

MARC의 XML 변환 동향

김 수 형*

目次

1. 배경	1) MARC.pm
1) MARC의 현황과 문제점	2) MARC XML
2) XML	3) Medlane XML MARC
2. MARC의 XML 변환 사례	3. 전망

1. 배경

1) MARC의 현황과 문제점

급속히 증가하고 있는 인터넷의 정보자원을 효과적으로 탐색하기 위한 기법의 개발이 진행되면서 새로운 검색모형으로 다양한 분야에서 각종 메타데이터가 제안되고 있다. 가장 오랜 역사를 지닌 대표적인 메타데이터 포맷으로 전통적인 도서관 서지 레코드를 위하여 이미 널리 사용되고 있는 MARC을 들 수 있다. 도서관은 소장 자료의 편목과 분류를 위해 표준화된 시스템을 이용하는데, 온라인 목록을 지원하는 많은 도서관의 협동을 이루는 기반구조가 MARC 레코드이다.

MARC 포맷은 현재 66개 국가에서 사용되고 있다. 그러나 MARC 포맷은 도서관 위주의 목록에 기반한 포맷으로서 도서관 이외의 출판사, 정보센터 등의 관련 기관과의 호환성에 약점을 지니고, 서지 레코드간의 연결기능이 부족하며, 국제적인 서지 레코드의 교환의 경우 별도의 변환프로그램을 필요로 하게 된다.¹⁾ 무엇보다 MARC은 1970년대의 정보처리환경을 토대로 개발되었기 때문에 오늘날의 인터넷 환경을 수용하기 어려운 약점을 지니고 있다.

MARC의 문제점들을 보완할 수 있는 메타언어로 SGML, XML 등을 들 수 있다. 특히 XML은 구성과 문법이 복잡하여 사용자가 습득하기 어려운 SGML을 간소화한 표준이라고 할 수 있으며 사용자가 문서의 요소와 속성, 개체를 선언할 수 있는 유연한 DTD를 지원한다.

* 동명대학교도서관사서(sadrain1@chollian.net)

1) 조윤희, "XML 기반 디지털도서관 구현에 관한 연구", '제7회 한국정보관리학회 학술대회논문집', 2000, pp. 79-82.

```

000 00935cam 22002654a 4500
001 12297319
005 20010327090511.0
008 010202s2001 nyua 001 0 eng
906 $a7$bcbc$corignew$d1$eocip$f20$gy-gencatlg
925 0 $aacquire$b2 shelf copies$xpolicy default
955 $apc20 to sa00 02-02-01; se45/se02 02-07-01; se05 to Dewey 02-08-01;
aa05 02-08-01; CIPver. sb17 3-26-01 ; to BCCD 03-26-01; sb00 03-27-01
010 $a 2001030152
020 $a0071371885
040 $aDLC$cDLC$dDLC
042 $apcc
050 00 $aHD30.37$b.S53 2001
082 00 $a005.7/2$221
100 1 $aSimon, Solomon H.
245 10 $aXML /$cSolomon H. Simon.
260 $aNew York :$bMcGraw-Hill,$c2001.
300 $axxviii, 259 p. :$bill. :$c24 cm.
440 0 $aEmerging business technology series
500 $aIncludes index.
650 0 $aBusiness enterprises$xComputer networks$ySoftware.
650 0 $aXML (Document markup language).
    
```

〈그림 1〉 MARC 레코드의 예제

```

<MARC>
<RECORD>
...
<개인명>Simon, Solomon H.</개인명>
<서명>XML</서명>
<저자명>Solomon H. Simon</저자명>
<발행지>New York</발행지>
<발행사>McGraw-Hill</발행사>
<발행년>2001</발행년>
<면수>xxviii, 259 p.</면수>
<삽도>ill.</삽도>
<크기>c24 cm.</크기>
...
</RECORD>
</MARC>
    
```

〈그림 2〉 그림 1의 예제 MARC을 XML로 표현

일반인들에게는 <그림 1>의 내용이 무슨 의미를 나타내는지 이해가 어려울 것이다. 만약 이것을 <그림 2>와 같이 XML로 구조화한다면 MARC에 대한 개념적 이해가 없는 사람일지라도 이것은 특정한 문서의 서지정보를 표현한 것이며, 서명은 'XML' 이고 저자는 'Solomon H. Simon' 라는 것 등을 쉽게 파악할 수 있다.

2) XML

인터넷 이용의 확산은 네트워크상의 정보자원의 폭발적인 증가를 가져왔으나 이를 효율적으로 이용하는 방법론은 아직 초기단계에 머물고 있다. 인터넷, 특히 WWW의 정보자원의 대부분을 차지하고 있는 HTML(Hyper Text Markup Language)은 애초에 정보 교환용 목적으로 설계된 포맷이 아니기 때문에 구조화된 정보검색과 교환이 어렵고, 다양한 멀티미디어 타입을 제공하지 못하며, 정교한 페이지 형태를 제어하기 어렵다. 그러나 1998년 XML(eXtensible Markup Language) 1.0이 제정되면서 HTML의 단순성과 기존 SGML(Standard Generalized Markup Language)의 복잡함에서 야기되는 문제점들을 동시에 해결할 수 있게 되었다. 즉, XML은 현재 구조화된 포맷의 현실적인 표준이라 할 수 있다.

XML은 웹 상에서 데이터 교환을 위해 제안된 표준언어로서 DTD를 통하여 문서 자체에 문서의 구조를 기술한다. 문서의 구조를 사용자가 원하는 대로 정의할 수 있으므로 이러한 구조의 유동성은 모든 형태의 데이터가 XML로 기술될 수 있도록 한다. 따라서 XML은 인터넷 상의 모든 데이터가 동일한 형태로 통합, 저장, 처리될 수 있는 기반을 제공한다. 요컨대, XML은 구조화된 정보를 교환하는데 용이할 뿐만 아니라 범용적인 시스템으로 작업을 제공하는 효과적인 수단이 된다고 할 수 있다.

XML의 정교한 구문법은 다른 포맷으로의 쉬운 변환을 가능하게 하므로 XML은 과거와 현재의 포맷을 연결하는 중요한 역할을 한다. 또한 XML은 MARC 데이터베이스를 더블린 코어나 GILS와 같은 포맷으로의 변환을 용이하게 할 수 있다.²⁾

2. MARC의 XML 변환 사례

현재 외국에서는 MARC의 문제점을 해결하기 위한 방안으로서 MARC 포맷을 XML 포맷으로 변환하기 위한 연구가 진행되고 있다. 그 대표적인 기관은 아래와 같으며, 대부분은 홈페이지를 통해 MARC의 XML 변환 프로그램과 소스를 외부에 공개하고 있다.

1) MARC.pm (<http://marcpm.sourceforge.net>)

SourceForge에서 개발한 MARC.pm은 MARC 포맷의 서지 레코드의 열람, 조작, 출력, 변환을 위

2) John Robert Gardner, "eXploring What's neXt : XML, Information Sciences, and Markup Technology", http://vedavid.org/xml/docs/eXploring_xmlandlibraries.html

```

<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<marc>
<record>
.....
<field type="100" i1="1" i2="">
<subfield type="a">Simon, Solomon H.</subfield>
</field>
<field type="245" i1="1" i2="0">
<subfield type="a">XML </subfield>
<subfield type="c">Solomon H. Simon.</subfield>
</field>
<field type="260" i1="" i2="">
<subfield type="a">New York :</subfield>
<subfield type="b">McGraw-Hill.</subfield>
<subfield type="c">2001.</subfield>
</field>
<field type="300" i1="" i2="">
<subfield type="a">xxviii, 259 p. :</subfield>
<subfield type="b">ill. :</subfield>
<subfield type="c">24 cm.</subfield>
</field>
.....
</record>
</marc>

```

〈그림 3〉 예제 MARC의 MARC.pm을 통한 XML 변환 예

한 Perl 5 모듈이다. 적용사례로서는 Adat Chaverim Library는 MARC.pm을 이용하여 MARC 레코드의 웹기반 검색 온라인목록을 생성했으며, Johns Hopkins 대학의 Milton S. Eisenhower Library에서는 MARC.pm을 이용하여 MARC 레코드의 049 필드를 제거했다.

2) MARC XML (<http://www.logos.com/marc/marcxml.asp>)

Logos Research Systems에서는 1998년에 DOS용 프로그램인 MARC XML을 개발, 공개하였다. 이것은 어떤 MARC 레코드라도 간단하게 well-formed XML 문서로 생성하고 XML을 정확한 MARC 레코드로 변환할 수 있다.

```

<?xml version="1.0" ?>
<marc>
<record type="cam" info="4a">
<control-field tag="001">12297319</control-field>
<control-field tag="005">20010327090511.0</control-field>
<control-field tag="008">010202s2001 nyua 001 0 eng</control-field>
.....
<data-field tag="100" ind1="1">
<subfield code="a">Simon, Solomon H.</subfield>
</data-field>
<data-field tag="245" ind1="1" ind2="0">
<subfield code="a">XML /</subfield>
<subfield code="c">Solomon H. Simon.</subfield>
</data-field>
<data-field tag="260">
<subfield code="a">New York :</subfield>
<subfield code="b">McGraw-Hill,</subfield>
<subfield code="c">2001.</subfield>
</data-field>
<data-field tag="300">
<subfield code="a">xxviii, 259 p. :</subfield>
<subfield code="b">ill. :</subfield>
<subfield code="c">24 cm.</subfield>
.....
</record>
</marc>

```

〈그림 4〉 예제 MARC의 MARC XML을 통한 XML 변환 예

3) Medlane XMLMARC (<http://xmlmarc.stanford.edu>)

Stanford 대학의 Lane Medical Library에서는 1999년 12월에 서지와 전거 포맷에 대한 DTD를 기반으로 MARC을 XML로 변환하는 JAVA 클라이언트/서버 프로그램인 XMLMARC 1.0을 개발, 배포하였다. Lane Medical Library는 실제로 이 프로그램을 사용하여 25만건 이상의 MARC 레코드를 XML로 변환하였다. 한편, Medlane XMLMARC 홈페이지에서는 도서관의 XML 활용에 관한 정보원 및 관련 연구동향을 소개하고 있다.

4) BiblioML (<http://www.culture.fr/BiblioML/>)

AJLSM의 Martin Sevigny에 의해 개발된 BiblioML과 AuthoritiesML은 국가 서지기관간 데이터 교환용 MARC인 UNIMARC의 서지 및 전거 레코드 교환에 대한 XML 포맷이다. BiblioML과 AuthoritiesML은 UNIMARC 레코드를 변환하여 XML 레코드로 생성한다.

5) Library of Congress (<http://lcweb.loc.gov/marc/marcsgml.html>)

미국의회도서관(Library of Congress)은 일찍이 SGML 포맷으로 USMARC 레코드를 정의하는 MARC DTD를 설계하여 1998년에 MARC을 SGML로 변환할 수 있는 Perl 5 모듈을 개발하였다. 이후 Perl 5를 사용하여 MARC을 XML로 변환할 수 있는 유틸리티를 공개하였다.

3. 전망

MARC을 XML로 변환하는 데 따르는 기대효과로서는 첫째, 서지 레코드의 작성 및 다른 포맷으로의 생성이 가능하고 둘째, 웹브라우저, 검색엔진, 그리고 잠재적으로 다른 변환이 필요없는 도서관 시스템에 의해 서지 레코드를 표현할 수 있으며 셋째, 데이터 손실없이 XML과 MARC 상호간의 변환이 가능하고 넷째, 아시아 문자의 표현, 로마나이즈 및 전거통제 등의 MARC의 많은 문제점들이 해소될 수 있으며 다섯째, 도서관에서 XML이 EDI(Electronic Data Interchange) 표준을 대신할 수 있으며 여섯째, 다른 데이터 소스도 XML구문을 사용함으로써 통합과 처리를 더 쉽게 할 수 있다고 하겠다.

MARC은 자기테이프가 저장매체로 주로 사용되던 시대에 개발된 포맷이어서 현재에 이르러 많은 문제점을 수반하게 되었다. 데이터 교환 및 저장매체 기술의 변화를 거의 반영하지 못하므로 새롭게 요구되는 내용을 표현하는 것이 불가능하기 때문이다. 이미 외국에서는 MARC 레코드 데이터를 차세대 포맷인 XML로 변환하는 작업이 진행되고 있다. 우리나라에서도 KORMARC 레코드를 XML로 변환할 수 있는 프로그램의 개발 및 구현이 요구된다.

3) K.T. Lam, "Moving From MARC to XML", <http://ihome.ust.hk/~lblkt/xml/marc2xml.html>

4) Dick Miller, "XML: libraries' strategic opportunity", <http://www.ljdigital.com/xml.asp>