

# XML기반 디지털도서관 구현에 관한 연구 : XMLMARC시스템 구축을 중심으로

## A study on the implementation of XML-based Digital Library : MARC to XMLMARC

조윤희, 중앙대학교 대학원 문헌정보학과

Yun-Hee Cho, Dept. of Library&Information Science, Graduate School of Chung-Ang University

본 연구는 XML기반 디지털도서관 구현을 위한 선행연구로서, 디지털도서관의 출현과 MARC 포맷의 이용현황에 관한 이론적인 검토와 MARC에서 XMLMARC로의 변환 사례를 검토하였다. 또한 XML기반 디지털도서관 통합서비스를 실현하려면, MARC 데이터의 표준 서지 생성, 웹 자원을 포함한 데이터의 자동생성, 벤더와의 데이터 교환, ILS에서 ILS로의 데이터 교환 환경의 구축이 선행되어야 함을 제안하였다.

### 1. 서론

전세계적으로 인터넷의 연간 증가율은 약 60퍼센트에 달하며, 인터넷을 이용한 전자 정보원의 양적 증가율 또한 60퍼센트를 넘고 있다. 이러한 웹의 양적 증가는 정보 표현에 HTML을 기반한 기술의 약점을 점차로 노출시키고 있다. HTML은 마크업을 이용한 문서의 화면표현과 링크를 통한 문서간의 이동만을 지원하는 단순한 구조로서, 인터넷의 양적 발전은 가셔왔지만, 질적 발전에는 큰 역할을 하지 못하고 있다.

1996년 W3C의 후원으로 새로운 표준을 작성하기 위한 전문가 그룹이 구성되어, 1998년 XML(eXtensible Markup Language) 표준이 완성되었다. 현재 전세계 소프트웨어 리더들인 Sun, Microsoft, DataChannel, Netscape, IBM, SAP, Adobe, SoftwareAG 등에서 채용되어, XML 지원 소프트웨어가 출시되고 있다.

한편 도서관 데이터베이스의 대표적인 레코드 포맷은 MARC로서, 현재 전세계 66개 국가에서 사용되고 있다. 그러나 MARC 포맷은 도서관 위주의 목록에 근거를 둔 레코드 포맷으로 도서관 이외의 출판사, 벤더, 색인초록기관 등 관련업계와의 상호교류나 호환이 되지 못하며, 소규모 서지 기관들이 이용하기에는 너무 복잡하고, 간단한 레코드를 작성하는 도서관에서 다른 서지 기관의 요구에 적합한 정보제공을 위한 멀티 레벨

레코드가 지원되지 않는다.

또한 서지 레코드간이나 서지 레코드내에서의 연결기능이 부족하며, 이를 해결하기 위해 약간의 연결필드와 연결식별기호가 추가되어 있으나, 현재 매우 부족한 실정이다. 현재 전세계 약 60여 국가 서지 기관은 자국의 특성에 맞는 다양한 형태의 MARC포맷을 개발하여 사용함으로써, 자국내 혹은 로컬 포맷으로 사용할 때는 문제가 되지 않으나, 국제적인 서지 레코드 교환의 경우, 별도의 변환프로그램을 필요로 하게 된다.

이러한 MARC의 문제점을 해결하기 위해 전세계적으로 개방된 표준인 XML 기반 XMLMARC 시스템인 디지털도서관 구현은 중요한 과제라고 할 수 있다.

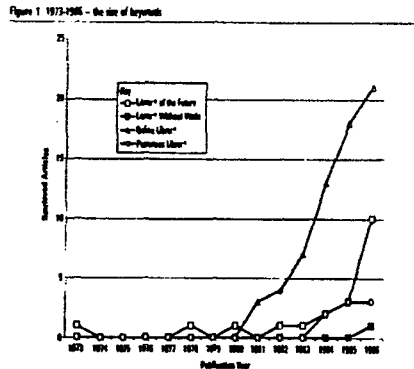
본 연구에서는 XML기반 디지털도서관 구현을 위한 선행연구로서, 디지털도서관의 출현과 MARC 포맷의 이용현황에 관한 이론적인 검토와 MARC에서 XMLMARC으로 변환 사례를 검토하여, XML기반 디지털도서관 구현을 위한 기초자료를 제공하는데 그 목적이 있다.

### 2. 디지털 도서관

디지털도서관이란 화상, 영상, 음향 등을 포함한 멀티미디어 도서의 수집, 축적, 배포를 디지털 신호의 형태로 종합해서 취급하는 도서관을 가리키는 용어로서, 환경의 변화에 따른 도서관 업무

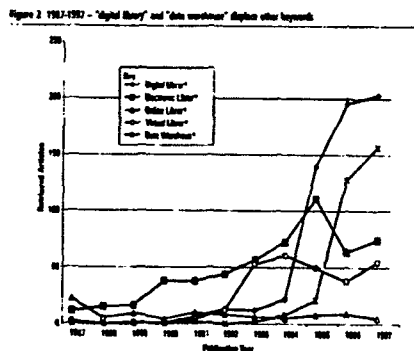
변화를 기술하는 용어로 사용되고 있다.

도서관을 지칭하는 다양한 용어 사용에 대한 비교 연구로서, Library Literature와 INSPEC, Ei Compendex를 대상으로 서지검색을 수행했다. 그 결과, 도서관 명칭사용에 관한 시대별 구분으로 제1기는 '70년부터 '86년까지 주로 사용된 용어는 "Online library"이고, 이것은 '80년 이후 본격적으로 사용되었다. 또한 "library without walls", "paperless library", "library of the future"는 '70년대 말부터 '90년대 초에 출현하였고, 그 현황은 다음 <그림 1>과 같다.



<그림 1> 제1기 : 1973-1986

제2기는 '87년부터 '97년까지로 제1기의 전통적인 용어대신에 "electronic library", "virtual library" 등의 새로운 용어가 출현하였고, 최근에는 "digital library"와 "data warehouse"가 지배적으로 사용되고 있으며, 그 사용현황은 <그림 2>와 같다.



<그림 2> 제2기 : 1987-1997

디지털 도서관에 관한 많은 문헌들은 도서관의 환경 변화에 따른 현상의 출현에 관한 "사실(facts)"이나 도서관과 정보학 발전을 위한 근거를 제공하는 것으로 시도된다. 디지털 도서관의 "의미"는 그것을 개념화하고 반향하는 모든 논술에 있어 뜨겁게 논쟁이 되고 있는 분야이다.

### 3. MARC 포맷의 현황

MARC 포맷은 35년의 역사를 가지고 있으며, 급변하는 컴퓨터와 통신산업의 변화와 견주어 볼 때, 그 지속적인 존립은 그 자체로서 상당히 주목할 만하다. 1966년 LC에 의해 MARC Pilot Project가 시작된 것을 기점으로 영국의 BNB, 프랑스, 독일, 이탈리아, 캐나다, 오스트리아를 포함한 각국의 국가 서지기관에서 Pilot 프로젝트가 MARC를 모델로 수행되었다. 이러한 프로젝트들은 국가 포맷의 개발과 구현을 시작으로, 대부분 다른 국가 포맷과의 호환성 유지에 노력하였다. 그러나 서지 약정과 국가 요구사항 특성의 차이로 인해 상세 사항에서 많은 차이가 있는 포맷을 개발하게 되었다.

이에 국제도서관연맹(IFLA)의 세계서지통정(UBC) 프로그램은 국가 서지 기관간 데이터 교환을 위해 "SuperMARC"으로 불려진 UNIMARC를 개발하였다. UNIMARC은 일반적으로 송신자 국가 포맷으로 인코딩된 후, 이어서 수신자 국가 포맷으로 데이터를 변환하여 전송하는 수단으로 기능하는 인식 포맷이다. 때문에 국가 서지기관은 각국의 국가 포맷에서 UNIMARC으로 변환하거나 UNIMARC에서 국가 포맷으로 변환하는 하나의 변환프로그램을 유지 개발하여야만 한다.

UNIMARC는 10여년 동안 국제간 교환 포맷으로서 사용되고 있으며, 국가 커뮤니케이션 포맷을 위한 모델로서 채택되었다. 그러나 UNIMARC 포맷을 증진하고, 그 적용을 조장하려는 IFLA의 노력에도 불구하고, 첫판 출간이후 10년 동안 UNIMARC으로 서비스를 제공하는 MARC 데이터베이스와 서비스를 제공하는 기관은 단지 5개 기관이었다. '94년 UNIMARC을 채택하고 있는 국가는 12개 기관이며, 3년 이내에 채택할 계획을 가지고 있는 기관은 9개 기관이었다.

'95년 IFLA UBCIM Programme에 의해 수행된 MARC 포맷 사용에 대한 조사 결과를 살펴보면, 전세계 66개 국가 국가도서관이 MARC기반 포맷을 채택하고 있는 것으로 나타났다. 이들 기관의 약 50%는 USMARC 기반으로 하고 있으며,

25%는 UNIMARC기반으로 하고, 나머지 25%는 각 국가 포맷을 기반하거나 기타로 개발된 것으로 보고되었다.

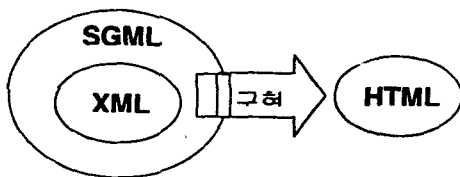
#### 4. XMLMARC

##### 1) XML과 그 응용

XML은 '98년 W3C 권고안으로 채택되었으며, SGML과 같이 메타언어이고, 비교적 간단한 스펙을 가지고 있다. XML의 실제 목적은 SGML에서 많이 쓰이는 핵심부분을 기본으로 인터넷에서 사용하기에 적합하도록 다양한 종류의 응용을 지원하며, SGML과 호환되도록 하였다. 또한 프로그램 작성에 용이하고, 사람이 이해할 수 있는 명확한 구조를 가진 공식적으로 간결하고, 쉽게 설계되었다.

XML의 특징은 사람이 이해하기 쉽고, 기계가 다루기 쉬운 구조를 제공하는 단순성과 W3C 권고의 표준으로 전세계 끝지의 소프트웨어 업체들로부터 지지를 받고 있는 개방성, 고정된 태그 없이, 새로운 태그들을 필요에 따라 사용자가 작성할 수 있는 확장성이다. 또한 태그, 속성, 요소가 같은 XML 구조는 내용의 의미를 해석할 수 있도록 정보를 제공함으로써, 고효율 검색엔진이나 지능형 데이터마이닝 등에 새로운 가능성을 제시해 주고 있다. HTML 태그가 주로 표현을 기술하는데 반해, XML 태그는 내용의 의미를 기술한다. XSL은 XML문서 수정없이 표현을 변경할 수 있으며, 하나의 XML문서의 여러 표현과 여러 XML문서의 동일한 표현이 가능한 내용과 표현이 분리된 언어이다. XML파서는 유니코드를 이해하여 다국어 문서를 지원하며, 이미지, 사운드, 비디오데이터에서 자바 컴포넌트, 액티브엑스 컴포넌트 등 어떠한 데이터의 수용도 가능하다. 현존하는 모든 데이터구조를 대체할 수 있고, 다양한 데이터 포맷을 지원함으로써, 인터넷을 하나의 거대한 XML 데이터베이스로 구축할 수 있다.

SGML, HTML, XML의 관계를 그림으로 표현하면 다음 <그림 3>과 같다.



<그림3> SGML, HTML, XML의 관계

##### 2) MARC에서 XMLMARC로 전환

스텐포드대학 Lane Medical Library의 Dick R. Miller는 도서관의 경험을 저해하기보다는 촉진하는 풍부한 포맷으로 가능하다면 곧 MARC를 교체할 필요가 있다고 언급했다. 그가 말하는 도서관의 경험을 저해하지 않는 형식이란 XML을 의미한다. 또한 Dorman은 XML이 어떻게 MARC를 대체할 것인가에 관하여 더욱 상세한 언급을 하고 있다.

도서관에서 MARC를 메타 언어로 전환하는 변화 징후는 '95년 LC가 USMARC 포맷 인코딩에 SGML 표준 사용에 대한 타당성 조사를 시작한 것에서 기인할 수 있다. 이어 SGML 포맷으로 USMARC 데이터를 정의하는 MARC DTD를 공표하였고, '98년 초 USMARC를 SGML로 변환하는 소프트웨어를 개발하였다.

LC에 의해 개발된 MARC DTD는 주로 MARC와 SGML간의 서지데이터 교환을 위해 설계되었다. 그러나 '98년 W3C에 의해 XML 권고안이 발표되었다. 이것은 SGML를 간소화한 표준으로서, 내용의 표현뿐만 아니라 의미론적 구조를 전달하는 웹 페이지 지원의 메타 마크업 언어이다. 많은 소프트웨어 업체들이 다양한 영역에서 XML을 적용하기 시작했으며, 특히 서지데이터는 그 가능한 영역 중의 하나가 되었다.

이를 종합하면, XML은 기반으로 도서관은 XMLMARC 서지레코드를 생성할 수 있게 되었고, 그것을 다른 포맷으로 변환할 수 있다. 또한 서지 레코드를 웹 브라우저, 검색엔진, 잠재적 변환이 필요 없는 도서관 시스템으로 직접적으로 표현할 수 있다. 서지 레코드는 데이터의 손실없이 XML과 MARC간 교환이 가능하며, MARC 포맷으로 인해 파생된 많은 문제들을 해결할 수 있게 된다.

##### 3) XMLMARC 구현 사례

###### ① Medlane XMLMARC

스텐포드대학의 Lane Medical Library는 '99년 12월 29일 XMLMARC 소프트웨어를 비 상업적 이용을 위해 무료배포를 공지했다. 이 Java client/server 프로그램은 유연한 맵과 단순성, 서지와 전자 포맷에 대한 상세한 DTD를 기반으로 MARC를 XML로 변환하도록 지원한다.

Lane Medical Library는 25만 이상의 MARC 데이터를 XML로 변환하는데 이 프로그램을 사용했다. 현재 XSL 스타일 시트 및 인터페이스를 이용한 디스플레이로 탐색엔진 Oracle 8i를 통해 XML

데이터로의 접근이 가능하다. XMLMARC은 MARC 레코드를 읽어 XML 포맷으로 변환해 주는 프로그램이다. 각 MARC 레코드는 제어번호(001)를 기반으로 화일명과 함께 분리된 XML 파일로 생성된다. 서지와 전거레코드에 대한 DTD와 맵은 모두 Lane Medical Library에 의해 정의되었으나, 필요에 따라 변경이나 교체가 가능하다.

맵은 DTD에 기반한 MARC 데이터 변환에 대해 상당히 유연한 변환 메커니즘을 제공한다. DTD와 맵은 비 Lane 레코드가 성공적으로 변환되기 이전에 교정될 필요가 있다는 것을 보여주며, 맵 자체는 매칭되는 DTD의 이용가능한 MARC 맵을 어떻게 설계하였는가 완전한 재정의의 허용하는 XML 포맷이다.

또한 2000년 4월 중순 Apache의 XML프로젝트 도구와 소스개방, 객체지향 데이터베이스 관리시스템인 Ozone을 이용한 색인에 관한 조사를 결정하고, 레코드색인에 가장 좋은 방법을 연구하고 있다. 유용성에 관해 Infoseek에서 검색엔진 Ultraseek Server 3.1로 실험을 했으며, 실험은 성공적이었다. 색인이 가능했고 데이터의 하위세트 검색이 가능하나 효율적인 "hit list"의 통계 디스플레이를 완전하게 관리하는 방법은 아직 연구되지 못했다.

#### ② 프랑스의 BiblioML

Ministere de la culture et de la communication후원으로 BiblioML 프로젝트는 AJLSM의 Martin Sevigny에 의해 수행되었다. BiblioML과 AuthoritiesML은 UNIMARC 서지 및 전거 레코드 교환에 대한 XML기반 포맷이다. BiblioML과 AuthoritiesML의 레코드는 UNIMARC 데이터베이스를 변환 코딩하여 XML레코드로 생성한다. BiblioML과 AuthoritiesML 변환코딩 운영에 관한 간단한 도구나 그와 관련된 하위세트는 다양한 응용에 의해 사용가능하다.

BiblioML DTD는 국제 표준인 UNIMARC 서지포맷을 기초로 작성되었으며, 모든 문헌 유형의 기술을 처리할 수 있다. BiblioML은 현재 단행본과 연속간행물을 제공하고 있으며, 6000 레코드 이상의 데이터베이스 예와 함께 BiblioML 문서에 대한 UNIMARC 레코드 변환에 응용하는 Java/XSLT 응용을 제공한다. 이 응용은 다운로드가 가능하고, 무료로 사용할 수 있다. 현재 BiblioML 응용은 계속적으로 개발중에 있으며, 미래의 완전한 UNIMARC 지원과 더 좋은 변환 도구, UNIMARC 전거 및 다른 XML 문서내에

간단한 서지 레코드의 마크업을 위해 DTD 프로파일을 제공할 계획에 있다.

## 5. 결론 및 향후과제

XML은 데이터와 문서 교환을 위한 표준 형식으로서 전세계적으로 빠른 호응을 얻고 있다. 현재 도서관은 웹을 포함한 디지털환경의 급변에 대해 상당히 불리한 위치에 놓여 있다. 가장 큰 이유로서, 도서관이 수천년에 걸친 수많은 정보를 보유하고 있음에도 불구하고, 통합도서관시스템(LIS)으로 전환하지 못한 데 있다. MARC 포맷은 복잡한 기계가독형으로 저장된 자료로서, 인터넷 환경의 변화에 따른 적용에 실패하고 있다. 이용자는 웹 환경에서 한번에 모든 검색을 완료할 수 있는 포털서비스를 선호하는데 반해 도서관은 통합시스템으로서의 서비스를 제공하고 있지 못하다. 디지털도서관은 MARC데이터로 표준 서지를 생산하면서, 주제에 관한 웹 자원을 포함한 데이터의 자동 생성, 벤더와 데이터를 교환하고 ILS에서 ILS로 상호 데이터를 교환하는 환경을 구축해야 할 것이다. 이를 위해 본 연구의 향후과제로, XML을 기반으로 하는 디지털도서관 구현을 실현하는 연구를 수행해야 할 것이다.

## 참고문헌

- Cover, Rovin. "BiblioML: XML for UNIMARC bibliographic records". [cited 2000.7.23]. <<http://www.oasis-open.org/cover/biblioML.html>>
- Cover, Rovin. "Mediane XMLMARC experiment : MARC to XML". [cited 2000.7.25]. <<http://www.oasis-open.org/cover/xmlmarc.html>>
- Delsey, Tom. "The evolution of MARC formats". [cited 2000.7.12]. <<http://www.acctbief.org/avenir/evmarc.htm>>
- Lam, K. T. "Moving from MARC to XML". [cited 2000.7.20]. <<http://home.ust.hk/~lbikt/xml/marc2xml.html>>
- Miller, Dick R. "XML: Libraries' strategic opportunity". [cited 2000.7.24]. <<http://www.ljdigital.com/xml.asp>>
- Stanford University. "Status of Medlane Project". [cited 2000.7.25]. <<http://xmlmarc.stanford.edu/Status.html>>
- Watstein, Sarah B., Calarco, Pascal V. and Ghaphery, Fames S. "Digital library : keywords". *Reference Services Review*. Vol.27, No.4(1999.12). pp.344-352.