

海外資料

XML : 도서관의 열린 새로운 가능성

XML: Libraries' Strategic Opportunity*

Dick R. Miller 著**

채지영 譯***

*Copyright 2000, used with permission of Library Journal, a publication of Cahners

Business Information, a division of Reed Elsevier

**스탠포드대 레인의학도서관 사서

***국회도서관 사서

XML은 데이터와 문서교환을 위한 보편적인 포맷으로 빠르게 지지를 얻고 있다. 현재 도서관은 월드와이드웹이 급속히 전개되고 있는 상황에 비해 많이 뒤쳐져 있다. 그러나 도서관들이 웹목록과 웹페이지 작성, 스캐닝 작업등의 노력을 함에도 불구하고, 컴퓨터를 다루지 못하는 사람들이 여전히 있고 이들에게는 그런 개발들이 무용지물이 되는 문제점이 있다. 핵심적인 도서관 데이터는 기술봉사 업무와 전산 사서들의 업무 영역에 주로 제한된 복잡한 MARC 포맷으로 저장된다.

이용자 스스로 해결해야 하는 문제점을 드러내게 된 도서관통합시스템(ILS)은 곧 표준의 필요성을 인식하게 되었다.

XML의 큰 가능성

XML은 자체의 특별한 유연성, 보편성 등으로 전례 없는 가능성을 제공한다. 기업들은 이 특성을 이용해 단일한 자원 검색을 더 좋아하는 고객들에게 매력적인 정보 포털을 준비하기 위해 노력하

고 있다. 그러나 공평성, 신뢰성, 친밀함, 비영리성과 같이 오랫동안 유지해 온 독특한 장점을 가지고 있는 도서관은 이런 장점들을 바탕으로 기업들이 도서관의 가치를 공유하고 있다고 그들의 고객들을 설득시키는 것보다 더 쉽게 XML기술을 통합할 수 있다.

확장가능한 인터넷 표준

사실상 정보내용의 표현을 위한 인터넷 표준이 되고 있는 XML은 1998년 2월, W3C(World Wide Web Consortium)에 의해 제안되었다. 이는, 각각 무제한의 태그를 포함할 수 있는 무제한의 마크업언어를 정의하기 위한 메타언어이며, 따라서 확장가능하다. 이는 1986년 ISO에 의해 인증되고 텍스트기반 문서의 출판에 널리 사용되는 SGML의 부분집합이다.

1996년에 이르러 SGML은 웹 애플리케이션을 다루기에 너무 복잡하고 HTML은 디지털 표리를 다루기에 너무 제한적이라는 사실이 명확해졌다. 사실, 널리 보급되고 있는 브라우저들은 상업적 이익을 위해 표준이 아닌 HTML 태그를 덧붙였고 이는, 웹상에 문서가 표현될 때 브라우저간에 호환되지 않도록 하는 문제점을 가져왔다. 이에, XML은 SGML을 간소화했고, HTML의 대중적 구문을 채택했으며 결국 HTML에 대신할 만한 XHTML을 도입하게 했다.

XML을 미리 코드화된 문서나 레코드들은 논리적인 구조와 물리적인 구조 모두를 가진다. 논리적으로, 요소라고 불리는 계층으로 구성되어 있으며, 이들 요소는 서브필드와 유사한 요소를 내포하는 필드와 링크되어 있다. 각 문서는 종속된 다른 요소를 가진 하나의 루트를 가진다. 요소들은 종료태그를 요구하는데, 이는 HTML과의 중요한 차이점이다. 시작과 종료태그는 한쌍의 산형(山形) 괄호(<>)로 범위가 정해져야 하며, 시작태그와 종료태그에 의해 범위가 한정되는 요소들이 각각의 내부에서 올바르게 종결되어 있을 때 잘 구성된 문서라고 말해진다.

물리적으로, 엔터티는 문서의 구성요소가 명명되고 독립적으로 저장될 수 있도록 하고, 정보를 재사용할 수 있게 하며, 또 이미지와 같이 XML이 아닌 데이터를 참조할 수 있게 한다. 일반적으로, 엔터티는 문서의 첫머리에 선언되며, 문서 안에서 참조된다.

비슷한 이해관계를 가지고 있는 그룹들은 공동의 필요를 충족시키기 위해 지엽적 부가기능을 추가한 DTD(Document Type Definition)를 개발할 수 있다. DTD는 기본적으로 XML관련 문서의 논리적 구조를 위한 템플릿을 형성하여 허용된 엔터티, 요소, 속성 그리고 그들 사이의 관계 각각을 선언한다. 또 데이터의 계층, 허용할 수 있는 속성가치 등을 표현한다. XML문서가 DTD를 따를 때, 일반적으로 유효하다고 말해진다. 이론상으로 XML문서는 DTD를 동반해야 하지만, 브라우저는 DTD없이도 태깅된 XML문서를 올바르게 읽을 수 있다. DTD보다 더 강력한 스키마는 XML문서를 더 효과적으로 유효하게 하도록 정수나 소수같은 복잡한 데이터형식을 가능하게 한다. W3C에 의해 연구중인 몇 개의 스키마 언어가 있고, 올해 그 계획만이 나올 예정이다.

분리와 통합

XML은 실제 분산된 웹상의 문서관리를 다루기 위해 활용된다. 관련기반기술로 XSL(eXtensible Stylesheet Language)과 XLink(XML Linking Language)가 있다. XSL은 W3C에 제안중인 상태로, 웹을 위해 고안된 스타일시트 언어이다. 이는 XML문서를 다른 문서 형태로 변환하거나 워드같은 스타일을 지원하는데 필요한 기술로, CSS(Cascading Style Sheets)보다 막강한 기능을 제공하게 된다. CSS는 간단하게 구조화된 XML문서를 보여주는 정도지만, XSL은 XML의 구조화된 데이터와 문서에 대해 강력한 포매팅 기능을 제공하며, XML의 내용과 표현을 분리한다. 이 내용과 표현의 분리는 XML의 가장 강력한 면 중 하나이다. 같은 데이터도 스타일시트가 다양한 목적으로 정의되는 것에 따라 그만큼 다르게 표시될 수 있다.

다른 관련기반기술로 문서사이의 하이퍼텍스트 링크를 조절하는 XLink가 있다. 이는 기존의 HTML에서의 링크 기능을 확장하여 1:N의 링크 기능까지 확대한 것이다. XML namespace는 XML 문서에서 요소형식과 속성이름으로 쓰일 수 있는 이름들의 집합이다. 이를 통해 정의된 문맥 안에서 이름들은 고유성을 보장받게 되고 이는, 둘 이상의 소스로부터 문서들을 조합해 낼 수 있으며, 여러 소스들이 섞인다고 해도 특성요소나 어떤 소스로부터 왔는지를 식별해 낼 수 있게 해준다.

효율성을 위한 토대

XML은 사서들의 흥미를 끌만한 강력한 힘, 정확성, 제어능력, 유연성을 제공한다. XML은 자체의 확장성, 정교한 하이퍼링크 등으로 장래에 개발을 위한 통합적인 토대로 사용될 수 있다.

XML은 주로 웹출판에서 사용되었지만, XML의 상호운용성때문에 컴퓨팅, 비즈니스(주로 전자상거래), 과학분야에서 광범위하게 쓰이고 있다. 도서관에서 특히 관심있는 것은 XML이 EDI(Electronic Data Interchange) 표준을 대신하고 있다는 것이다. 근본적으로 다른 데이터 소스도 XML구문을 사용함으로써 통합과 처리를 더 쉽게 할 수 있다. 또한, XML사용은 데이터 배포에 있어 30-50퍼센트의 비용절감효과가 있는 것으로 평가됐다.

XML을 지원하는 신상품들이 거의 매일 나오고 있다. 오라클은 5월에 iFS (Internet File System)를 발표했는데, 이는 XML DTD로부터 관계데이터베이스를 자동적으로 생성하고 XML문서를 쉽게 출력하게 한다. 인텔도 5월에 하드웨어단계에서 XML문서처리의 능력을 높이기 위해 고안된 신상품을 발표했다. 마이크로소프트사는 인터넷 익스플로러 5.0에서 XML을 지원했고, 넷스케이프사는 네비게이터를 XML 플랫폼으로 변환하기를 바란다고 발표했다.

SAX(Simple API for XML), W3C에 의해 개발되고 있는 프로그래밍 인터페이스 규격인 DOM(Document Object Model)을 포함하여 무료로 이용할 수 있는 소프트웨어들이 많이 있다. 이들은 문서의 내용, 구조, 형식을 동적으로 업데이트하고 접근할수 있는 프로그램과 스크립트를 가능하

게 한다.

아래는 도서관에서 사용되고 있는 광범위한 XML 애플리케이션과 이를 사용하므로 얻을 수 있는 이점을 설명한다.

XML기반의 도서관통합시스템(ILS)

SGML과 달리 XML은 유니코드를 가진다. 유니코드는 현재 널리 쓰이고 있는 1바이트 문자 대신 2바이트 문자를 지원하므로, 65,000개 이상의 문자를 제공할 수 있다. 또한 XML은 4바이트로 한 문자를 나타내는 ISO 10646 문자인코딩포맷을 채택하면서 20억개 이상의 문자까지 제공할 수 있는 확실한 가능성을 보장받았다. 이 표준은 XML이 문서안에서나 컴퓨터작동시스템과 애플리케이션 안에서나 필독기호, 특수문자, 비로마문자를 보통 텍스트처럼 다룰 수 있게 해준다. 이것은 데이터 네트워킹의 국제화에 있어서뿐 아니라 도서관에도 결성적으로 중요한 점이다.

유니코드의 지원과 종립적 플랫폼이라는 특성으로 XML은 하드웨어, 소프트웨어, 그리고 네트워크 프로토콜로이 계속 변화할 때에도 데이터 수명을 연장시킬 수 있다. Stanford의 Lane Medical Library는 이메일과 더 이상 이용할 수 없는 프로그램으로 작성된 쓸모 없는 포맷의 전자문서를 사용 가능하게 변환하기 위해 XML을 사용하고 있다.

데이터베이스관리는 복잡한 문제이다. XML은 객체로 다루어줄 수 있는 복잡한 데이터구조에 명확한 식별자를 제공한다. 데이터베이스 관리와 결합에 있어 XML사용의 실현가능성을 가장 잘 보여주는 것은 National Library of Medicine(NLM)의 예이다. NLM은 컴퓨터 시스템 현대화 프로젝트의 일부분으로 XML을 MEDLINE서지인용데이터 배포를 위한 포맷으로 채택했다. NLM은 포맷을 변화시키면서 오탈자 결정과 같은 데이터의 조직적 변화 기획에 삼았다. XML은 2001년 MEDLINE의 유일한 배포 포맷이 될 것이다. NLM은 또한, MeSH(Medical Subject Headings)의 XML 버전을 만들 계획이다.

DialogWeb도 지금 XML 데이터베이스 인터페이스를 사용하고, WIPO(World Intellectual Property Organization)도 XML이 문서기탁을 위한 더 좋은 포맷이라고 발표했다.

단순한 계층구조에도 불구하고, XML은 복잡한 서지데이터를 놀랄만큼 잘 다룰수 있다. XML의 우수성에 비추어 사서들은 MARC와 AACR2를 서고할 필요가 있다. MARC를 위한 DTD를 개발하는 것은 어려울 수도 있지만, 1980년대 개발된 MARC를 그대로 둘 수는 없을 것이다. MARC은 21세기의 목록규칙과 상관관계가 있는 포맷으로 변환시키는 것이 가능해 보인다.

1998년 9월이 시작되면서, Lane Medical Library는 Medlane 프로젝트를 착수했다. 이는 다른 웹소스와의 통합을 위해 XML로 목록 레코드를 변환하는 것을 포함한다. Lane은 MARC의 새구조화와 단순화를 위해 DTD를 개발했고 변환가능성을 설명하는 XMLMARC 소프트웨어를 발표했다. 현재, 프로젝트는 색인과 인터페이스개발에 초점이 맞춰져 있다.

원신한 XML기반 도서관통합시스템(ILS)은 3-5년 내에 실현가능하다. ILS 벤더들은 현존 시스템

에 XML을 통합하기 시작하고 있다. ILS는 DTD나 서지, 전거, 소장, 상호대치 등의 ILS 모듈을 구성하는 레코드를 위한 스키마의 개발을 촉진시키기 위해 W3C의 모델을 이용할 수도 있었지만, 이용하지 않았다. 왜 아직도 도서관 시스템은 W3C의 모델을 이용하지 않는가?

XML은 복잡하고 관리할 수 없는 문제들에 대해 훌륭한 해결책을 제공한다. ILS나 MARC이 XML 기반이라면, 우리는 현재 시스템보다 훨씬 훌륭한 시스템의 기능향상과 이종 시스템간의 통합에 더 집중할 수 있었을 것이다. 확장 가능하고 보편적인 포맷으로의 변환은 오래 시간을 들여 개발한 MARC포맷이 퇴화하지 않도록 하고 사서들이 그들의 본연의 임무인 이용자봉사 업무에 좀 더 충실할 수 있도록 할 것이다.