자 하는 목적에 따라 생성 할 수 있도록 구성하였다.

둘째, Z39.50 Z Client Protocol을 이용하여 분담목록과 데이터를 공유하고, 별도로 전거데이터를 생성할 수 있도록 컨텐츠 개념을 도입하고 있다.

4.3 XMLAS³³⁾

4.3.1 도입시기 및 활용

XMLAS는 목록 소프트웨어 인 'VintageLAS 5.1'을 2004년에 업그레이드하면서 고서용 KORMARC 포맷을 지원하고 있다. VintageLAS는 1995년 당시 오름정보시스템과 (구)삼보정보시스템의 협력으로 VintageLAS 3.1를 발표하여 보급하였고 그 후 VintageLAS는 각 국립대학교를 제외한 대학도서관에서 가장 많이 사용하는 목록 소프트웨어로 자리매김하고 있다.

이 시스템의 운영환경 서버는 UNIX 환경이고, RDBMS는 Oracle[®]를 지원한다. XMLAS에서는 고서용 KORMARC에 의해 고서 서지데이터를 입력하면 XMDL은 XMLAS의 서지정보를 가져다가 DTD를 메타데이터로 변환하여 전자 도서관 모듈에서 XML을 지원하고 있다. XMDL(XML Base Millennium Digital Library)은 자원통합관리를 위한 전자 자료관리 프로그램이다. 이는 메타데이터 및 DTD데이터에서 제공된 메타데이터의 관리를 담당한다. XMDL은 Windows환경에 운영되고, 개발언어는 Visual basic 6.0 ASP(Active Server Page)이며, IIS 및 SQL Server(Oracle[®] 가능)를 이용하여 메타데이터 및 멀티미디어데이터를 통합 관리한다. 인쇄 자료는 MARC를 기반으로 하고 전자자료는 DC를 기반으로 한다.

4.3.2 고서용 KORMARC의 입력 및 변환 화면

XMLAS에서 고서 데이터를 처리할 때는 고서용 KORMARC에 의해 신규로 입력하는 방법과 고서용 KORMARC에 의해 입력되지 않은 데이터를 수정하여 업데이트하는 방법의 두 가지가 있다. 단행본포맷에서 입력한 고서데이터는 고서용 KORMARC으로 설정한 후에 변환이 가능하다. 고서용 KORMARC으로 변경한 다음에는 단행본에 입력한 사항 중 일부를 고서용에 알맞게 수정하거나 신규 생성 작업이 가능하도록 하였

³³⁾자료를 보유한 동국대학교 중앙도서관은 1996년 'VintageLAS' ver3.1시스템을 당시 삼보정보시스템과 공급계약을 체결한바 있다. 2004년 하반기 XMLAS 및 XMDL으로 업그레이드 했다.

다.

XML 변환화면은 MODS XSL를 이용한 구조이다. 즉 XML변환은 DTD를 별도로 관리하면서, XML변환 시에 DTD변환을 거친 XML형태의 데이터를 다시 MODS형태의 변환을 거쳐 웹 화면에 보여주는 구조이다.

이러한 XMLAS 목록시스템의 실제 화면을 보여주기 위해 <그림9>는 XMLAS 고서용 목록 프로그램의 서지 데이터를 입력 한 화면이고, <그림10>은 MODS 보기화면을 제시하였다.

레코드상태	레코드형태	서지수준	입력수준	목록기술형식	연관레코드	간사년유형
c	w	m		k		n
간사년 1	간사년 2	간사국명	삽도표시	a	b	관구
머미	자료형식	내용형식	수정레코드	판종	C00	문학형식
건기	언어	kor				

001 000000416856
 005 20050307115103
 007 oe
 040 d 211020
 049 0 l D023992 f D
 082 a 173
 090 a 173 b 박64o
 100 i a 박영진, f 조선, e 편수
 245 10 a 鄺解圖像童文先習 / d 朴永鎮(朝鮮) 編修; e 權世鎮(朝鮮) 校正.
 250 a 木板本
 260 a [韓國]: b [刊寫者未詳], c [刊寫年未詳].
 300 a 不分券1冊(19張): b 圖, 四周雙邊, 半郭 20.1 x 16.1 cm, 有界, 半葉 8行16字, 註雙行, 開註, 上內向黑魚尾; c 27.0 x 18.9 cm.
 500 00 a 國漢對譯本
 500 10 c 版心題: 童文先習
 500 00 a 紙質: 楮紙
 650 4 a 經部, 小學類
 653 a 童문선습
 700 i a 권세진
 740 0 a 동문선습

< 그림 9> XMLAS 고서용 목록 데이터 입력화면

<그림9>는 XMLAS 목록형식에서 고서용 화면에 대한 파일편집 기능을 보여주는 화면이다. 이는 간편한 입력화면구조로 되어 입력된 데이터가 변환한 구조를 보여주고 있다. 형태화필드와 부호화정보필드가 따로 관리가 되도록 설계되었으며, MARC의 편집기능 및 반입, 반출기능 그리고 보고서기능을 지니고 있어 변환된 데이터에 대한 교열 및 수정이 편리하도록 개발되어 있다.



<그림10> MODS 보기 화면

<그림10>은 XMLAS의 전자도서관 모듈에서 해당 서지사항을 변환하는 화면이다. 전자도서관 관리모듈인 XMDL에서 먼저 제어번호에 따른 서지정보를 불러오고 이 정보를 토대로 DTD구조를 요소화 한다. XML형태로 DTD는 데이터 필드 내에 담겨진 정보를 메타데이터 형태로 저장한다.

<부록3>은 동국대 소장 「묘법연화경」이 DTD요소에 의해 MARCXML을 응용 XSL을 거쳐 MODS데이터로 변환된 자료이다. 즉 XMLAS에서 XMDL데이터를 DTD변환을 거친 XML형태의 데이터를 다시 MODS형태의 변환을 거쳐 웹 화면에 보여주는 구조이다.

4.3.3 XMLAS 변환 프로그램의 특징

XMLAS는 VintageLAS 5.1Plus에서 업그레이드 된 것으로 고서용 KORMARC 포맷을 지원하며 목록시스템의 구조는 MARC21의 환경을 모델로 하고 있다.

XMLAS 목록모듈에서 고서용 데이터를 입력 후 전자도서관 서버에 기본정보와 KORMARC데이터를 넘겨주면 그 안에 별도의 데이터스키마 구조를 정해 놓고 XMDL에서 이를 토대로 XML형태로 변환한다. 변환된 XML 데이터는 다시 한번 전자도서관의 MODS보기에서 MODS 형태로 스타일시트를 생성하고, 데이터가 XSL을 거쳐 MODS 형식을 빌어 웹으로 나타나는 것이다.

변환과정에 있어 매핑 요소들은 MARC2XML Version 3을 사용하였다. XMLAS의 고서용 KORMARC 데이터는 2004년 후반 단행본포맷에서 고서용 포맷으로 변환한 것이다. 따라서 표목부와 형태기술부분에 있어 고서용 KORMARC에서 규정한 기술규칙대로 입력 설계되었다.

URL정보를 지원하여 웹자원 간의 연계가 가능하도록 하고, 고서용 KORMARC과 MARC21의 기본 목록형식을 지원한다. 원문보기에 있어서는 XMDL과 연동하여 고서용 KORMARC으로 입력된 서지데이터 필드를 XMDL에서 원문관리와 연계하며, 웹 화면에서는 이용자서비스에 중점을 두어 별도 MODS포맷을 XML보기용으로 서비스하고 있다.

4.4 고서용 KORMARC XML변환모듈과 SOLARS 및 XMLAS 비교평가

고서용 KORMARC XML 변환모듈과 SOLARS 및 XMLAS 프로그램을 비교 평가하기 위해, 운영환경 개발언어 XML 변환방식 등의 요소를 <표5>로 제시하고 그 세부내용을 비교하면 다음과 같다.

<표5> 고서용KORMARC의 XML변환모듈과 SOLARS 및 XMLAS의 비교

비교 요소	고서용KORMARC의 XML변환모듈(A)	SOLARS(B)	XMLAS(C)
운영환경	Stand Alone	Client/Server	Client/Server
개발언어	Visual Basic6.0	C #(Sharp)	Visual Basic6.0
XML변환방식	XSL marc handler	XSL	DTD XSL
MARCXML지원	MARCXML2.0	자체지원	MARCXML3.0
XML표준 지원	XML 1.0	XML 1.0	XML 1.0
단행본MARC데이터 변환지원	지원 및 변환가능	변환가능	변환가능
메타데이터 스키마 형식	DC RDF	DC Namespace	DTD 및 스키마
MARC 편집기능	자체변환가능	MARC21 지원	MARC21지원
관리도구	편집기능 부족	CMS기능	XMDL
확장성 및 호환성	인터페이스 연계	별도DB수정	XML데이터구축여부

첫번째, 운영환경은 고서용 KORMARC의 XML 변환모듈(이하 A라 칭함)은 Windows 98SE(Standard Edition) 버전이상 환경에서 실행되며, SOLARS(이하 B라 칭함) 및 XMLAS(이하 C라 칭함)는 서지관리 서버 환경이 UNIX이고 RDBMS는 Oracle8i기반으로 하고 있다.

두 번째, 개발 언어를 비교하면, A는 개발언어는 Visual Basic 6.0이고 Visual Basic 관련 DLL이 요구된다. B는 C #(Sharp)을 사용하며, C는

Visual basic 6.0 ASP(Active Server Page)이다.

세 번째, XML 변환방식에 있어서 A는 XSL MARC handler를 응용·개발하여 XML변환에 각 레코드 요소를 변환 시도 한 것이고, B는 XSL에 의거해 레코드 자체를 XML로 설정하여 레코드 변환을 수행하며 이 과정에서 MARC의 반입, 반출 과정을 위해 일부 변환에 따른 규칙을 차용하였다. C는 XMDL에서 별도로 XML변환을 수행하므로 향후 XML의 표준이 새롭게 발표되고 및 기술개발에 따른 인터페이스 구조 설계에 따라 응용이 자유로운 변환 방식이지만 XML변환은 DTD를 별도로 관리하며, XML변환 시에 새로운 MODS변환을 XMLMODS를 설정하여 보여 주는 것이다.

네 번째, MARCXML을 지원하는 것은 A는 MARCXML 2.0이고, B는 자체지원이며, C는 MARCXML 3.0이다.

다섯 번째, XML표준을 지원하는 것은 A와 B 및 C 모두가 XML 1.0이다.

여섯 번째, 단행본 MARC 데이터의 지원 및 변환에 있어서 A는 지원 및 변환이 가능하고, B와 C는 변환이 가능하다.

일곱 번째, 메타데이터 스키마의 형식에 있어서 A는 DC RDF이고, B는 DC Namespace이며, C는 DTD구조 이다.

여덟 번째, KORMARC의 편집기능에 있어서 A는 자체변환이 가능하고, B와 C는 MARC21을 지원한다.

아홉 번째, 관리도구기능 측면 차원에서는 A가 편집기능이 부족하고, B는 CMS기능이 있으며, C는 XMDL에 관리기능이 있다.

열 번째, 확장성 및 호환성에 있어서 A는 인터페이스에 의한 연계가 가능하고, B는 DB의 수정이 있어야 가능하며, C는 XML데이터의 변환 구축이 가능하지만 데이터 테이블을 별도 생성해야 한다.

위에서 비교한 내용을 종합하여 특징적인 것을 몇 가지로 요약하면 다음과 같다.

고서용 KORMARC XML변환모듈과 SOLARS 및 XMLAS프로그램의 공통점은 첫째, 고서용 KORMARC 형식을 지원하며. 특히 고서용이 표준화 이전에 생성된 레코드를 지원하거나 변환할 수 있는 도구가 있다. 둘째, 데이터변환에 있어 기능은 한 레코드 단위에서 부터 여러 개의 레코드단위 변환을 지원한다. 셋째. 인터넷 환경에서 지원하는 프로토콜을 지원 개발하고 있다. 넷째, 모두 표준화된 규약을 근거로 하여 설계 및 개발되어 일정한 XSL 스타일시트를 지원하면 메타데이터로의 변환이 가능하다

이들 프로그램의 차이점은 첫째, 각 개발자의 의도에 따라 고서용 KORMARC XML변환모듈과 XMLAS는 Visual Basic과 ASP언어로 개발되었고, SOLARS는 C#(샵)³⁴⁾언어로 Java환경을 지원하고 있다. 따라서 이들 프로그램 중 고서용 KORMARC XML변환모듈만이 PC에서 서버의 도움없이 운용이 가능하며, 나머지는 Client/Server환경에서만 운영이 가능하다. 둘째, XMLAS와 SOLARS는 Content Management의 기능을 지닌 개별요소를 관리하는 기능을 지원하지만, 고서용 KORMARC XML 변환모듈은 Content Management의 기능을 지원하지 않는다.

이들 프로그램 중 XML변환모듈이 지닌 장점은 첫째, 개발 및 구현 환경에 있어서 국제적인 표준 모듈인 MARCXML과 XML1.0 스펙을 준수하여 향후 다른 프로그램과의 상호연계 및 확장성을 지니고 있다. 둘째, XML 데이터변환에 있어 표준변환 방식 인 MARCXML ver2.0과 고서용 KORMARC이 제정되기 이전에 구축된 단행본포맷의 고서데이터를 변경할 수 있도록 개발되었다. 셋째, 고서용 KORMARC에서 명시한 부호화정보필드, 형태기술필드 및 가변길이필드 전체에 대해 DC요소와 매핑을 시도한 변환모듈로서, 이는 향후 새로운 메타데이터 형식과의 인터페이스 개발에 있어서 호환성이 높은 표준사례로 제시될 수 있다

³⁴⁾C언어 일종이며, 유니코드를 지원한다.

5. 결론

고서용 KORMARC의 XML 변환모듈 구현에 관한 연구를 통해 밝혀진 사실을 종합하면 다음과 같다.

첫째, 고서용 KORMARC는 고서의 기술적인 요소를 기술하도록 설계되어 2000년 KSX6006-6으로 제정되었다. 고서용 KORMARC는 단행본용 KORMARC 구조를 바탕으로 하면서도 고서 자체의 특성을 정확하게 기술하기 위해 매우 복잡하게 구조화되어 있고, 네트워크 자원의 통합적 기술이나 새로운 데이터 요소의 추가 및 기존 포맷을 수정하는데 있어서는 확장성이 결여되고 있다. 또한 정보자원의 소재에 접근에 대한 정보를 생성할 수가 없으며 필드간의 연결이 어려운 점도 지적되었다. 이러한 고서용 KORMARC의 문제점을 보완할 수 있는 메타언어로서는 웹상에서 구조화된 문서를 전송 가능한 형식으로 설계 표준화된 XML이 가장 적합한 것으로 평가되었고, 메타데이터로서는 데이터의 호환성을 유지하고 네트워크자원의 기술에 필요한 일련의 데이터요소를 규정하여 자원의 신속한 검색을 목적으로 하는 Dublin Core와 XML구문론을 사용하고 DC와의 호환이 가능한 RDF Syntax가 가장 적합한 것으로 평가되었다.

둘째, 고서용 KORMARC의 고서용 XML 변환모듈을 설계하여 구현하는 작업은 고서용 KORMARC데이터를 XML데이터로 변환하여 KORMARC의 단점들을 보완하고 데이터 통합을 위한 정보기술의 응용성을 점검하는데 그 목적을 두었다.

고서용 XML 데이터로 변환하기 위한 변환 요소 설계에 있어서는 고서용 KORMARC 데이터의 완전한 변환을 고려하여 고서용 KORMARC의 구조를 분석하고 변환형식은 DC요소를 기반으로 하고 RDF Syntax를 일부 적용한 변환요소 매핑을 통한 구조구문을 작성하였다. 변환 요소명 구조구문은 변환요소에서 DC요소, 연결자(-), 태그명, 연결자(-), 데이터 요소로 기호화되어 있다. 그중 DC요소는 변환요소의 명칭이 해당 주제를 알 수 있도록 구분 한 것이다. 태그명은 고서용 KORMARC의 3자리 숫자를, 데이터요소는 고서용 KORMARC의 영문 지시자를 그대로 차용하였다. 요소명칭은 고서용 KORMARC의 부호화정보필드, 형태기술필드 및 가변길이필드에 나타난 모든 명칭을 부여하여 총 815개의 요소가 변화되

었으며 DC요소로 분류하여 고서가 지닌 특성 중 많은 비중을 차지하는 요소는 format 23.93%, title 14.11%, type 12.52%순으로 나타났다.

셋째, 고서용 KORAMRC 데이터를 고서용 XML 데이터로 전환하고 XML데이터 형태로 보여주는 변환모듈은 고서용 KORMARC 데이터의 정보자원이 손실이나 오류가 없도록 하는데 치중하였다. 기본운영환경은 Windows 98SE(Standard Edition) 버전이상의 환경에서 실행되도록 하였으며, 개발언어는 Visual Basic 6.0이고 Visual Basic 관련 DLL을 사용하였다. 웹 브라우저는 MS Internet Explorer6.X이상을 지원하도록 하였다. MARCXML conversion 및 디렉터리를 참조하였다. 변환모듈의 실행 파일명은 MARC2XML.exe이다. 변환모듈 개발알고리즘은 고서용 KOR MARC로 입력된 데이터가 변환모듈 XML변환데이터로 처리되도록 하였다.

넷째, 변환모듈의 구현화면은 파일, 변환(MARC->XML변환), MARC 편집의 3가지 메뉴로 구성되어 있다. 파일메뉴는 KORMARC데이터 및 일반 텍스트용 파일을 불러오는 기능에서 메타데이터의 변환 결과를 저장하였고, 변환메뉴는 MARC데이터를 XML로 변환하여 변환된 원 소스 데이터를 화면과 웹 뷰어에서 볼 수 있도록 하였다. MARC 편집메뉴는 007 태그의 내용을 추가 편집할 수 있는 기능으로 구성하였다.

다섯째, 고서용 KOR MARC XML변환모듈과 SOLARS와 XMLAS프로그램의 운영환경, 개발언어, XML 변환방식 등의 요소를 비교평가하였다. 이들 프로그램의 공통점은 첫째, 고서용 KORMARC 형식을 지원하며. 특히 고서용이 표준화 이전에 생성된 레코드를 지원하거나 변환할 수 있는 도구가 있다. 둘째, 데이터변환에 있어 기능은 한 레코드 단위에서부터 여러 개의 레코드단위 변환을 지원한다. 셋째, 인터넷 환경에서 지원하는 프로토콜을 지원 개발하고 있다. 넷째, 모두 표준화된 규약을 근거로 하여 설계 및 개발되어 일정한 XSL 스타일시트를 지원하면 메타데이터로의 변환이 가능하다

이들 프로그램의 차이점은 첫째, 각 개발자의 의도에 따라 고서용 KORMARC XML변환모듈과 XMLAS는 Visual Basic과 ASP언어로 개발되었고, SOLARS는 C#(샵)³⁵⁾언어로 Java환경을 지원하고 있다. 따라서 이들 프로그램 중 고서용 KORMARC XML변환모듈만이 PC에서 서버의

³⁵⁾C언어 일종이며, 유니코드를 지원한다.

도움없이 운용이 가능하며, 나머지는 Client/Server환경에서만 운영이 가능하다. 둘째, XMLAS와 SOLARS는 Content Management의 기능을 지닌 개별요소를 관리하는 기능을 지원하지만, 고서용 KORMARC XML 변환모듈은 Content Management의 기능을 지원하지 않는다.

이들 프로그램 중 XML변환모듈이 지닌 장점은 첫째, 개발 및 구현 환경에 있어서 국제적인 표준 모듈인 MARCXML과 XML1.0 스펙을 준수하여 향후 다른 프로그램과의 상호연계 및 확장성을 지니고 있다. 둘째, XML 데이터변환에 있어 표준변환 방식 인 MARCXML ver2.0과 고서용 KORMARC이 제정되기 이전에 구축된 단행본포맷의 고서데이터를 변경할 수 있도록 개발되었다. 셋째, 고서용 KORMARC에서 명시한 부호화정보필드, 형태기술필드 및 가변길이필드 전체에 대해 DC요소와 매핑을 시도한 변환모듈로서, 이는 향후 새로운 메타데이터 형식과의 인터페이스 개발에 있어서 호환성이 높은 표준사례로 제시될 수 있다

참 고 문 헌

- 강순애. 『문헌조직과 정보운용』 서울:아세아문화사,1989. pp101-106
- _____. 고문헌 정보시스템 구축 및 전망. 「한국문헌정보학회지」 31(4), pp83-112
- _____. 한국문헌자동화목록형식 및 기술규칙(안) -고서용제정과 목록기술의 발전 방향에 관한 연구. 「서지학연구」 제18집, pp27-58.
- _____. 『고문헌조직과 정보활용』 서울: 아세아문화사, 2005.
- 국립중앙도서관. 『한국문헌자동화목록형식-고서용』 서울: 국립중앙도서관, 2000
- _____. 『한국문헌자동화목록기술규칙-고서용』 서울: 국립중앙도서관, 2000.
- 김이겸, 김태수. 메타데이터변환과 자원기술구조의연구 「정보관리학회지」 15(3)
- 김태수. "더블린코어(DublinCore)와 메타데이터"
- <http://dewey.yonsei.ac.kr/course/metadata/DC.hwp> (2006.4.15) 2000.
- 김현희, 장혜원. 1999. 디지털도서관 문서양식으로서의 XML과 HTML특성 및 검색기능비교 「정보관리학회지」 16(2)
- 도태현. 『한국의 목록규칙 변천사』 서울:한국도서관협회, 2003.
- _____. 한국의 자동화목록법 관련 규칙 발전과정에 관한 연구-KORMARC을 중심으로, 「한국문헌정보학회지」 36(2), pp55-71
- 윤구호, 천혜봉, 김태수, 김이겸. 한국문헌자동화목록법(KORMARC)에 관한 연구-고서용포맷 및 기술규칙 「정보관리학회지」 10(1), pp3-12.
- 류부현. 국립중앙도서관제정 고서용기술규칙(안)에 관한 연구 . 「서지학연구」 제18집 pp59-76.
- 리상용. XML을 활용한 고문헌의 원문디지털화방안에 관한연구 「한국문헌정보학회지」 37(1) pp171-201.
- 문헌정보처리연구회. 『메타데이터형식과구조;문헌정보처리연구회시리즈7』 서울: 동연구회, 1987
- 박문렬. 1999. 한국문헌자동화목록기술규칙의 통합성에 관한 연구 「서지학연구」 제18집. pp 7-26.
- 박재혁. 고서자동화목록형식에관한 연구 「서지학연구」 제18집 pp77-101.
- 박하영 외. XML을 이용한 TEI문헌구조화 연구 「지식처리연구」 1(2), 2000
- 서울대학교중앙도서관. 2002. 『 디지털 도서관 구축을 위한 XML 스키마 메타데이터 구현에 관한 연구 ; 최종보고서』 서울 : 서울대학교중앙도서관
- _____. 『서울대학교중앙도서관 디지털 자료 구축을 위한 표준화 연구 ; 최종보고서』 서울 : 서울대학교 중앙도서관, 2002.

- 서울:서울대학교중앙도서관, 2002.
- 안영선,이은철,오삼균 공저 . 2000. 학위논문의 XML DTD설계에 관한 연구.
『정보관리학회지』 17(4)
- 이재선. KORMARC의 현재와 미래 「도서관연구회논문집」 2002, pp185-190.
- 이지원 외. 2000. RDF를 적용한 더블린코어와 vCard의 연결방안연구
『지식처리연구』 1(2).
- 정진식. 『자료편목록 : 이론과 실제』 서울 : 한성대학교출판부, 2005.
- 조윤희, 이두영 RDF 기반 인터넷자원 메타데이터 설계에 관한 연구
『정보관리학회지』 17(3) pp147-170.
- 한국전산원. 1998. 『SGML/XML문서시스템구현에 관한 연구』
논문
- 구영옥. 한국고문서의 기술요소 선정과 고문서 XML DTD 설계
숙명여자대학교대학원 석사학위논문. 2002.
- 김경미. 국내고문헌 처리를 위한 한국형 EAD개발에 관한 연구
전주대학교대학원 석사학위논문, 2000.
- 김달옥. 고서용 OPAC시스템의 기능 비교분석 -국립중앙도서관,
규장각, 존경각, 장서각을 중심으로-, 부산대학교대학원 석사학위논문, 2004.
- 김상미. MARC기반 서지레코드 핵심데이터 요소에 관한 연구,
중앙대학교대학원 석사학위청구논문, 1999.
- 김성은. KORMARC의 DTD 및 변환 프로그램 개발연구,
연세대학교대학원 석사학위논문, 1997.
- 김은혜. 동양고서목록의 전산화를 위한 표준화 연구
숙명여자대학교대학원 석사학위논문, 1986.
- 김정하. 고서의 효율적인 정보조직체계, 숙명여자대학교 대학원
석사학위논문, 2003.
- 김중진. XML을 이용한 디지털도서관의 서지정보시스템 설계 및 구현,
중앙대학교대학원 석사학위논문, 2001. pp61-65.
- 박천수. 디지털도서관에서의 XML Repository 시스템, 충남대학교대학원
석사학위논문, 2001
- 안혜순. 문서검색시스템에서 DOM을 이용한 XML 문서 편집기의 설계 및
구현 경일대학교산업대학원 석사학위논문, 2001.
- 양충식. 디지털 도서관 시스템에서의 가상문서 브라우징을 위한 XML 문서처리
충남대학교대학원 석사학위논문, 2000.

- 엄혜련. TEI Header와 KORMARC의 비교분석 및 매핑 알고리즘개발
숙명여자대학교대학원 석사학위논문, 1998.
- 이지영. 고서의 디지털화 색인에 관한 연구 이화여자대학교 대학원
석사학위논문, 1998.
- 임점미. XML문서의 데이터베이스 저장도구 설계 및 구현 성신여자대학교
교육대학원 석사학위, 2000.
- 장혜원. XML과 HTML로 제작한 웹문헌의 특성 및 검색기능 비교연구
명지대학교대학원 석사학위논문, 1998.
- 정나영. PDF를 적용한 한국의 고문서 디지털화 방안에 관한 연구
숙명여자대학교대학원 석사학위논문, 2002.
- 조윤기. 다양한 XML문서에 대한 통합관리 시스템 충북대학교대학원
박사학위논문, 2002.
- 조윤희. XML/KORMARC통합시스템 구현에 관한 연구
중앙대학교대학원 박사학위논문, 2001.

인터넷자원

- KORMARC www.nl.go.kr
- MARCXML <http://www.loc.gov/standards/marcxml/>
- XML Schema <http://www.w3.org/1999/06/REC-xml-디렉터리>
- XSLT <http://www.w3.org/TR/2002/WD-xslt20-20021115>
- Xlink <http://www.w3.org/TR/xlink>
- Namespace <http://www.w3.org/TR/REC-xml-names>
- MODS <http://www.loc.gov/standards/mods/>
- RDF Syntax <http://www.w3.org/TR/PR-rdf-syntax/>
- Xpath <http://www.w3.org/TR/path>
- http://dewey.yonsei.ac.kr/courses/metadata/rdf/rdf_model_syntax_spec_kor.html
- DCRB www.folger.edu/bsc/dcrb/dcrbrev.htm

ABSTRACT

An Implementation of XML Transformation Module of Old Books on KORMARC Format

Shin, Ick Chul

Major in Information Science

Dept. of Library and Information Science

Graduate School

Hansung University

Nowadays, advancement in computer technology has brought various web information technologies as objects of bibliographical description. Limitation on the existing MARC standards has been demonstrated by the movement of sharing information through internet. New information system environment has integrated with web resources, in order to share digital information resources. It, however, requires maintaining compatibility with the existing MARC bibliography management program, and using expansive metadata.

The purpose of this research is to overcome the shortcomings of KORMARC by transforming KORMARC data for old books to XML data, and review the application of information technology for data integration.

The research conducted three different methods to make a XML transformation program of KORMARC for old books.

First, a theoretical approach on meta language, which is made for XML KORMARC format, description rule for old books, and for description of old literatures, and meta data DublinCore (DC) and RDF Syntax will be carried out.

Second, DC type XML elements were designed so that XML transformation program of KORMARC for old books can transform on the web. XML elements were designed through the structural analysis of KORMARC for old books. The XML transformation model was developed based on the international standards.

Third, a comparison between SOLARS and XMLAS was made. Amongst the domestic index programs and KORMARC XML

transformation program for old books, both SOLARS and XMLAS support a format for old books, and can convert into XML data.

Conclusions driven by the research are the following:

1) KORMARC for old books were designed in year 2000 as KSX6006-6, in order to describe the descriptive elements of old books. KORMARC for old books have a complicated structure because it focused on accurate description of the characteristics of old books, and at the same time based the structure of KORMARC on independent volumes. In addition to the complicated structure, it lacks expansion for changing formats and adding new data, and integrating description for network resources. Also, criticism has been made that it is difficult to create information access to source of information resources, and that it is difficult to connect among fields.

As meta language, standardized XML, which has a format to send structured documents on web, was evaluated to be one of ways to overcome KORMARC's problems. As metadata, DC, XML phrase theory and RDF Syntax was evaluated to be the most appropriate method. DC has compatibility with other data, and can stipulate the needed data for describing network resources. Also, DC's purpose is to look up the data quickly. RDF Syntax has good compatibility with DC.

2) In order to transform XML data for old books, the structure of KORMARC for old books were analyzed in designing the transformation elements. Transformation format was based on DublinCore elements, and structure phrases were written by mapping transformation elements with some of RDF Syntax application.

3) Transformation program has focused on limiting information loss of KORMARC data and errors. The program converts KORMARC data for old books to XML data for old books, and demonstrated the data in the formation of XML data.

4) Screen shot of transformation program consists of three menu: file, conversion (MARC XML), and MARC editing. File menu has saved the conversion result of metadata from file opening function for KORMARC data and general text. Conversion menu has shown the one source data, which has been converted from MARC data to XML, on the screen and web viewer. MARC editing menu is composed of a function that can addition edit the context of 007 tag.

The merits of the XML transformation program among these programs are first, the program has met the international standard

program MARCXML and the spec of XML 1.0 in development and screen shot environment that it has mutual linkage and expansion with other programs. Second, it enabled a change for old book data of independent formats, which was established before MARCXML ver.2.0 and KORMARC for old books.

Third, it has attempted mapping with DC elements for signaling information field, format description field and changeable length field stated in KORMARC for old books. This can be an example demonstrating high compatibility in developing interfaces with new metadata formats.

고서용 KORMARC의 XML 변환모듈 구현에 관한 연구

신 익 철

오늘날은 컴퓨터기술의 발전으로 서지기술의 대상물로서 새로 출현한 다양한 웹정보자원의 기술과 인터넷상에서 이러한 정보자원을 공유하려는 움직임은 기존의 표준화된 목록규칙인 MARC의 한계를 드러내게 되었다. MARC는 자료처리를 위한 목록에 근거를 둔 메타데이터로서 기종이 다른 컴퓨터간의 호환성이 보장되지 않고 새로운 레코드 자원의 필드정보를 확장하는데 어려움이 있으며, 레코드간의 연결기능이 부족 한 것이 단점으로 지적되어 왔다. 디지털화된 정보자원을 공유하기 위해 웹 자원과 통합한 새로운 정보시스템환경은 기존에 구축된 MARC서지관리프로그램과 호환성을 유지하고 확장성을 가진 메타데이터 활용이 요구되었다.

따라서 본 연구는 고서용 KORMARC데이터를 XML데이터로 변환하여 KOR MARC의 단점들을 보완하고 데이터 통합을 위한 정보기술의 응용성을 점검하는데 그 목적을 두었다.

고서용 KORMARC 포맷의 XML 변환모듈 구현을 수행하기 위해 다음과 같은 방법으로 연구를 진행하였다.

첫째, 고서용 KORMARC 목록형식 및 기술규칙, 고문헌 기술을 위한 메타언어로 XML 및 메타데이터인 DublinCore(DC)와 RDF Syntax에 대한 이론적 배경을 고찰하였다.

둘째, 고서용 KORMARC의 XML변환모듈이 웹 상에서 변환이 가능하도록 고서용 KORMARC의 구조분석을 통해 DC type의 XML요소를 설계하였고, 국제 표준규격을 바탕으로 고서용 XML변환모듈을 개발하였다.

셋째, 고서용 KORMARC XML변환모듈과 국내에서 사용중인 목록프로그램 중 고서용 형식을 지원하고 있으며 XML데이터로 변환기능이 있는 SOLARS 및 XMLAS를 비교하였다.

이와 같은 방법으로 연구하여 얻은 결론은 다음과 같다.
고서용 KORMARC는 고서의 기술적인 요소를 기술하도록 설계되어 2000년

KSX6006-6으로 제정되었다. 고서용 KORMARC는 단행본용 KORMARC 구조를 바탕으로 하면서도 고서 자체의 특성을 정확하게 기술하기 위해 매우 복잡하게 구조화되어 있고, 네트워크 자원의 통합적 기술이나 새로운 데이터 요소의 추가 및 기존 포맷을 수정하는데 있어서는 확장성이 결여되고 있다. 또한 정보자원의 소재에 접근에 대한 정보를 생성할 수가 없으며 필드간의 연결이 어려운 점도 지적되었다. 이러한 고서용 KORMARC의 문제점을 보완할 수 있는 메타언어로서는 웹상에서 구조화된 문서를 전송 가능한 형식으로 설계 표준화된 XML이 가장 적합한 것으로 평가되었고, 메타데이터로서는 데이터의 호환성을 유지하고 네트워크 자원의 기술에 필요한 일련의 데이터 요소를 규정하여 자원의 신속한 검색을 목적으로 하는 Dublin Core와 XML구문론을 사용하고 DC와의 호환이 가능한 RDF Syntax가 가장 적합한 것으로 평가되었다.

2) 고서용 XML 데이터로 변환하기 위한 변환 요소 설계에 있어서는 고서용 KORMARC 데이터의 완전한 변환을 고려하여 고서용 KORMARC의 구조를 분석하고 변환형식은 DC요소를 기반으로 하고 RDF Syntax를 일부 적용한 변환요소 매핑을 통한 구조구문을 작성하였다. 변환 요소명 구조구문은 변환요소에서 DC요소, 연결자(-), 태그명, 연결자(-), 데이터 요소로 기호화되어 있다. 그중 DC요소는 변환요소의 명칭이 해당 주제를 알 수 있도록 구분한 것이다. 태그명은 고서용 KORMARC의 3자리 숫자를, 데이터 요소는 고서용 KORMARC의 영문 지시자를 그대로 차용하였다. 요소명칭은 고서용 KORMARC의 부호화정보필드, 형태기술필드 및 가변길이필드에 나타난 모든 명칭을 부여하여 총 815개의 요소가 변화되었으며 DC요소로 분류하여 고서가 지닌 특성 중 많은 비중을 차지하는 요소는 format 23.93%, title 14.11%, type 12.52%순으로 나타났다.

3) 고서용 KORAMRC 데이터를 고서용 XML 데이터로 전환하고 XML데이터 형태로 보여주는 변환모듈은 고서용 KORMARC 데이터의 정보자원이 손실이나 오류가 없도록 하는데 치중하였다. 기본운영환경은 Windows 98SE(Standard Edition) 버전이상의 환경에서 실행되도록 하였으며, 개발언어는 Visual Basic 6.0이고 Visual Basic 관련 DLL을 사용하였다. 웹 브라우저

는 MS Internet Explorer 6.X 이상을 지원하도록 하였다. MARCXML conversion 및 디렉터리를 참조하였다. 변환모듈의 실행 파일명은 MARC2XML.exe이다. 변환모듈 개발알고리즘은 고서용 KOR MARC로 입력된 데이터가 변환모듈 XML변환데이터로 처리되도록 하였다.

4) 변환모듈의 구현화면은 파일, 변환(MARC->XML변환), MARC편집의 3가지 메뉴로 구성되어 있다. 파일메뉴는 KORMARC데이터 및 일반 텍스트용 파일을 불러오는 기능에서 메타데이터의 변환 결과를 저장하였고, 변환메뉴는 MARC데이터를 XML로 변환하여 변환된 원 소스데이터를 화면과 웹 뷰어에서 볼 수 있도록 하였다. MARC 편집메뉴는 007 태그의 내용을 추가 편집할 수 있는 기능으로 구성하였다.

5) 고서용 KOR MARC XML변환모듈과 SOLARS와 XMLAS프로그램의 운영환경, 개발언어, XML 변환방식 등의 요소를 비교 평가하였다. 이들 프로그램의 공통점은 첫째, 고서용 KORMARC 형식을 지원하며, 특히 고서용이 표준화 이전에 생성된 레코드를 지원하거나 변환할 수 있는 도구가 있다. 둘째, 데이터변환에 있어 기능은 한 레코드 단위에서 부터 여러 개의 레코드단위 변환을 지원한다. 셋째, 인터넷 환경에서 지원하는 프로토콜을 지원 개발하고 있다. 넷째, 모두 표준화된 규약을 근거로 하여 설계 및 개발되어 일정한 XSL 스타일시트를 지원하면 메타데이터로의 변환이 가능하다 이들 프로그램의 차이점은 첫째, 각 개발자의 의도에 따라 고서용 KORMARC XML변환모듈과 XMLAS는 Visual Basic과 ASP언어로 개발되었고, SOLARS는 C#(샵)³⁶⁾언어로 Java환경을 지원하고 있다. 따라서 이들 프로그램 중 고서용 KORMARC XML변환모듈만이 PC에서 서버의 도움없이 운용이 가능하며, 나머지는 Client/Server환경에서만 운영이 가능하다. 둘째, XMLAS와 SOLARS는 Content Management의 기능을 지닌 개별요소를 관리하는 기능을 지원하지만, 고서용 KORMARC XML 변환모듈은 Content Management의 기능을 지원하지 않는다.

이들 프로그램 중 XML변환모듈이 지닌 장점은 첫째, 개발 및 구현환경에 있어서 국제적인 표준 모듈인 MARCXML과 XML1.0 스펙을 준수하여 향후 다른 프로그램과의 상호연계 및 확장성을 지니고 있다. 둘째, XML 데이터변

³⁶⁾C언어 일종이며, 유니코드를 지원한다.

환에 있어 표준변환 방식 인 MARCXML ver2.0과 고서용 KORMARC이 제정되기 이전에 구축 된 단행본포맷의 고서데이터를 변경할 수 있도록 개발되었다. 셋째, 고서용 KOR MARC에서 명시한 부호화정보필드, 형태기술필드 및 가변길이필드 전체에 대해 DC요소와 매핑을 시도한 변환모듈로서, 이는 향후 새로운 메타데이터 형식과의 인터페이스 개발에 있어서 호환성이 높은 표준사례로 제시될 수 있다