대용량 XML 문서 처리를 위한 기본도구

XML을 여러 응용에서 활용하기 위해서는 문서 작성을 도와주는 도구 및 관련 S/W들이 필요하다. 여기서는 XML 문서 처리의 가장 기본이 되는 XML 파서와 DTD 편집기, XML 편집기, XSL 편집기 등의 저작 도구를 설명하고, 일반적인 XML 문서를 위한 브라우저와 수학, 화학, 음악 등 특수 분야의 브라우저에 대해 소개한다.

■ 채진석/ 인천대학교 전자계산학과 교수

XML 파서 제품

XML 파서는 검증용(validating) 파서와 비검증용(non-validating) 파서의 두가지 종류가 있다. 일반적으로 검증용 파서는 DTD가 있는 유효한 문서를 파싱하는 데 사용되고, 비검증용 파서는 DTD가 없이 사용된 문법에 맞는 문서를 파싱하는 데 사용되는데, 검증용 파서는 비검증용 파서에 비해 구현하기가 어렵다. 〈표 1〉은 지금까지 개발된 주요 비검증용 파서 제품을 보여주고 있다.

〈표 2〉는 지금까지 개발된 주요 검증용 파서를 보여주고 있다.

〈표 1〉 비검증용 파서 제품들

파서명	제작사/제작자	언어	URL
Aelfred	Microstar	Java	http://www.microstar.com/aelfred.html
Expat	James Clark	C	http://www.jclark.com/xml/expat.html
XP	James Clark	Java	http://www.jclark.com/xml/xp
Lark	Textuality	Java	http://www.textuality.com/Lark
TclXML	Zveno/Steve Ball	Tcl	http://www.zveno.com/zm.cgi/in-tclxml
Xparse	Jeremie	JavaScript	http://www.jeremie.com/Dev/XML
xmlproc	Lars Marius Garshol	Python	http://www.stud.ifi.uio.no/~larsga/download/python/xml/xmlproc.html

〈표 2〉 검증용 파서들

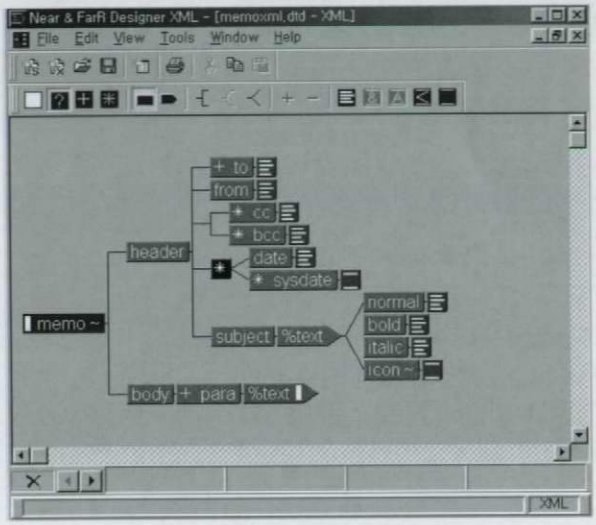
파서명	제작사/제작자	언어	URL
MSXML	Microsoft	Java	Java Virtual Machine으로 통합
XJParser	DataChannel	C	http://www.datachannel.com/products/xjparser.html
Larval	Textuality	Java	http://www.textuality.com/Lark

〈표 2〉의 검증용 파서 중에서 마이크로소프트의 MSXML은 익스플로러 4.0에 탑재되었고, DataChannel에서 MSXML을 확장한 XJParser는 익스플로러 5.0에 탑재된 것으로 알려져 있다.

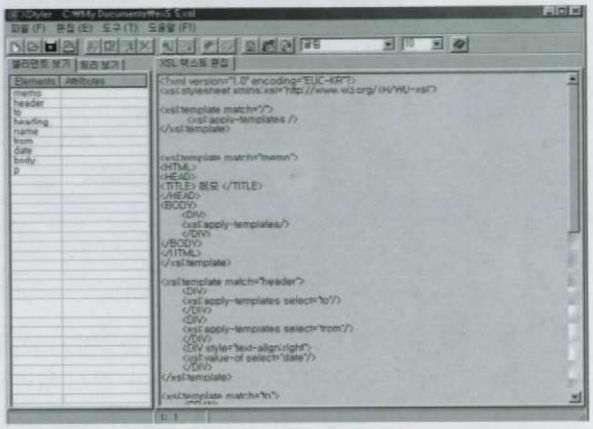
DTD 편집기 제품

DTD 편집기 중 대표적인 것으로는 Microstar Software (http://www.mircrostar.com)에서 개발한 Near & Far Designer를 들 수 있다. 이 제품은 SGML과 XML을 위한 DTD 개발 도구로서, 초보자들이라도 DTD를 개발할 수 있도록 그래픽을 사용하여 시각적으로 XML 요소 트리를 보여주고 있다.

그러나 SGML과 XML 문서를 위한 DTD를 함께 개발할 수 있는 Near & Far Designer는 749달러로 가격이 비싸서 개인적으로 구입하여 사용하기는 무리가 있었으나, 최근에는 199달러의 XML 전용 Near & Far Designer가 출시되어 저렴한 가격에 사용할 수 있게 되었다. 〈그림 1〉은 Near & Far Designer로 메모를 위한 DTD를 작성한 화면을 보여주고 있다.



〈그림 1〉 Near & Far Designer를 사용한 DTD 작성 예



〈그림 2〉 XStyler의 화면

XML 편집기 제품

현재 출시되어 사용할 수 있는 주요 XML 편집기는 〈표 3〉과 같다.

〈표 4〉는 〈표 3〉에서 소개한 XML 편집기의 특성을 보인다.

〈표 3〉 XML 편집기 제품들

제품명	제작사	URL
XML Notepad	미국 마이크로소프트	http://msdn.microsoft.com/xml/notepad/intro.asp
XML Pro	미국 Vervet Logic	http://www.vervet.com
CLIP! XML Editor	한국 Techno 2000 Project	http://www.t2000.co.kr
XEditor	한국 K. O. Tech	http://www.kot.co.kr

〈표 4〉 XML 편집기의 특성

구분	XML Notepad	XML Pro	CLIP!	XEditor
유효한 문서 작성	△	○	○	○
문법에 맞는 문서 작성	○	○	○	○
한글 지원	×	×	△	○
속성 값 작성 및 편집 지원	×	×	×	○
특수 문자 표현	×	×	×	○

XML Notepad는 공개 소프트웨어로 현재 XML Notepad Beta 1.5를 사용할 수 있다. 〈표 4〉에서 보는 바와 같이 XML Notepad에서는 한글이 전혀 지원되지 않고 있으며 주로 DTD가 없는 문법에 맞는 문서 작성용 도구이다.

XML Pro는 XML 문서 작성용 도구로서 한글이 지원되지 않으며, 가격은 약 150달러 정도이다. XML Pro와 Near & Far Designer를 함께 구입하면 299달러에 구입할 수 있다.

CLIP! XML Editor는 유효한 XML 문서를 사용자 친화적인 작업 환경에서 쉽고 빠르게 작업할 수 있도록 도와주는 도구로서 현재 버전 1.5가 출시되어 있다. 그러나 버전 1.5에서는 영문 XML 문서는 아무런 문제없이 작성할 수 있지만, 한글 XML 문서의 경우 화면에는 표시되지만 저장은 되지 않는 문제가 있다. 가격은 약 100달러 정도이다.

XEditor는 CLIP! XML Editor와 비슷한 사용자 인터페이스를 제공하고 있으며, CLIP! XML Editor의 한글 문서가 올바로 저장되지 않는 문제점을 완벽하게 해결하였다. 현재 평가판이 출시되어 있으며, 회사 홈페이지에서 내려 받아 사용해 볼 수 있다. 조만간 상용 제품도 출시될 것으로 기대된다.

XSL 편집기 제품

XML 문서를 위한 스타일시트 편집기로는 XEditor를 만든 K. O. Tech의 XStyler가 있다. XStyler에서는 새로운 XSL 문서를 만들 수도 있고, 기존의 XSL 문서를 열 수도 있다. 〈그림 2〉는 XStyler에서 XSL 규칙을 입력한 예를 보여주고 있다.

〈그림 2〉에서 보는 바와 같이 XStyler에서는 텍스트 형태로 XSL 규칙을 입력하여야 한다. 따라서 XStyler에서는 이러한 XSL 규칙 작성을 도와주기 위해 XSL 문서가 보이는 창에서 마우스 오른쪽 단추를 클릭하면, 현재 사용 가능한 XSL 태그와 HTML 태그를 보여주고 사용할 수 있도록 하고 있다.

또한 왼쪽 편에 엘리먼트 보기와 트리 보기를 제공해 주어

서 XSL 규칙을 입력할 때 도움을 받을 수 있다. 또한 XML 문서를 위한 XSL 문서를 완성하고 나면 '링크' 단추를 눌러 XSL 문서와 XML 문서를 연결시킬 수 있으며, '원문 보기' 단추를 눌렀을 때 나타나는 XML 문서가 익스플로러 5.0에서 어떻게 보일 것인지도 미리 볼 수도 있다.

XML 브라우저 제품

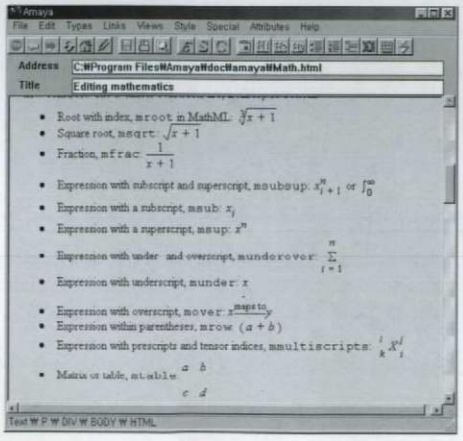
〈표 5〉는 현재 사용할 수 있는 대표적인 XML 브라우저를 보여주고 있다

〈표 5〉 XML 브라우저 제품들

제품명	제작사/제작자	용도
Explorer 5.0	Microsoft	일반적인 XML 문서 브라우징
Amaya	W3C	MathML 문서 편집 및 브라우징
JUMBO	Peter Murray-Rust	CML 문서 편집 및 브라우징
MusicML	Connection Factory	MusicML 문서 브라우징

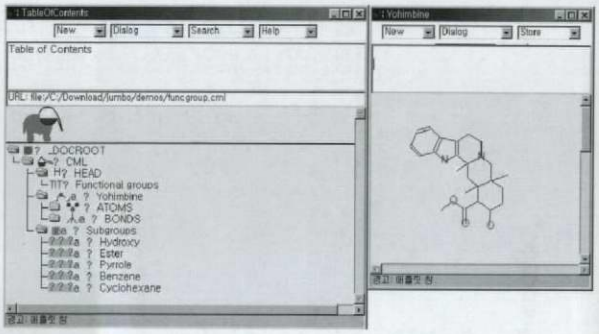
마이크로소프트의 익스플로러 5.0은 최초의 일반 XML 브라우저라고 할 수 있다. 익스플로러 5.0이 출시된 후 비로소 사람들은 XML의 실체와 가까이 만나게 되었으며, 일반적인 홈페이지가 HTML이 아닌 XML로 작성될 수 있는 계기가 마련되었다. 아직 넷스케이프 Communicator에서는 XML이 지원되지 않고 있지만 넷스케이프에서도 XML을 지원하는 브라우저가 개발되기를 기대한다.

Amaya(http://www.w3.org/Amaya)는 W3C에서 개발된 수학식을 입력할 수 있는 마크업 언어인 MathML (Mathematical Markup Language)을 편집하고 브라우징할 수 있는 브라우저이다. 〈그림 3〉은 수학식이 나타난 Amaya의 화면을 보여주고 있다.



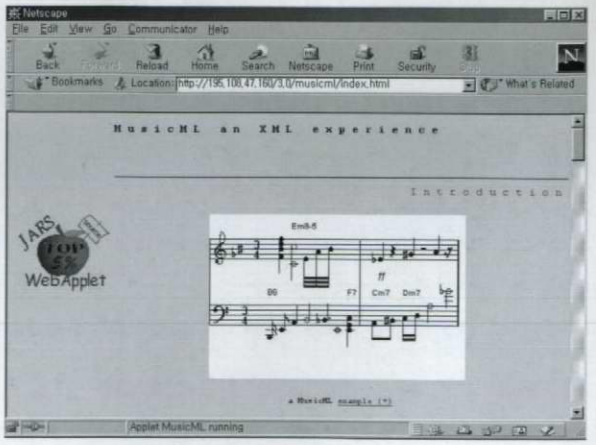
〈그림 3〉 수학식이 나타난 Amaya의 화면

JUMBO(http://www. vsms.nottingham.ac.uk/vsms /java/jumbo)는 University of Nottingham의 Virtual School of Molecular Sciences의 Director로 있는 Peter Murray-Rust가 개발한 화학식을 편집하고 브라우징할 수 있는 CML(Chemical Markup Language) 전용 브라우저로 Java로 구현되었다. 〈그림 4〉는 JUMBO를 사용하여 Yohimbine의 분자 구조를 브라우징하는 예를 보여주고 있다.



〈그림 4〉 Yohimbine의 분자 구조를 보여 주는 JUMBO의 화면

MusicML(http://195.108.47.160/3.0/musicml/index.html)은 네덜란드의 Connection Factory에서 개발한 음악 데이터 기술 언어이다. MusicML(Music Markup Language)로 기술된 악보를 넷스케이프 등의 일반적인 웹 브라우저에 브라우징하기 위해 MSXML을 파서로 사용하여 Java applet 형태의 MusicML 브라우저를 구현하였다. 〈그림 5〉는 MusicML 문서를 넷스케이프에서 실행되는 Java applet을 통해 본 결과 화면을 보여주고 있다.



〈그림 5〉 MusicML을 넷스케이프에서 실행되는 Java applet으로 본 결과 화면