

Topic Map 기반의 MARC 적용 방안 연구

장 화수*, 고 일주*

A Study on MARC Based Topic Map

Hwa-Su Jang*, Il-Ju Ko*

요 약

문헌정보처리 표준화도구인 MARC는 포맷의 문제점과 다양한 웹자원 메타데이터 정보조직의 문제점으로 인하여 웹 기반의 XML표준포맷의 도입을 시도하였고, MARCXML로 변환되어 시스템간 상호운용되고 있으나, MARCXML은 서지정보의 의미특성이나 메타데이터의 표현을 고려하지 않고 단순히 MARC 레코드의 표현을 XML 구조로 변환한 것일 뿐이다. 시맨틱의 핵심기술로 부각되고 있는 Topic Map은 XML기반의 표준기술언어인 ISO의 XTM을 이용해 정보와 지식의 분산 관리를 지원하는 기술이다. 학술정보자원에 대한 DB 구축 시 Topic Map언어인 XTM을 이용한다면 이미 개발된 여러 메타데이터 등을 한곳으로 통합 하면서도 신축성과 확장성을 제공하는 것이 용이하게 된다. 하지만, 기존 시스템에서 새로운 Topic Map을 구축하는 것은 많은 비용과 시간이 소요 되는 등 어려운 일이다.

본 연구에서는 기 구축된 학술DB로부터 Topic Map에서 재활용할 수 있는 요소들을 추출하기 위한 정보 소스로서 데이터베이스 스키마와 MARC에서 언급하는 메타데이터를 이용하는 것은, XML의 특징인 시스템간 상호운용성을 확보함과 동시에 기초 학문자료의 복잡한 관계의 개념구조, 자료유형 및 자료간의 의미적 상관관계 등을 표현에 있어 효율적인 개발방법임을 제안한다.

▶ Keyword : 도서관 시스템, MARC, KORMARC, MARCXML, 시맨틱웹, 토픽맵, xtm, 온톨로지

• 제1저자 : 장희수
* 숭실대학교 미디어학부

1. 서론

학술정보와 지식을 분류, 구축 하는 일은 한 번에 완성하고 종료되는 작업이 아니고, 지속적으로 이용자 연구에 의해서 수정과 보완 작업을 거치게 된다. 이러한 특성 때문에 대부분의 DB들은 다수의 메타데이터 관련 표준 활동과 더불어 국가지식정보자원으로 관리가 되기도 하지만, 각 기관에서 정의하고 있는 메타데이터 스키마들로 관리가 이루어 지기도 하는데, 이질적인 각각의 서비스로 존재함으로써 각 기관간에 메타데이터들의 상호 운용과 DB통합검색에 어려움이 있다.

MARC는 문헌정보처리 표준화기구로서 각국의 주요 도서관을 중심으로 발전해 오고 있다. 최근의 연구들을 보면 MARC포맷의 문제점과 다양한 웹자원 메타데이터 정보조직의 문제에 대한 해법으로 XML표준포맷의 도입을 시도하였다. 실제 W3C에 의해서 제안된 XML Schema로부터 미국 국회도서관에서는 MARCXML을 Z39.50프로토콜에 적용하여 운용하고 있으며, 우리나라 또한 국립중앙도서관, 국회도서관, KERIS 등에 적용하여 운용하고 있다. 하지만, XML속성 형태의 표현은 단순히 MARC 레코드 표현을 XML 구조로 변환한 것 뿐이며, 서지정보의 메타데이터 요소를 고려하고 있지 않다. 이것은 XML이 정보요소의 의미특성 또는 메타데이터의 표현을 목적으로 하고 있음에 비추어 볼 때, MARCXML은 XML의 근본 취지를 살려내고 있지 못함으로써 정보요소의 의미 표현이 가능한 XML 형식의 MARC레코드 표현에 대한 연구가 요구되고 있다. 시맨틱의 핵심기술로 부각되고 있는 Topic Map은 이미 유용성이 입증된 책 색인 정보를 전자화하기 위한 노력으로 시작되었고, 그 기본적인 기능을 바탕으로 필요한 요소들을 추가하여 ISO 13250 표준을 제정하기에 이르렀다. 이러한 Topic map에서 XML로 표현되는 표준이 바로 XTM(XML Topic Map)이다. XML을 처리하는 표준 API로서 DOM과 SAX 등이 존재한다면 Topic Map을 처리하는 표준 API로 TMAPI 1.0이 있다. XTM은 Topic Map의 각 요소들을 XML요소들로 정의하고 이에 대한 검증은 DTD나 XML스키마를 통해 수행하도록 되어 있다. Topic Map은 토픽과 토픽들 간의 관계를 직접 연결하고 토픽의 인스턴스까지도 다 연결된 상태이기 때문에 토픽을 재배치하면 그 인스턴스가 함께 따라 가게 되어 있고 필요하면 인스턴스 수정도 물론 가능하다. 이러한 기능은 Topic Map을 수정하면 새로운 서비스를 제공하는 결과를 가져오기 때문에, 학술정보자원에 대한 DB 구축 시 XTM을 이용하였다면 이미 개발된 여러 메타데이터 등을 한곳으로 통합 하면

서도 신축성과 확장성을 제공하는 것이 용이하게 된다.

본 논문에서는 서지정보를 포함하고 있는 도서관 정보시스템의 MARC를 메타데이터의 의미를 기계가독형으로 표현할 수 있는 Topic Map을 이용하여 KORMARC 포맷을 재 설계하는 시스템을 실험적으로 구현하였고, 이러한 Topic Map 기반의 변환모듈을 통하여 학술정보자원에 대한 DB구축과 상호운용 및 메타 통합방안에 대한 대안을 제시한다.

II. 상호운용성을 위한 MARC, Dublin Core의 비교

MARC(Machine-Readable Cataloging)는 서지정보 기술을 위해 지속적으로 발전되어 왔으며, 풍부한 서지정보의 요소를 포함하고 있어 대부분의 도서관 정보시스템이 MARC 형식을 지원하고 있다. 이러한 까닭에 서지정보 기술은 MARC의 풍부한 서지정보 요소와 상호 호환을 보장하면서, 의미 표현도 가능한 메타데이터 수준의 방식이 요구되고 있다.

본 논문에서는 MARC의 형식과 서지정보요소를 그대로 보존하면서 개념화하여, MARC 구조체계와 Topic Map과의 매핑을 통한 Topic Map 모델링을 시도하고자 한다. 개념화된 서지정보 요소는 메타데이터 수준에서의 XML기술의 하나인 Topic Map의 XTM으로 표현이 가능하고, MARC 형식과 상호 변환될 수 있으므로 기존의 도서관 정보시스템의 기능을 고도화 하는데 활용할 수 있고, 다른 메타 데이터 체계로의 변환이나 문헌정보 구축 등에도 이용할 수 있다. <그림 1>은 대표적인 MARC 레코드의 예이다.

```
00896nmnn                2200229
45000010016000000050015000160070015000310080041000460410008000870580011000950900
00800106245009000114246012000204280004400324300001400368516001600382538004200398
653009200440773007000532856006600602850001100658NURL_AR0090297420071106170905c
r mn ---auauu071106s2007 ko ze a kor 0 akor a65824 a65800a20
세기 후반 디자인 트렌드의 형성요인과 색채 트렌드 분석h[전자자료]/d김현경(Kim
Hyun Kyung),e 김영인(Kim Young In) 저3 iParallel title:aThe Analysis of the
Influent Factors on Design Trends and Color Trends in the Late 20th Century
a(서술)b한국디자인학회,c2007.02e(서술-f누리미디어) a K:b종합 a전자저널(PDF파
일) a접근방법 : World wide web 및 Adobe Acrobat akeyword트렌드의 형성요인,a
디자인트렌드,a색채트렌드,atrend influential factors,adesign trends,color trends0 t디
자연구,d디자인연구 통권 제69호(Vol.20
No.1)(2007.2).gpp.5-20x1226-804640uhttp://www.dbpia.co.kr/view/ar_view.asp?arid=902
9740 b)2600
```

<그림 1> KORMARC 레코드의 예

<그림 1>에서 알 수 있는 바와 같이, MARC 레코드는 순차레코드 형식으로 레코드의 내용을 이해하기 위해서는 이를 재구성을 하여야 하는 문제점이 있다.

의미 태그를 활용하여 정보자원의 메타데이터를 표현할 수

있는 XML 기술이 보편화됨에 따라, MARC의 정보요소를 XML로 표현하기 위한 많은 노력이 있었다. 미국 국회 도서관은 MARC의 데이터요소를 XML환경에서 사용할 수 있도록 XML 스키마 기반의 MARCXML을 개발하였다. MARCXML의 주 목적은 MARC 레코드 데이터를 XML 형식으로 재현하고, XML 기술을 응용하여 레코드 데이터를 변환하거나 표현하는 것이다. MARCXML 프레임워크의 핵심 요소는 MARC 데이터를 포함하는 XML 스키마이다. 이 스키마는 MARC 데이터를 더블린 코어나 다른 메타데이터 형식으로 변환할 때에 MARC 레코드 처리의 교량 역할을 한다. <그림 2>는 미국 Library of Congress에서 배포하는 MARCXML의 MARC 레코드 데이터 표현의 예이다. MARCXML은 <그림 2>에서 보는 바와 같이, MARC 요소를 XML 속성형태로 표현하고 있다. 이와 같은 XML 속성형태의 표현은 단순히 MARC 레코드 표현을 XML 구조로 변환한 것일 뿐이며, 서지정보의 메타데이터요소를 고려하고 있지는 않다. 이것은 XML이 정보요소의 의미특성 또는 메타데이터의 표현을 목적으로 하고 있음에 비추어 볼 때, MARCXML은 XML의 근본 취지를 살려내고 있지 못함을 알 수 있다. MARCXML은 서지정보 시스템에서 요구하는 정보요소 기술사항을 수용하고 있지 못하기 때문에 실용화하는데 한계가 있으며, 정보요소의 의미 표현이 가능한 XML 형식의 MARC레코드 표현에 대한 연구가 요구되고 있다.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <collection xmlns="http://www.loc.gov/MARC21/slim">
- <record>
- <leader>01142cam 2200301 a 4500</leader>
- <controlfield tag="001">92005291</controlfield>
- <controlfield tag="008">DLC</controlfield>
- <controlfield tag="005">19930521155141.9</controlfield>
- <controlfield tag="008">92021951993 caua j 000 0 eng</controlfield>
- <datafield tag="010" ind1="" ind2="">
- <subfield code="a">92005291</subfield>
- </datafield>
- <datafield tag="020" ind1="" ind2="">
- <subfield code="a">0152038655 :</subfield>
- <subfield code="c">$15.95</subfield>
- </datafield>
- <datafield tag="040" ind1="" ind2="">
- <subfield code="a">DLC</subfield>
- <subfield code="c">DLC</subfield>
- <subfield code="d">DLC</subfield>
- </datafield>
- </record>
- </collection>
```

<그림 2> MARC 데이터의 MARCXML 표현의 예

더블린코어는 MARC를 비롯한 기존의 메타데이터가 네트워크 상의 정보자원을 표현하는데 구조적인 문제로 인하여 많은 비용과 시간이 소요되어 이를 대체할 수 있는 단순한 구조의 형식으로 제안된 것이다. 더블린코어 서지인용정보의 입력 지침을 보면, 15개의 메타 데이터 요소로 이루어져 있으며 MARC와의 대응관계를 보면 <표 1>과 같다.

<표 1> MARC, Dublin Core의 주요 대응 필드

구분	레이블	MARC
콘텐츠 기술요소	표제(Title)	245
	주제(Subject)	653
	설명(Description)	250, 520
	출처(Source)	507
	언어(Language)	041
	관련자료(Relation)	760, 762
	내용범위(Coverage)	856
지적 속성 (Intellectual Property)요소	제작자(Creator)	100
	발행처(Publisher)	260
	기타제작자(Contributer)	700, 710
	이용조건(Right)	260
물리적 기술요소	발행일자(Date)	773
	자료유형(Type)	516
	형식(Format)	300
	식별자(Identifier)	856

III. Topic Map을 통한 지식 표현

토픽 타입을 정의하는 과정에서 일반 데이터 리소스와 토픽 인스턴스, 토픽 타입을 구분하는 것은 매우 중요하다. 오직 토픽으로 구현된 정보자원들 만이 이름, 어커런스, 다른 토픽과의 관계 등의 속성을 가질 수 있다. 이는 일반 데이터 리소스와 토픽 인스턴스를 구분할 수 있는 중요한 기준으로 작용한다. 만약 어떤 정보 자원이 이름이나 어커런스, 관계 등의 속성을 필요로 하지 않는다면 그 정보자원은 단순한 인스턴스로 표현된다. 이러한 개념들은 다음과 같이 예를 들어 설명할 수 있다.

- 저자 : 주저자, 공동저자
- 제목 : 원제목, 부제목
- 키워드 : 논문상에 나열되는 여러 키워드 들

저자, 제목, 키워드처럼 인스턴스 요소들을 가지는 카테고리들을 타입으로 분류하고, 만약 어떤 인스턴스도 가지고 있지 않다면 그것은 타입이 아닌 것으로 간주한다. 즉 개념 또는 개체에 그것에 귀속되는 인스턴스가 존재하지 않는다면, 그것은 타입으로 볼 수 없으며, 타입으로 정의되지 않은 모든

개체들은 인스턴스 항목으로 분류한다. MARC는 해당 주제 영역의 중요 개념 정의, 유의어, 번역된 어휘 등을 풍부하게 포함하고 있다는 점에서, 토픽맵으로 변환 가능한 인스턴스 정보들을 추출해 낼 수 있는 완벽한 자원이다.

논문 서지 사항		KORMARC 태그명	TOPIC MAP 매핑
출판등록	ISBN/ISSN	773	Occurrence
저자 정보	저자명	245	topic
내용 정보	논문 제목	245	topic
	키워드	653	topic
발행사항	발행처	260	topic
	간행물명	773	topic
	발행날짜	773	Occurrence
	권호사항	773	topic
형태	페이지 수	773	Occurrence
URL		856	Occurrence

<표 2> KORMARC에 대한 논문의 서지 사항과 TOPIC MAP 매핑

논문 자체에 대한 정보를 가지고 있는 서지 정보 중심의 논문 Topic Map을 모델링하기 위해서 논문의 서지사항 중에서 앞에서 언급한 Dublin Core의 15개 메타데이터 중심으로 토픽을 선정한다. <표 2>는 논문의 서지정보를 정리한 것이다. 논문은 기본적으로 ISBN 또는 ISSN을 가지고 있다. 논문의 공통적인 서지 정보로는 크게 저자정보, 내용정보, 발행사항, 그리고 형태사항으로 구별할 수 있으며, 저자정보에는 한,영문 저자명과 소속기관명이 있다. 내용정보에는 한글, 영어 제목과 국,영문 초록, 그리고 한,영 키워드, 초록이 해당된다. 다음 발행정보에는 발행처와 발행 날짜, 그리고 권호 사항이 있으며, 형태사항으로 페이지수 또는 수록 페이지가 필요하다. 위 <표 2>에서 제시된 서지사항 모두가 논문과 의미적으로 관계된 토픽들이 될 수 있지만 본 연구에는 의미 정보 기반의 논문 온톨로지 검색 시스템 구축을 목표로 하기 때문에 내용과 관계된 논문의 제목과 키워드, 권호, 간행물 그리고 발행처까지를 Topic Map모델링 범위로 제한한다. 따라서 선정된 토픽들 중 주요 접근 포인트가 되는 제목, 저자, 키워드, 권호, 간행물, 발행처를 중심으로 Topic Map의 관계를 정의한다.

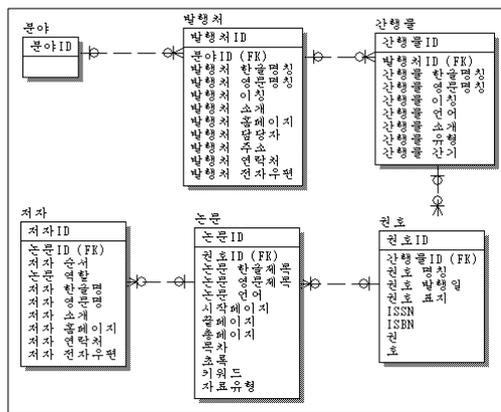
IV. 학술 메타데이터 Topic Map 구축

구축된 학술DB로부터 Topic Map에서 재활용할 수 있는 요소들을 추출하기 위한 정보 소스로서 데이터베이스 스키마,

MARC에서 언급하는 메타데이터, 주제분류로 방향을 잡는다.

일반적으로 어떠한 시스템을 완전히 새로 구축하는 것 보다는 기존에 존재하는 것을 최대한 재활용하여 이용하는 것이 효율적인 방법이다. Topic Map을 작성할 때도 모든 요소들을 새로이 작성하는 것 보다는 작성하려는 분야에 대해 구축된 정보자원이 존재할 경우 이를 최대한 재활용하여, 모든 요소들을 추출한 다음 Topic Map으로 변환하고 이용하는 것이 시간과 비용을 절약할 수 있는 효율적인 방법일 것이다.

일반적으로 학술 콘텐츠 DB를 구축하기 위하여 전자도서관에서는 어떤 파일 형식으로 제작할 것인지를 결정해야 한다. 전문DB는 이미지 형태나 텍스트 형태로 구축할 수 있다. 텍스트 파일을 입수할 수 있는 경우는 전문DB를 텍스트 형태로 혹은 텍스트 형태와 이미지 형태를 동시에 구축하고, 텍스트 파일을 입수할 수 없는 경우 이미지 형태로 구축하는 것이 일반적이다. 세계적으로 PDF 방식으로 전문 DB가 구축되고 있는 경향에 따라 전문DB의 파일형식으로 PDF가 있으며, 뷰어(Viewer)로 사용하고 있다. 그리고 학술 메타데이터의 서지항목들을 관리하기위한 전자도서관 시스템의 기본적이거나 공통적인 스키마 설계 구성은 <그림 3>과 같다.



<그림 3> 전자도서관 시스템의 스키마 설계구성

전문DB를 검색하기 위해서는 색인이 구축되어야 한다. 그리고 색인과정을 거쳐 도서관의 엔진에 의해 검색되고, 결과는 양방향 경로에 의하여 정보를 제공 받을 수 있어야 하는데 이를 위하여 서로 상이한 DBMS간의 검색의 호환을 유지하기 위한 통신규약으로 Z39.50표준 프로토콜을 이용하기도 한다.

기존의 인터넷 검색 서비스는 색인 된 키워드에 기반을 둔 검색이 주종을 이루었다. 사용자는 데이터베이스에 키워드나 여과된 값, 혹은 블리언 연산자를 이용하여 정보 소스에 대한

- [2] The Library of Congress "MARC21 XML Schema", 2007, URL: "<http://www.loc.gov/standards/marcxml/>"
- [3] 이현실, 전양승, 한성국, "MARC의 개념 모델링 연구", 한국도서관정보학회, 한국도서관정보학회지, vol.36, no.3 pp.275-289, 2005.
- [4] 조윤희, 이두영, "XML/KORMARC 시스템 구현에 관한 연구" 한국정보관리학회, 정보관리학회지, vol.18, no.4 pp.217-233, 2001.
- [5] 이정연, 김정민, 최석두, 김이경, "기초학문자료 메타데이터 설계 분석 및 온톨로지 적용방안 연구", 한국문헌정보학회지, 제41권, 제2호, pp.291-316, 2007.
- [6] Ahmed, Kal. "Topic Map Design Patterns For Information Architecture. XML", 2003, URL: "<http://www.techquila.com/tmsinia.html>".
- [7] ISO/IEC. 2002. ISO/IEC 13250: Topic Map. ISO/IEC.
- [8] Steve, "The TAO Topic Maps" 2005, URL: "<http://www.gca.org/papers/xml europe2000/papers/s11-01.html>".
- [9] TopicMaps.org Authoring Group, "XML Topic Maps(XTM) 1.0 Specification", 2001, URL: "<http://www.topicmaps.org/xtm/1.0/index.htm>"
- [10] Bernstein, A., Klein, M. : Towards High-Precision Service Retrieval. Proceedings of The International Semantic Web Conference(2002) 84-101
- [11] Steve P. : The TAO of Topic Maps : Finding the Way in the Age of Information. Proceedings of XML Europe 2000, GCA(2000).
- [12] Garshol, L.M. (2004). Metadata? Thesauri? Taxonomies? Topic maps! Ontopia. Available at: <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>
- [13] Oh et al. (In press).. Ontology-Driven Knowledge Organization Enhancing UDDI Web services in Koreausing Topic Maps, Proceedings of the 68th Annual ASIST meeting. October 28-November2, 2005. Charlotte, North Carolina. USA.
- [14] Pepper, S. 2002. The tao of topic maps. Oslo: Ontopia. Available at: <http://www.ontopia.net/topicmaps/materials/tao.html>