# 공개정보에 대한 XML 적용연구

김현태
한국원자력연구소
*Korea Atomic Energy Research Institute*
*htkim@kaeri.re.kr*

## 1. 서론

원자력통제 분야에서 공개정보 (Open Source Information)는 추가의정서에 의해 중요성이 더욱 강조되고 있다. 최근까지 공개정보 교환에 Microsoft 사의 Word 를 사용하고 있다. Word 가 대다수의 PC 사용자에 의해 사용되고 있으나 Word 파일들에서 검색을 수행하거나 자료를 변경, 변형하는 등의 정보 관리 측면에서는 부족한 점이 있다. 문자에 기반을 둔 메타 언어 (Meta Language)인 XML (eXtensible Markup Language)은 통합의 용이성과 유연성으로 인하여 정보 관리 및 정보 교환의 표준으로 인식되어 차세대 정보기술의 핵심으로 부상하고 있다. 국제원자력기구(IAEA)의 추가의정서 전산체계는 수집된 정보를 XML 로 변환하여 처리하고 있다[1].

본 논문에서는 XML 을 공개정보에 적용하여 Word 파일들을 XML 데이터베이스로 변환 하는 것과 개선된 자료 교환 방안 그리고 기존의 GIS 와의 통합에 대하여 다루었다.

| |
|---|
| Title: (영어)<br>Source Date: (영어)<br>Original Source: (영어)<br>Korean Title: (한글)<br>Keywords: (영어)<br><br>***Content*** (영어, 한글)<br><<>><br><br>…<br><br>Title:<br>Source date:<br>Original Source:<br>Korean Title:<br>Keywords:<br><br>***content***<br><<>>. |

표 1. Word 에 의한 공개정보 교환 형식

## 2. 공개정보교환 형식

현재 공개정보 저장 교환에 사용되는 자료 형식은 표 1 과 같다. 영어 자료인 경우에는 Title, Source Date, Original Source, 빈 줄 (blank line) 그리고 ***content*** 를 사용하고 있다.

한글 자료인 경우에는, Source Date, Original Source 에 추가하여 Korean Title 과 해당 공개 정보의 주요 내용을 기술하는 Keywords, 빈줄 그리고 ***content*** 를 사용하고 있다. Korean Title 에는 인터넷 주소가 추가된다.

사진이나 도표가 있는 경우에는 ***content*** 에 포함시킨다. 사진이나 도표가 몇 개 ***content*** 에 포함될지를 사전에 알 수 없다. ***content*** 에는 작성자의 연락처가 있는 경우가 있으며 내용에 관련된 연락처도 있는 경우가 있다.

"<<>>"는 각각의 자료 끝을 나타낸다. 추가 자료가 있는 경우에는 "<<>>"와 "Title:" 사이에 빈 줄을 넣는다. 추가 자료가 없으면 "<<>>"에 "."을 붙인다. 즉, "<<>>."는 파일의 끝을 나타낸다.

Title:, Source Date:, Original Source:, Korean Title:, Keywords:, 빈 줄(blank line), ***content***, 그리고 <<>> 의 순서는 고정 되어 있다. Korean Title 과 Keywords 는 한글 자료인 경우에만 사용 된다.

Word 사용자가 많기 때문에 이와 같은 방식의 자료 교환이 효과적이기는 하지만 검색에 어려움이 있으며 다른 형태로의 자료 변형이 쉽지 않다. Word 파일들을 데이터베이스로 사용하기에는 XML 데이터베이스에 비해 부족한 점이 많다.

또한 자료형식이 제대로 되어 있는지를 확인하는 작업은 수작업에 의존하고 있다. 즉 ":"과 "<<>>"그리고 "<<>>." 등의 확인을 수작업으로 하고 있다.

이러한 단점을 보완하기 위하여 XML (eXtensible Markup Language) 적용을 고려하게 되었다.

## 3. 공개정보에 XML 적용

XML 은 다른 언어를 기술할 수 있는 언어(메타 언어)이다. 우리에게 익숙한 HTML 과 비슷하게 보이지만 현격한 차이가 있다. HTML 은 정보를 화면에 표현(formatting)하는데 사용된다. 정보를 구성하고 의미를 기술하는 데에는 HTML 이 적합하지 않다. 이에 반해 XML 은 자료의 구조를 정하는데 사용된다. XML 은 사용자가 태그(tag)를 정의하는 확장형이다. 자료구조를 사용자가 정하는 것이다. XML 이 자료 구조를 제공하지만 자료를 설명하고 표현할 수 없기 때문에 다른 기술을 상용해야 한다. XML 자료를 표현(formatting)하는 방법 중의 하나가 HTML 을 사용하는 것이다.

XML 에는 두 가지 유형이 있다. 하나는 "잘 구성된(well-formed)"문서 유형이고, 다른 하나는 "유효한(valid)" 문서 유형이다. 구문론적으로 올바른 (syntactically correct) 문서를 잘 구성된 문서라고 하고 의미론적으로 올바른 (semantically correct)한 문서를 유효한(valid) 문서라고 부른다. 잘 구성된 문서는 XML 문법을 따라야 한다. 유효한 문서는 먼저 잘 구성된 XML 문서이어야 한다. 업무처리상 의미 있는 문서 구조를 미리 XML Schema 등으로 정의하고 이를 잘 구성된

XML 문서에 적용하여 유효성을 확인하여야 한다.

잘 구성된 XML 문서는 MS IE 등으로 확인 할 수 있다. 유효한 XML 은 업무환경에 의미 있는 문서가 되어야 하기 때문에 이를 반영하는 XML Schema 등을 작성하는 데에는 상당한 기간의 적응이 필요하며 업무환경 변화에 따라 바뀔 가능성이 있다.

그림 1 은 표 1 의 문서에 대응하는 XML 문서의 유효성을 확인하는 XML Schema 를 간략하게 하여 Word 2003 에서 읽은 것이다. 그림 1 은 시작태그와 대응하는 마침태그 들을 보여준다[2].

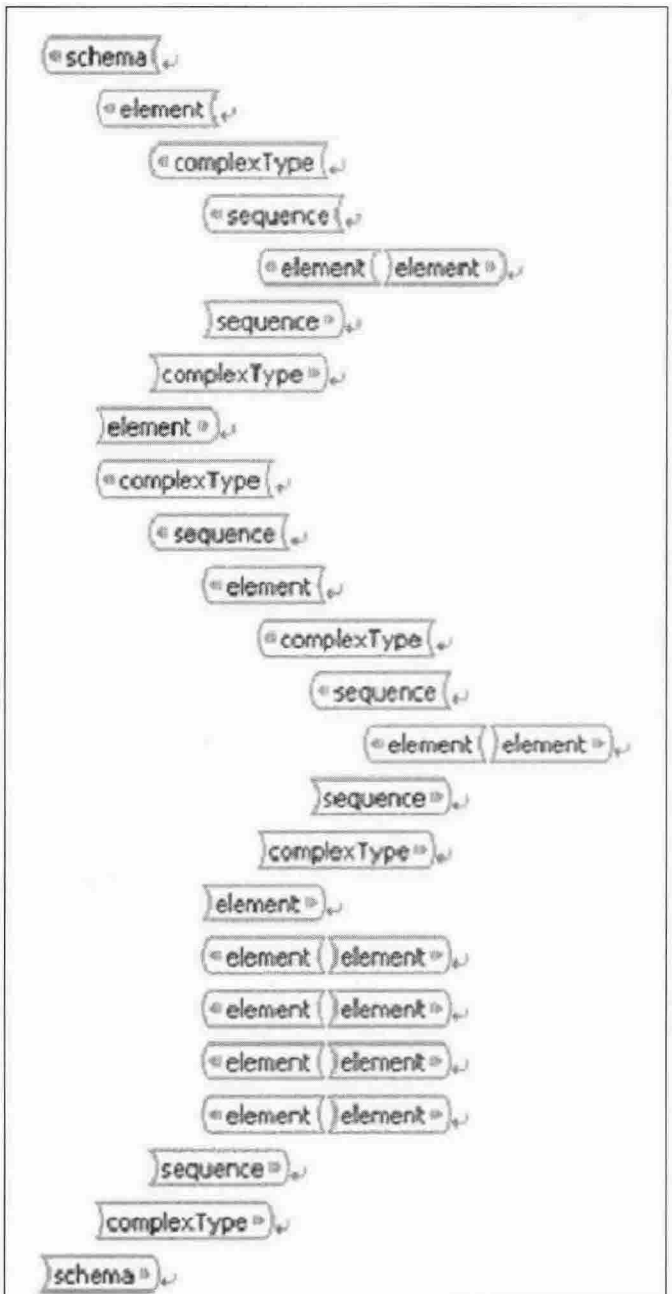


그림 1. 공개정보에 적용된 XML Schema 의 요약

## 4. GIS (Geospatial Information System)와 통합

개발된 GIS 에서는 DAP (Data Access Page)를 UI (User Interface)로 사용하고 있다[3]. DAP 에서 XML 자료를 접속할 수 있기 때문에 XML 로 구성된 공개정보 데이터베이스를 web server 에 저장하거나 XML 자료를 사용중인 MS 사의 Access 에 포함시켜서 Access 기반의 GIS 와 통합할 수 있다[4]. 상기 GIS 개발에 사용된 ESRI 사 ArcGIS 8.1 의 최신판인 Arc GIS 9 은 ESRI 사가 공개한 Geodatabase XML Specification 을 적용하여 XML 을 정보 교환에 사용하고 있다[5]. 특히 shapefile interchange[3]와 유사하게 simple feature interchange 에 XML 을 사용하고 있다. 지리공간정보체계(GIS)와 공개정보관리체계를 XML 기반으로 구축하여 통합할 필요가 있다.

## 5. 결론

정보 체계가 XML 을 적극 수용하는 방향으로 발전하고 있다. IAEA 추가의정서 전산체계도 XML 을 사용하고 있다. Word 를 공개정보 교환에 사용하기에는 부족한 면이 많기 때문에 공개정보를 XML 로 구성하여 정보 교환에 사용하고 XML 기반의 데이터베이스를 구축하는 것이 필요하다. 대부분의 정보체계가 지향하고 있는 GIS 도 XML 를 수용하고 있으므로 GIS 를 XML 기반으로 전환하기 위한 준비가 필요가 있다. 기존의 지리공간정보체계에서 UI 로 사용하고 있는 DAP 에서 XML 데이터베이스에 접속할 수 있으므로 Access 기반의 GIS 와 XML 데이터베이스를 통합할 필요가 있다.

## 참고자료

[1] Elsa Augustenborg, THE IAEA ADDITONAL PROTOCOL INFORMATION SYSTEM DESIGNED FOR EVOLVING NEEDS, 44TH Annual Meeting of Institute for Nuclear Materials Management, J.W. Marriott Desert Ridge Resort & Spa, Phoenix, Arizona, USA, July 13~17, 2003.

[2] Evan Lenz, et al., Chapter 2 The WordprocessingML Vocabulary in *Office 2003 XML*, O'REILLY, http://www.oreilly.com/catalog/officexml/chapter/ch02.pdf

[3] Hyun-Tae KIM, A Study of Geographic Database Application integrated with Nuclear Site Image Information Database, 45th Annual Meeting of Institute for Nuclear Materials Management, Renaissance Orlando Resort at Sea World, Orlando, Florida, USA, July 18~22, 2004.

[4] Frank C. Rice, Exploring XML and Access 2002, Microsoft Corporation, 2001, July (MSDN Home > MSDN Library > Office Solutions Development > Microsoft Access > Technical Articles).

[5] Environmental Systems Research Institute, Inc., Open Geodatabase Interchange Using XML, ArcNews, Spring 2004 (Vol. 26 No. 1) p.VI.