

MARC의 개념 모델링 연구

A Study of the Conceptual Modeling of MARC

이 현 실(Hyun-Sil Lee)*

전 양 승(Yang-Seung Jeon)**

한 성 국(Sung-Kook Han)***

〈 목 차 〉

- | | |
|--------------------------|----------------------|
| I. 서론 | III. MARC 기반의 개념 모델링 |
| II. MARC 기반의 서지 정보 기술 체계 | 1. MARC 개념 모델링의 |
| 1. 서지정보 기술의 표준 : MARC | 요구 사항 분석 |
| 2. MARCXML | 2. MARC 구조의 개념 모델링 |
| 3. MODS | IV. 결 론 |

초 록

본 연구에서는 데이터의 포괄성과 상세성이 있는 MARC를 기반으로 서지정보 개념 모델을 구축하였다. MARC21, MARCXML, MODS를 분석하여 MARC의 개념 모델링에 고려하여야 할 요구 사항을 도출하였으며, 이를 충족할 수 있도록 구축한 개념모델은 MARC 기반에 XML의 태그를 사용할 수 있고, 관계형 데이터베이스 구축에 활용할 수 있는 집합관계의 모델이다. 집합체 구조는 다양한 배열의 형태로 구현할 수 있기 때문에 서지정보 시스템 개발에 유용하게 사용할 수 있을 것이다. MARC가 서지 정보 레코드 구조만을 정의하고 있기 때문에 풍부한 의미의 표현에는 한계가 있으므로 더블린 코어 등 MARC수준의 서지 정보 요소를 확장한 메타데이터 시스템을 도입할 필요가 있다.

주제어: MARC, XML, 서지정보, 개념모델링, 집합관계 모델, 메타데이터 시스템

ABSTRACT

In this paper, the conceptual model for bibliographic information consistent and compatible with MARC is presented. The modeling requirements are derived from MARC-related models such as MARC21, MARCXML and MODS. To meet these requirements, this paper proposes the conceptual model based on MARC formalism. The model composed with aggregation relationships among bibliographic data elements can use semantic tags of XML. As the model can be realized into diverse structures, it will be effectively applied for the development of bibliographic information management systems. Since MARC defines only record format and has the limitations in semantic representation, the metadata system that can expand bibliographic data elements in MARC into metadata level is strongly required.

Key Words: MARC, XML, Bibliographic Information, Concept Modeling, Aggregation Relation, Meta Data System

* 원광대학교 중앙도서관 사서 (hyunsil@wonkwang.ac.kr) (제1저자)

** 원광대학교 컴퓨터공학과 석사과정(globaljeon@wonkwang.ac.kr) (공동저자)

*** 원광대학교 컴퓨터공학과 교수(skhan@wonkwang.ac.kr) (공동저자)

• 접수일: 2005년 8월 18일 • 최종심사일: 2005년 8월 25일 • 최종심사일: 2005년 9월 5일

I. 서론

정보기술의 비약적인 발전으로 인하여 구조화된 정보뿐만 아니라 그래픽, 동영상, 음성 등 다양한 비정형 멀티미디어 데이터의 체계적인 표현이 가능하게 되었다. 이러한 기술을 기반으로 하여 정보자원에 내재한 의미 요소를 기술하기 위한 메타데이터나 온톨로지 등의 개념정보 모델링 기술이 보편화 되고 있다. 컴퓨터가 정보자원의 의미구조나 개념구조를 이해하고 처리함으로써, 시맨틱 웹(Semantic Web) 등에서 목표로 하고 있는 지능화된 지식처리 시스템을 실현할 수 있게 된 것이다.

도서관 정보시스템에서는 자료의 목록과 검색을 위한 서지정보 기술과 서지정보의 상호 공유 및 전달 등의 표준화 방법으로 전통적인 MARC형식을 사용하고 있지만 MARC형식은 컴퓨터 기술의 초창기에 설계된 것으로 기계중심의 단순 레코드 형식으로 다양한 서지정보를 조직적으로 표현하는 데는 한계가 있음이 지적되어 왔다. 서지정보간의 계층적 구조와 같은 다양한 관계표현의 어려움, 구조의 복잡성과 경직성으로 인한 확장성 결여, 디지털 정보자원의 표현 한계와 도서관 이외의 정보환경과의 통합어려움 등 여러 문제가 제기되고 있다.¹⁾ 특히, 서지정보의 의미구조 표현 능력의 결여는 지식기반의 정보시스템 실현에 걸림돌이 되는 것이다.

의미 태그를 활용하여 정보자원의 메타데이터를 표현할 수 있는 XML 기술이 보편화됨에 따라, MARC의 정보요소를 XML로 표현하기 위한 많은 노력이 있었다. 미국 국회 도서관은 MARC21의 데이터요소를 XML환경에서 사용할 수 있도록 XML 스키마 기반의 MARCXML을 개발하였다.²⁾ 또한 정보의 메타데이터 표현을 위한 DCMI(Doublin Core Metadata Initiative), MARC21의 기술적 메타데이터 표준인 MODS(Metadata Object Description Schema), 전거 기술을 위한 MADS(Metadata Authority Description Schema) 등 여러 방법이 제시되고 있다. 최근에는 서지 데이터의 구조적 프레임워크를 정의한 FRBR(Functional Requirements for Bibliographic Records) 모델이 제시되기도 하였다.³⁾ 하지만, MARCXML은 MARC 데이터 요소의 XML 스타일 표현 방식으로 메타데이터 개념이 결여되어 있고, MODS는 일종의 DCMI의 확장으로 MARC의 주요 요소 중 일부만을 수용하고 있어 표현의 완전성이 미흡한 상태이다. MADS는 MODS와 연동하기 위한 전거데이터 기술형식이며, FRBR은 서지정보 기술 방식 아닌 구조적 프레임워크이다.

한편, MARC는 서지정보 기술을 위해 지속적으로 발전되어 왔으며, 풍부한 서지정보의 요소를

1) 이현실, 한성국, "OWL을 이용한 온톨로지 기반의 목록시스템 설계 연구," 정보관리학회지, 제21권 제2호(2004. 6), p.251.

2) 조윤희, XML/KORMARC 통합시스템 구현에 관한 연구(박사학위논문, 중앙대학교 대학원 문헌정보학과, 2001), pp.32-37.

3) 이성숙, FRBR 모형의 수용에 관한 연구(박사학위논문, 연세대학교 대학원 문헌정보학과, 2004), pp.12-14.

포함하고 있어 대부분의 도서관 정보시스템이 MARC형식을 지원하고 있다. 이러한 까닭에 서지 정보 기술은 MARC의 풍부한 서지정보 요소와 상호 호환을 보장하면서, 의미 표현도 가능한 메타 데이터 수준의 방식이 요구되고 있다. 나아가서 서지정보의 온톨로지 체계 구성이 필요하다.

이에 본 연구에서는 MARC의 의미표현을 가능하게 하는 MARC 형식의 개념 모델화 방식을 제시하고자 한다. 제시하는 MARC의 개념모델링 방식은 MARC의 형식과 서지정보요소를 그대로 보존하면서 개념화하여, 서지정보요소의 통합된 조직 체계를 실현하는 방법을 시도한다. 개념화된 서지정보 요소는 메타데이터 수준에서의 서지정보 기술을 가능하게 하며, MARC 형식과 상호 변환될 수 있으므로 기존의 도서관 정보시스템의 기능을 고도화 하는데 활용할 수 있고, 다른 메타 데이터 체계로의 변환이나 문헌정보 온톨로지 구축 등에도 이용할 수 있을 것이다.

본 연구의 진행 절차는 먼저, MARC를 중심으로 한 기존의 서지정보 표현 방식과 XML 기반의 MARC 확장 방식을 비교 분석한다. 그리고 이를 근거로 MARC 모델링의 요구사항 고찰하고, 이를 충족하는 방법으로 MARC를 개념모델을 구축한다.

II. MARC 기반의 서지 정보 기술 체계

MARC는 서지정보 기술의 표준으로 자리하고 있지만, 컴퓨터기술의 급속한 발전에 따라, 새로운 기술 수용이 요구되고 있다. 특히, XML을 기반으로 하는 의미 표현 기술이 보편화되면서, 표현 형식과 내용이 XML 기반으로 전환되고 있다. 본 장에서는 효율적인 서지정보 기술 방식을 정립하기 위하여 MARC와 XML 기반의 MARC 확장을 분석한다

1. 서지정보 기술의 표준 : MARC

MARC는 1960년대 후반 LC에서 서지관련 정보의 표현과 교환을 위한 표준 형식으로 처음 개발되었다. MARC는 원래 자기테이프 상에서 서지데이터 전송형식을 지원하고 인쇄목록의 합리적인 관리를 위한 목적으로 개발되었으나, MARC 형식을 이용한 문헌 정보의 컴퓨터처리 유용성이 입증됨에 따라 서지용 포맷을 비롯하여 전거용 포맷, 소장용 및 분류용 포맷 등으로 도서관 업무에 적합한 여러 형식이 개발되었다. 또한 각국은 도서관 업무의 효율성을 위해 USMARC, CAN/MARC, UKMARC, KOMARC 등 자국의 실정에 적합한 MARC 형식을 발전시켰다. 최근에는 인터넷의 확산과 더불어 네트워크 정보자원에 대한 서지정보 기술의 필요성이 증대됨에 따라 USMARC를 개정한 MARC21 표준안⁴⁾이 제정되었다. MARC21은 단행본을 비롯한 연속간행물, 전자자료, 지

4) Library of Congress, *MARC Standards*, <<http://www.loc.gov/marc/marc.html>> [cited 2005. 7. 10].

도, 악보, 영화, 음성/비디오 레코딩 등의 다양한 형태의 정보자원에 대해 서지정보를 종합적으로 표현할 수 있도록 설계되어, 서지정보 시스템의 표준으로 확고한 위치를 차지하고 있다.

MARC21은 5종류의 계열 표준안으로 구성되어 있지만, 각 형식의 MARC 레코드는 다음과 같이 3가지 주요 요소로 구분된다.

- 리더(leader): 레코드 처리에 필요한 정보를 제공하는 데이터 요소로서 각 자리에 특별한 의미가 부여되어 있는 24문자로 구성된다.
- 디렉토리(directory): 레코드내의 각 가변필드의 태그(tag), 필드길이, 시작위치 등을 표현하는 12자리 정보요소들로 구성되는 가변 필드 목록이다.
- 가변 필드(variable field): 가변 필드는 디렉터리에 나타난 3자리 숫자태그로 시작되는 필드로 구성된다. 가변 필드는 기능에 따라 가변 제어 필드(variable control field)와 가변 데이터 필드(variable data field) 순으로 구성된다. 가변 제어 필드는 00X 태그형식을 갖고며 지시자나 서브필드를 갖지 않으며 단일 데이터요소나 고정길이 데이터요소의 열로 구성된다. 001은 레코드의 제어번호를 표현하며, 003은 제어번호를 갖는 기관을 식별한다.

01X-8XX 형식의 태그를 갖는 가변 데이터 필드는 지시자와 서브필드의 2종류로 구성된다.

- 지시자(indicator): 가변 데이터 필드에 기술하는 처음 두 자리 숫자로써 필드의 부가적인 정보 표현에 사용된다. 모든 가변 필드는 지시자를 가져야 하며 미정의 또는 정보가 없을 경우에는 '#' 기호로 표시한다.
- 서브필드(subfield): '\$'와 알파벳으로 구성된 데이터요소 식별자(data element identifier)로 서브필드 코드와 후행하는 개별 데이터요소로 형성된다. MARC21의 가변 데이터 필드의 정의에 따라, 하나 이상의 서브필드를 가질 수 있다.

MARC21은 서지정보의 정보 구조화 보다는 정보요소의 기술과 저장을 위주로 하여, 일반 자료 구조나 형식언어와 같이 형식적 정의는 없다. MARC의 각 필드요소의 성격에 따라 반복 가능성이 규정되고 있기 때문에 MARC의 형식체계를 정의하는 것은 복잡한 문제이지만, MARC 레코드 형식의 전반적 구성 체계는 <그림 1>과 같이 정의될 수 있다.

MARC 레코드 주요 구성요소 간의 관계를 보여주는 <그림 1>의 MARC 레코드 구조 정의는 MARC 구조를 모델화하고 해석하는데 활용할 수 있을 것이다.

```

<MARC21Record>::<Leader><Directory><VariableField>
<Directory>::<DirectoryElement>*
<DirectoryElement>::<Tag><Length><Position>
<VariableField>::<ControlField><DataField>*
<ControlField>::<ControlNumber><ControlFieldElement>
<DataField>::<Tag><Indicator><SubfieldCode>*
<Indicator>::<FirstIndicator><SecondIndicator>

```

〈그림 1〉 MARC21의 레코드 구조 형식

01041cam 2200265 a 4500001002000000003000400020005001700002400800
4100041010002400082020002500106020004400131040001800175050002400
1930820018002171000032002352450087002672460036003542500012003902
6000370040230000290043950000420046852002200051065000330073065000
1200763` ## #89048230 #/AC/r91~DLC~19911106082810.9~891101s199
0 ## # #maua ## # #j ## # # # # #000 #0 #eng # #~# # \$a # # #89
048230 #/AC/r91` # # \$a0316107514 : \$c \$12.95` # # \$a0316107506
(pbk.) : \$c \$5.95 (\$6.95 Can.)` # # \$aDLC #dLCC #dDLC`00 \$aG
V943.25 \$b.B74 1990`00 \$a796.334/2 \$220`10 \$aBrenner, Richard J.,
\$d1941~10 \$aMake the team.\$pSoccer : \$ba heads up guide to super
soccer! / \$cRichard J. Brenner.`30 \$aHeads up guide to super soccer.
`## # \$alst ed.`## # \$aBoston : \$bLittle, Brown, \$cc1990.`## # \$al27
p. : \$bill. ; \$c19 cm.`## # \$a"A Sports illustrated for kids book."`## #
aInstructions for improving soccer skills. Discusses dribbling, heading,
playmaking, defense, conditioning, mental attitude, how to handle pro
blems with coaches, parents, and other players, and the history of so
ccer.`#0 \$aSoccer \$vJuvenile literature.`#1 \$aSoccer.`\`

〈그림 2〉 MARC 레코드의 예

〈그림 2〉는 대표적인 MARC 레코드의 예이며, 〈그림 3〉은 〈그림 2〉에 저장된 내용에 해당 MARC 필드를 첨가하여 이해하기 쉽도록 재구성한 것이다. 그림들에서 알 수 있는 바와 같이, MARC 레코드는 사용 저장의 효율성을 높이기 위해 구분자(delimitate)를 사용하고 있으나 수정, 갱신 등의 처리효율이 떨어지는 순차레코드 형식으로 레코드의 내용을 이해하기 위해서는 이를 재구성을 하여야 하는 문제점이 있다. MARC 형식은 근본적으로 정보의 모델링을 고려하지 않고 필요에 따라 정보 요소를 추가, 보완하고 있기 때문에 정보요소간의 형식적인 체계성을 기술하고 있지 못하다.⁵⁾ 즉, MARC는 정보 요소의 명확한 정의와 정보 요소간의 의미관계를 표현해 줄 수 있는 정보의 형식 모델로서는 미흡한 것이다. 뿐만 아니라, 〈그림 2〉의 레코드 형식은 발달된 정보 기술의 환경에 비하여 비효율적이며, 정보의 상호공유 및 전달을 지향하는 MARC의 기본 목표와

5) Roy Tennant, "A Bibliographic Metadata Infrastructure for the 21st Century," *Library Hi Tech*, Vol.22, No.2(2004), pp.175-181

도 일치하지 않는다. 이외에도 MARC 레코드 체계는 여러 가지 문제를 내포하고 있어 그 체계로는 지식기반의 문헌관리 시스템이나 지능형 문헌 서비스 등의 개발이 어렵다. 그러나 오랜 기간 동안의 수정과 보완을 통하여 검증 개발된 MARC의 다양한 정보요소를 수용한다면 메타데이터나 온톨로지를 개발하는데 중요한 역할을 할 수 있을 것이다.

Leader	01041cam 2200265 a 4500
Control No.	001 ##89048230
Control No. ID	003 DLC
DTLT	005 19911106082810.9
Fixed Data	008 891101s1990 maua j 001 0 eng
LCCN	010 ## \$a ##89048230
ISBN	020 ## \$a 0316107514 : \$c \$12.95
ISBN	020 ## \$a 0316107506 (pbk.) : \$c \$5.95 (\$6.95 Can.)
Cat. Source	040 ## \$a DLC \$c DLC \$d DLC
LC Call No.	050 00 \$a GV943.25 \$b .B74 1990
Dewey No.	082 00 \$a 796.334/2 \$2 20
ME:Pers Name	100 1# \$a Brenner, Richard J., \$d 1941-
Title	245 10 \$a Make the team. \$p Soccer : \$b a heads up guide to super soccer! / \$c Richard J. Brenner.
Variant Title	246 30 \$a Heads up guide to super soccer
Edition	250 ## \$a 1st ed.
Publication	260 ## \$a Boston : \$b Little, Brown, \$c c1990.
Phys Desc	300 ## \$a 127 p. : \$b ill. ; \$c 19 cm.
Note: General	500 ## \$a "A Sports illustrated for kids book."
Note: Summary	520 ## \$a Instructions for improving soccer skills. Discusses dribbling, heading,

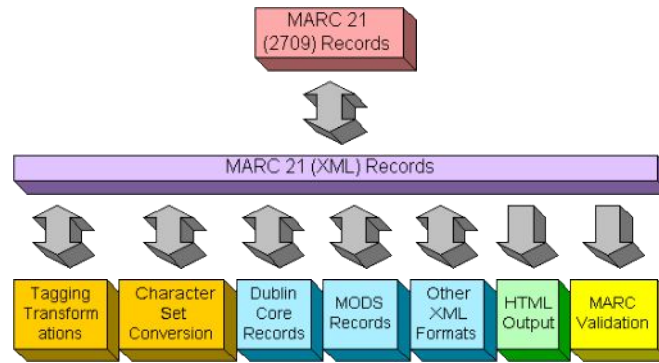
〈그림 3〉 필드 구성의 MARC 레코드

2. MARCXML

정보기술이 XML기반 환경으로 변화됨에 따라, 1990년 중반 LC의 Network Development and MARC Standards Office에서도 MARC 데이터를 XML 환경에서 운용할 수 있는 프레임워크를 개발하였다.⁶⁾ 이 프레임워크는 MARC 데이터를 유연하게 표현하며 필요에 따라 기능을 확장할 수 있도록 개발 되었으며, 스키마(schema), 스타일시트(styleshet), 소프트웨어 툴 등을 포

6) Library of Congress, *MARCXML*, <<http://www.loc.gov/standards/marcxml/>> [cited 2005. 7. 10]

함하고 있다. MARCXML의 전체 구조는 <그림 4>와 같다.



<그림 4> MARCXML의 구조

MARCXML 프레임워크의 핵심요소는 MARC 데이터를 포함하는 XML 스키마이다. 이 스키마는 MARC 데이터를 더블린 코어나 다른 메타데이터 형식으로 변환할 때에 MARC 레코드 처리의 교량 역할을 한다. MARC21에서 변경이 발생하여도 MARCXML 스키마는 수정할 필요가 없도록 MARC의 근본구조를 수용하고 있다. MARCXML 프레임워크에서는 MARC21(2709)과 MARC21(XML) 사이의 무손실 상호변환이 가능하며, 이를 위한 도구가 개발되어 있다.⁷⁾

MARCXML 레코드는 웹 브라우저에서 바로 표현될 수 있고, XML 도구를 이용해서 분석 처리하여 용이하게 응용시스템을 구현할 수 있다.⁸⁾ <그림 5>는 MARCXML의 MARC 레코드 데이터 표현의 예이다. MARCXML의 주 목적은 MARC 레코드 데이터를 XML 형식으로 재현하고, XML 기술을 응용하여 레코드 데이터를 변환하거나 표현하는 것이다.⁹⁾ MARCXML은 <그림 5>에서 보는 바와 같이, MARC 요소를 XML 속성형태로 표현하고 있다. 이와 같은 XML 속성 형태의 표현은 단순히 MARC 레코드 표현을 XML 구조로 변환한 것일 뿐이며, 저지정보의 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 MARC 레코드 표현에 대한 연구가 요구되고 있다.

7) Sally H. McCallum, "Library of Congress Metadata Landscape," *ZifBB*, Vol.50, No.3(2003), p.184

8) Workshop on XML-Based Library Applications, <library.ust.hk/xmlworkshop/2-metadata.pps> [cited 2005. 6. 200]

9) Joaquim Ramos de Carvalho et al, "Meta-Information about MARC : an XML Framework for Validation, Explanation and Help Systems," *Library Hi Tech*, Vol.22, No.2(2004), pp.133-134.

```

<record xmlns="http://www.loc.gov/MARC21/slim">
  <leader>00967cam 2200277 a 4500</leader>
  <controlfield tag="001">3471394</controlfield>
  <controlfield tag="005">19990429094819.1</controlfield>
  <controlfield tag="008">931129s1994 wauab 001 0 eng </controlfield>
  <datafield tag="020" ind1=" " ind2=" ">
    <subfield code="a">0898863872 (acid-free, recycled paper) :</subfield>
    <subfield code="c">$14.95</subfield>
  </datafield>
  <datafield tag="040" ind1=" " ind2=" ">
    <subfield code="a">DLC</subfield>
    <subfield code="c">DLC</subfield>
    <subfield code="d">DLC</subfield>
  </datafield>
  <datafield tag="050" ind1="0" ind2="0">
    <subfield code="a">GV1046.G3</subfield>
    <subfield code="b">G47 1994</subfield>
  </datafield>
  <datafield tag="082" ind1="0" ind2="0">
    <subfield code="a">796.6/4/0943</subfield>
    <subfield code="2">20</subfield>
  </datafield>
  <datafield tag="100" ind1="1" ind2=" ">
    <subfield code="a">Slavinski, Nadine,</subfield>
    <subfield code="d">1968-</subfield>
  </datafield>
  <datafield tag="245" ind1="1" ind2="0">
    <subfield code="a">Germany by bike :</subfield>
    <subfield code="b">20 tours geared for discovery /</subfield>
    <subfield code="c">Nadine Slavinski,</subfield>
  </datafield>
  <datafield tag="260" ind1=" " ind2=" ">
    <subfield code="a">Seattle, Wash. :</subfield>
    <subfield code="b">Mountaineers,</subfield>
    <subfield code="c">c1994.</subfield>
  </datafield>
  <datafield tag="300" ind1=" " ind2=" ">
    <subfield code="a">238 p. :</subfield>
    <subfield code="b">ill., maps :</subfield>
    <subfield code="c">22 cm.</subfield>
  </datafield>
  <datafield tag="440" ind1=" " ind2="0">
    <subfield code="a">By bike</subfield>
  </datafield>
  <datafield tag="500" ind1=" " ind2=" ">
    <subfield code="a">Includes index.</subfield>
  </datafield>
  <datafield tag="650" ind1=" " ind2="0">
    <subfield code="a">Bicycle touring</subfield>
    <subfield code="z">Germany</subfield>
    <subfield code="x">Guidebooks.</subfield>
  </datafield>
</record>

```

〈그림 5〉 MARC 데이터의 MARCXML 표현의 예

3. MODS

MODS¹⁰⁾는 LC에서 2002년 처음 발표하여 현재 3.0버전이 공개되어 있는 정보의 객체기술 방법으로서, MARC21의 서지정보요소를 기반으로 한 메타데이터 형식의 표준안이다. MODS는 MARC의 중요 서지정보요소를 XML 스키마로 정의하고 있으며, 전거 MARC를 위해 MADS를 별도로 제공하고 있다.

MODS는 MARCXML과는 달리 서지정보요소에 대한 메타데이터 모델을 기반으로 설계되었으며, 다음과 같은 특징이 있다.

- XML 기반으로, 서지정보 요소간의 계층구조와 의미를 표현할 수 있고 구조에 확장성과 유연성이 있으며, XML 응용환경을 그대로 활용할 수 있다.¹¹⁾
- MARC의 많은 서지정보에 대한 기능과 의미를 재그룹화 하여 핵심 메타데이터 요소를 추출하였다. MODS에서는 <titleInfo>, <classification> 등 19개의 상위 메타데이터 요소를 정의하고 있으며, 이 요소들은 하위요소 및 속성과 함께 사용된다. MODS는 MARC21의 서지정보와 호환되므로 특수 경우를 제외하고는 상호 무손실 변환이 가능하다.
- 서지정보 메타데이터를 XML 스키마를 이용해서 사용자가 이해하기 쉽도록 의미 태그화하였다.¹²⁾ 예를 들어, MARC 필드260 서브필드b 대신에 <publisher>태그를 사용하고 코드화 된 값 대신에 의미태그를 사용하는 등 MARC의 기호적 성격을 사용자 친화적인 형태로 혁신하였다.
- XML에서 제공하는 xlink나 ID속성을 이용해서 서지정보간의 연관관계를 표현할 수 있다.
- MODS가 전자자원을 고려하여 개발되었기 때문에 자원의 형태에 따른 서지정보 요소들이 많이 포함되어 있다.¹³⁾

<그림 6>이 MODS 문서의 대표적인 형태이다. MODS는 MARC21에 비교할 때 의미 기반의 서지 정보 요소 기술 방식을 제공한다고 할 수 있지만, 근원적인 메타데이터 개념 모형으로는 미흡하다. MODS의 서지정보 요소는 MARC21의 부분집합으로 표현의 완전성이나 호환성에서 있어서도 문제가 제기되고 있다.

10) Library of Congress, MODS, <<http://www.loc.gov/standards/mods/>> [cited 2005. 7. 10].

11) Sally H. McCallum, "An Introduction to the Metadata Object Description Schema(MODS)", *Library Hi Tech*, Vol.22, No.1(2004), pp.82-88

12) Rebecca Guenther, *The Metadata Object Description Schema (MODS)*, NISO Metadata Workshop, 2004, <http://www.niso.org/news/events_workshops/MD-2004_resources/guenther.ppt> [cited 2005. 6. 2]

13) Rebecca S. Guenther, "Using the Metadata Object Description Schema(MODS) for Resource Description : Guidelines and Applications," *Library Hi Tech*, Vol.22, No.1(2004), pp.89-98

```

<mods xmlns="http://www.loc.gov/mods/">
  <titleInfo><title>Germany by bike : 20 tours geared for discovery </title></titleInfo>
  <name type="personal">
    <namePart>Slavinski, Nadine.</namePart>
    <namePart type="date">1968-</namePart>
    <role><roleTerm type="text">creator</roleTerm></role>
  </name>
  <typeOfResource>text</typeOfResource>
  <originInfo>
    <place><placeTerm type="code" authority="marc">wau</placeTerm>
    <place><placeTerm type="text">Seattle, Wash. :</placeTerm></place>
    <publisher>Mountaineers.</publisher>
    <dateIssued>c1994</dateIssued>
    <issuance>monographic</issuance>
  </originInfo>
  <language><languageTerm type="code" authority="iso639-2b">eng</languageTerm> </language>
  <physicalDescription><extent>238 p. : ill., maps : 22 cm.</extent></physicalDescription>
  <note type="statement of responsibility">Nadine Slavinski.</note>
  <note>Includes index.</note>
  <subject authority="lcsht">
    <topic>Bicycle touring</topic>
    <geographic>Germany</geographic>
    <topic>Guidebooks.</topic>
  </subject>
  <classification authority="lcc">GV1046.G3 G47 1994</classification>
  <classification authority="ddc" edition="20">796.6/4/0943</classification>
  <relatedItem type="series">
    <titleInfo><title>By bike</title></titleInfo>
  </relatedItem>
  <identifier type="isbn">0898863872 (acid-free, recycled paper) :</identifier>
  <identifier type="lcn">93047676</identifier>
  <recordInfo>
    <recordContentSource>DLC</recordContentSource>
    <recordCreationDate encoding="marc">931129</recordCreationDate>
    <recordChangeDate encoding="iso8601">19990429094819.1
    </recordChangeDate>
    <recordIdentifier>3471394</recordIdentifier>
  </recordInfo>
</mods>

```

〈그림 6〉 MODS 표현 예

Ⅲ. MARC의 개념 모델링

MARC는 풍부한 서지정보 요소를 가지고 있고 대부분의 도서관 정보 시스템이 이를 지원하고 있는, 도서관 정보 시스템의 표준이다. 본 장에서는 앞에서의 MARC 기반 서지정보 모델링 분석 결과를 토대로 MARC 개념 모델링의 요구사항을 정립하고, 이를 충족시킬 수 있는 MARC 구조

의 개념 모델을 구축한다.

1. MARC 개념 모델링의 요구 사항 분석

MARC는 여러 문제점이 노출되고 있지만, 서지정보 기술의 표준으로 모든 도서관 정보 시스템의 기반이 되고 있다.¹⁴⁾ 이러한 관점에서, MARC기반의 메타데이터나 온톨로지 모델링은 MARC의 기능을 캡슐화하는 래퍼(wrapper)와 최신의 정보 기술과 접목하는 중재자(mediator)의 역할을 하게 한다. 따라서 MARC의 개념 모델링은 서지정보 체계를 발전시키는데 핵심 역할을 수행하게 될 것이다.

MARC는 앞서 고찰한 바와 같이 풍부한 서지 정보 요소를 가지고 있으며, 필요에 따라 서지정보 요소를 추가하면서 발전되어 왔다. 지시자와 서브필드간의 종속성을 제외하고 서지정보 요소는 상호 독립적이고, 명시적인 관계가 존재하지 않는 특징이 있다.¹⁵⁾ 즉, MARC의 서지정보 기술은 선언적 표현(declarative representation)의 성격을 가지고 있다. 이러한 MARC의 특징과 MARCXML, MODS의 MARC 모델을 분석한 결과 MARC 개념 모델링에 고려하여야 할 요구사항을 다음과 같이 도출할 수 있다.

- MARC의 서지정보 요소를 완벽하게 지원하고 호환성이 유지되어야 한다. MARC 개념 모델이 유용성을 갖기 위해서는 MARC 체계를 완전히 수용할 수 있어야 하며, 상호 호환이 가능하여야 한다. MODS에서는 MARC 서지정보 요소의 부분집합만을 표현할 수 있기 때문에, 정보 모델의 완전성이 떨어지고, 상호 호환성에도 한계가 있다.
- 표현과 구조에서 관점의 분리(separation of concerns)가 이루어져야 한다. 개념 모델은 다양한 형태로 표현될 수 있어야 하지만 조직 구조에 일괄성이 유지되어야 한다. 이를 위해서는 XML 기반의 모델이 되어야 한다.
- MARC에 대한 전문지식이 없어도 서지정보를 기술할 수 있어야 한다. MARCXML은 단순히 MARC의 태그를 XML 형식으로 기술하고 있고 때문에 MARC 전문가가 아니면 활용할 수 없는 가용성의 문제가 제기 되었다.¹⁶⁾ MARC의 서지정보 요소를 지식베이스화 할 수 있는 모델이 되어야 한다.

14) Alan Hopkinson, "Traditional Communication Formats : MARC is Far from Dead," *International Cataloguing and Bibliographic Control*, Vol.28, No.1(1999, 1/3), pp.17-21.

15) Peter C. Weinstein, "Ontology-Based Metadata : Transforming the MARC Legacy," *Proceeding of the Third ACM Digital Library Conference* (New York : ACM Press, 1998), pp.254-263.

16) Jian Qin, "Representation and Organization of Information in the Web Space : From MARC to XML," *Information Science*, Vol.3, No.2(2000, 1), pp.83-88.

- 속성 사용을 최소화하고 태그만으로 표현할 수 있어야 한다. MODS는 상위 19개 메타데이터 요소를 정의하고 있으나, 제반 필수 정보는 속성의 형태로 표출되도록 하였다. 이로 인하여 가독성이 저하되고 활용에 어려움이 제기되고 있다.
- 메타데이터 수준의 표현 능력을 가져야 한다. 태그 코드나 기호적인 요소를 배제하고 이해하기 쉬운 의미 태그를 사용하여야 한다.
- 확장 가능하여야 한다. MARC의 확장성을 고려하여 추가 서지정보 요소를 수용할 수 있어야 하고, 서지정보 온톨로지 등과 연계할 수 있는 유연한 확장 능력을 보유하여야 한다. 더불어, 다른 메타데이터 체계와도 연계할 수 능력이 있어야 한다.

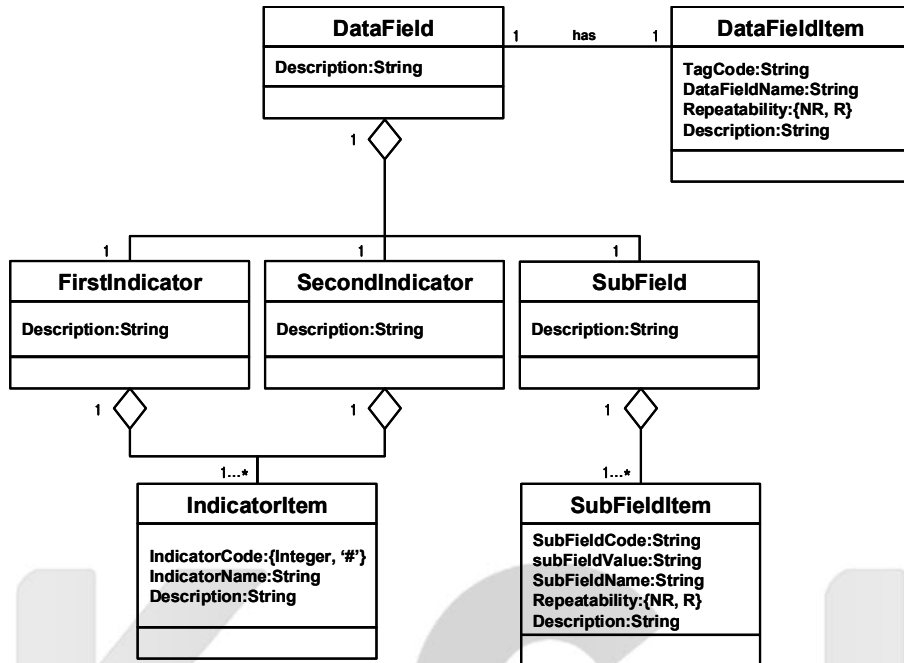
더블린코어와 같은 독자적인 메타데이터 체계 구성이 아닌, MARC 기반의 서지정보 개념 모델은 MARC의 고유 특성을 기반으로 정보 모델링 기법을 접목해야 하며, 사용자와 시스템의 여건을 고려하여 구축된 모델의 유효성도 고려하여야 한다.

2. MARC 구조의 개념 모델링

MARC는 서지정보 요소의 임의 선택에 의한 가변 길이 레코드(Variable-Length Record) 구조를 지원하도록 설계되어 있기 때문에, 정형화된 모델을 설정하는 것이 쉽지 않다. 이러한 이유로 인하여 MARC 기반의 서지정보 시스템은 ER 모델에 기초한 관계형 데이터베이스 구축이 어려웠다. 본 연구에서는 앞에 분석한 요구사항을 기초로 XML의 태그를 사용할 수 있으면서 관계형 데이터베이스 구축에 활용할 수 있는 MARC 구조의 개념 모델을 도출하였다.

MARC 구조를 정의한 <그림 1>을 살펴보면, MARC의 서지정보 요소는 집합(aggregation)의 관점에서 모델화할 수 있다. 예를 들어, <DataField>는 <Tag>, <Indicator>와 <SubField>의 집합 관계로 표현할 수 있다. <그림 1>에서는 이들 서지정보 요소 사이에 순서 관계가 있지만, 실제 구현되는 시스템에서는 순서 관계는 무의미 하므로 집합 관계로 모델화 할 수 있다. 집합 관계를 중심으로 MARC의 서지관계를 모델화하면, <DataField>, <FirstIndicator>, <SecondIndicator>, <SubField>와 같이 집합체를 형성하는 요소와 <Tag>, <FieldCode>, <FieldName>, <Description>와 같이 집합의 원소가 되는 2가지 요소로 구분할 수 있다. 집합체를 형성하는 서지정보 요소는 클래스(Class) 형태의 객체(Object)로 모델화되고, 집합체 원소가 되는 서지 정보요소는 객체의 속성(Property)으로 간주될 수 있다. 이러한 관점에서 가변 필드를 중심으로 모델화한 MARC 구조를 UML(Unified Modeling Language)¹⁷⁾ 형식으로 도식화하면 <그림 7>과 같다

17) UML은 여러가지 다이어그램들을 제시함으로써 소프트웨어 개발과정의 산출물들을 비주얼하게 제공하고, 개발자들과 고객 또는 개발자들 간의 의사소통을 원활하게 할 수 있도록 하고 있다. UML은 시스템을 모델링 할



〈그림 7〉 MARC 구조 모델링

〈그림 7〉에서는 UML 모델링에 적합하도록 MARC 데이터 요소의 명칭을 재구성 하고, 데이터 형(data type)을 지정하였다. 또한 필드의 반복성을 표시하는 〈Repeatability〉와 상세 내용을 기술 하는 〈Description〉을 추가 하였다. 그리고 제1지시자나 제2지시자를 구분하는 〈IndicatorType〉 속성을 도입하였다. 〈FirstIndicator〉와 〈SecondIndicator〉를 하나로 통합하여 보다 간단한 모델의 구축도 가능하지만, 제1지시자와 제2지시자가 필수인 경우의 MARC의 특성을 고려한다면, 별도의 구성요소로 모델링하는 것이 합리적이다. 이 때문에 〈FirstIndicator〉와 〈SecondIndicator〉는 〈IndicatorItem〉의 집합체로 구성하였고, 〈SubField〉는 〈SubFieldItem〉의 집합체로 구성하였다. 〈DataField〉는 〈DataFieldItem〉, 〈FirstIndicator〉, 〈SecondIndicator〉와 〈SubField〉의 집합 구조체로 구성된다.

MARC 형식에서 필드의 반복성은 중요하다. 그러나 필드의 반복성은 필드의 특성에 의존하여 임의로 지정되고 있기 때문에, 이를 위한 정형화된 모델을 구축하는 것은 불가능하다. 필드 반복성이 필드 종속적이므로 본 연구에서는 이를 필드의 개별 속성으로 하였다. 〈DataFieldItem〉과

수 있는 다양한 도구들을 제공하기 때문에, 도메인을 모델링하기가 훨씬 용이할 뿐만 아니라 모델링한 결과를 쉽게 파악할 수 있게 된다. 또한 산업계 표준으로 채택되었기 때문에 UML을 적용한 시스템은 신뢰성 있는 시스템으로 평가받을 수 있다. Terms Korea and whatis.com Inc., 〈<http://www.terms.co.kr>〉 [cited 2005. 8. 9].

〈SubFieldItem〉은 필드 반복성을 표시하는 〈Repeatability〉 속성을 갖도록 하였고, 이에 대한 처리는 응용 시스템에서 이루어지도록 하였다. MARC 형식이 코드화된 숫자와 문자로 이루어져 그 내용을 이해하기가 어렵기 때문에 상세 내용을 추가할 수 있는 〈Description〉 속성을 추가하였다. 〈Description〉 속성은 응용시스템에 따라 다양한 목적으로 활용할 수가 있을 것이다.

〈그림 7〉의 클래스 객체와 속성은 XML 기반 시스템에서 독립된 의미 태그 형태로 사용할 수 있다. XML에서 집합 구조체는 내포된 계층구조 형식으로 표현되므로 보다 체계적인 서지정보의 기술이 가능하다. 집합체 구조는 다양한 배열의 형태로 구현할 수 있기 때문에 서지정보 시스템 개발에 유용하게 사용할 수 있을 것이다.

IV. 결 론

지금까지 MARC는 도서관 정보시스템에서 서지정보 기술의 표준으로 주요 역할을 수행하여 왔다. 하지만 이것은 컴퓨터 기술의 초창기의 단순 레코드 형식으로 설계된 것으로, 발달된 정보기술의 환경에 비하여 효율성에 한계가 있음이 지적되고 있다. 이를 해결하기 위한 여러 방안이 제시되고 있지만 MARC처럼 풍부한 서지 정보 요소를 제공하고 있지는 못하다.

본 연구에서는 데이터의 포괄성과 상세성이 있는 MARC의 개념 모델링이 서지정보 체계를 발전시키는데 핵심 역할을 수행하게 될 것이라는 전제하에 MARC 기반의 서지정보 개념 모델을 구축하였다. 이를 위해 먼저 MARC의 특징과 MARCXML, MODS 등의 MARC 모델을 분석하여 MARC 개념 모델링에 고려하여야 할 요구 사항을 다음과 같이 도출할 수 있었다.

- 첫째, MARC의 서지정보 요소를 완벽하게 지원하고 호환성을 유지할 수 있어야 한다.
- 둘째, 표현과 구조에서 관점의 분리가 이루어질 수 있도록 XML 기반의 모델이 되어야 한다.
- 셋째, MARC에 대한 전문성이 없어도 서지정보를 기술할 수 있도록 MARC의 서지정보 요소를 지식베이스화 할 수 있는 모델이 필요하다.
- 넷째, 속성 사용을 최소화하고 태그만으로 표현할 수 있어야 한다.
- 다섯째, 메타데이터 수준의 표현 능력을 가져야 한다.
- 여섯째, 확장 가능하여야 한다.

이상의 요구사항을 충족할 수 있도록 본 연구에서 도출한 개념모델은 MARC 기반에 XML의 태그를 사용할 수 있고, 관계형 데이터베이스 구축에 활용할 수 있는 집합관계의 모델이다. 이것은 MARC의 서지정보를 집합 관계 중심으로 집합체를 형성하는 요소와 집합의 원소가 되는 2 가지 요소로 구분한 것인데, 집합체를 형성하는 서지정보 요소는 클래스(Class) 형태의 객체(Object)로 모델화되고, 집합체 원소가 되는 서지 정보요소는 객체의 속성(Property)로 간주될 수 있다. 이러

한 관점으로, MARC 데이터 요소의 명칭을 재구성 하고, 데이터형을 지정하였으며, 필드의 반복성을 표시하는 <Repeatability>와 상세 내용을 기술하는 <Description>을 추가하여 UML 형식으로 모델링하였다.

설계된 UML 모델링의 클래스 객체와 속성은 XML 기반 시스템에서 독립된 의미 태그 형태로 사용할 수 있다. XML에서 집합 구조체는 내포된 계층구조 형식으로 표현되므로 보다 체계적인 서지정보의 기술이 가능하다. 집합체 구조는 다양한 배열의 형태로 구현할 수 있기 때문에 서지정보 시스템 개발에 유용하게 사용할 수 있을 것이다.

본 연구에서 설계한 MARC의 개념 모델은 메타데이터 수준의 정보 표현이 가능하고, XML 기술을 활용하여 다양한 형태로 레코드 구조만을 용이하게 변환할 수 있다. 그러나 MARC가 서지정보 레코드 구조만을 정의하고 있기 때문에 풍부한 의미의 표현에는 한계가 있다. 이를 해결하기 위해서는 더블린 코어 등 MARC 수준의 서지 정보 요소를 확장한 메타데이터 시스템을 도입할 필요성이 있음을 제안한다.

<참고문헌은 각주로 대신함>

КСІ