What Is... XML?

Brian Kelly explains XML. This article appears in the Web, and not the print version of Ariadne.

About XML

What is XML?

XML stands for the Extensible Markup Language. XML has been designed to address a number of deficiencies in HTML.

Which deficiencies in particular?

HTML is not extensible. Submitting proposals for extensions to HTML can be a very lengthy process. Browser software vendors can short-circuit the standardisation process by introducing their own extensions, but this has caused problems, as we have seen with controversial extensions such as Netscape's `<BLINK>` and Microsoft's `<MARQUEE>` elements. In addition the browser vendors have shown little interest in supporting specific communities such as the mathematical and scientific communities who would like to make mathematical formulae, chemical symbols, etc. available on the web, without having to resort to the use of images.

How does XML help?

XML is designed to be *extensible*. If you wish to design an office application you can markup a memo thus:

```
<MEMO>
<TO>John Smith</TO>
<FROM>Jane Brown</FROM>
<GREETING>Hello John</GREETING>
<CONTENT>Thanks for the information about XML.
It was very useful.</CONTENT>
</MEMO>
```

Notice that the *structural* elements of the document are described, not the appearance. This will enable the information to be used in a variety of ways. For example a collection of memos could be viewed using the `TO` or `FROM` fields (as you can do in many email programs), structured searching can be carried out or a browser for the visually impaired could communicate the information in a meaningful way.

But how is the appearance of the memo defined?

Using XSL, the XML Style Sheet Language. Just as HTML can be used to define a (very limited) set of structural objects (paragraphs, headings, etc.) leaving CSS to describe the appearance of these objects, XML defines the structure leaving XSL to describe the appearance.

Browser Support

Do browsers support XML?

Microsoft's Internet Explorer 4.0 provides partial support for XML. It supports an XML application called CDF, the Channel Definition Format which is used to "push" information to users. Microsoft have said that Internet Explorer version 5 will provide more complete support for XML. Netscape have also announced support for XML in version 5 of their browser, as illustrated in Figure 1.

The example shown in Figure 1 is taken from <URL: http://www.mintert.com/xml/mozilla/>.

The source code of the document is illustrated below.

Figure 1 Rendering an XML Document in Mozilla

```
...

<album>
<artist>Jackson Browne</artist>
<title>Running on empty</title>
<tracklist>
   <track>Running on empty</track>
   <track>The Road</track>
    ...
   <track>Stay</track>

</tracklist>
<label>Asylum Records</label>
<year>1977</year>
</album>
```

The definition of the layout is given is a linked style sheet file.

In this example we can envisage applications which search for the artist, title or tracks, count the number of tracks, etc. In an XML-aware browser it could be possible to click on a record title to display the tracks. All of these applications would, of course, be very difficult to implement in HTML.

But I'll have to wait until such browsers become widely used before providing my information in XML format?

Not necessarily. Figure 2 shows a screen image of an XML document which has been rendered in Internet Explorer.

In this document, which is based on one available at **<URL: http://www.hypermedic.com/style/tips/tipindex.htm>** the resource contains an XML document which is illustrated below:

```
<document>
<to>To: John Smith</to>
<from>From: Jane Brown</from>
<greeting>Hello John</greeting>Thank you
for your message.
<insult>Your suggestion is
<emphasis>very<extraem>
very</extraem></emphasis> useful!</insult>
Let's talk.
<!--a comment-->
</document>
```


Figure 2 Rendering an XML Document in A Web Browser

A Javascript program converts the XML document to HTML with a style sheet on the fly. (Note that unfortunately this example only works with Internet Explorer because Netscape supports a non-standard document object model).

This example converts the XML document to HTML at the client, using JavaScript. Another approach is to use Java to render XML elements. We are also likely to see XML to HTML conversion happening at the server or using intermediate proxy gateways.

XML Applications

You've described how structured documents can be stored in XML format. What other applications are available?

As an example the Math(s) Markup Language (MathML) [1] is an XML application which became a W3C Recommendation recently. Although most browsers do not yet support MML, a number of Java and ActiveX applications have been developed which can display MML documents, as can be seen in Figure 3.
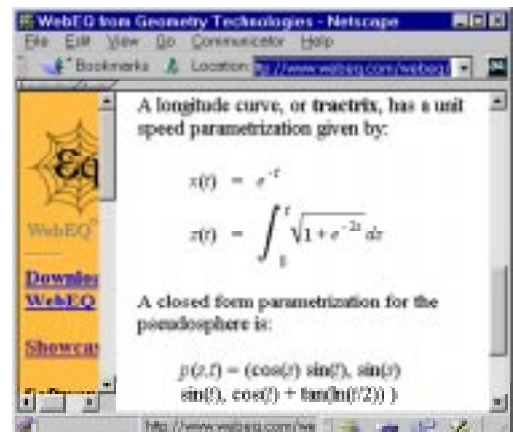

Figure 3 Use of MML to Represent A Mathematical Formula

The use of Java and ActiveX to render XML applications is another way of deploying XML with the current generation of browsers. A paper [2] at the WWW 7 conference suggested Java applets known as *displets* as a way of rendering XML documents. An example is illustrated in Figure 4.

You can download the Java application yourself, together with some example applications from <URL: http://www.cs.unibo.it/~fabio/displet/>.
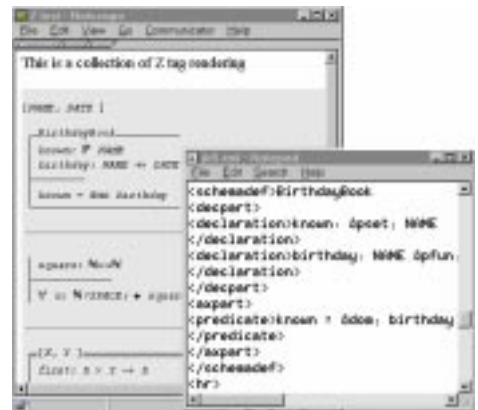

Figure 4 Using Java applets known as *Displets* **to render XML elements**

What other applications are available?

A number of XML applications have already been developed including:

CDF (Channel Definition Format) [3]
> The Channel Definition Format is an XML application which has been developed by Microsoft and submitted to the W3C.

CML (Chemical Markup Language) [4]
> The Chemical Markup Language, developed by Peter Murray-Rust, Nottingham University.

PGML (Precision Graphics Markup Language) [5]
> A proposed 2D imaging model in XML of the PostScript language and the Portable Document Format (PDF).

RDF (Resource Description Framework) [6]
> A framework for describing metadata applications.

OSD (Open Software Description) [7]
> A suggested XML application for automated distribution and updating software.

SMIL (Synchronized Multimedia Integration Language) [8]
> An XML application which enables independent multimedia objects to be integrated into a synchronized multimedia presentation. SMIL is a proposed W3C Recommendation.

What should I be doing to prepare for XML?

If you are an information provider on the Web you should follow the following tips:

- Quote all attribute values e.g. use **`<IMG SRC="logo.gif">`** and not **`<IMG SRC=logo.gif>`**

- Be consistent in use of case in HTML element names e.g. don't use **`<H1>`** in one part of the document and **`<h1>`** somewhere else. and definitely **`<STRONG>don't</strong>`** use different case for start and end tags. Ideally you should define guidelines covering use of case within your organisation.

- Ensure that you validate your documents, or use an authoring tool which creates valid HTML.

You can expect to see a variety of tools which will correct errors in HTM documents and convert them to XML format. Dave Raggett, of the W3C, has written a utility called **`Tidy`** [9] which may be useful. However it is probably advisable to try to ensure that your documents are created correctly in the first place - you cannot guarantee, for example, that the owner of a badly-formed document will be available to make corrections.

If you're developing software you should ensure that your software follows these guidelines.

What else do I need to know?

This article has given a brief introduction to XML. Further information on XML is given below. However as well as XML there are a number of related protocols which you need to be aware of.

XSL [10], the XML Style Sheet Language, will describe the appearance of XML resources. **XLink** [11] will provide a rich hyperlinking mechanism. **XPointer** [12] will provide access to components of XML resources. Further information on XSL, XLink and XPointer will be given in a future "*What Is?*" column.

Further Information

How do I find out more?

Further information on XML is available at the following locations:

W3C's XML home page at <URL: http://www.w3.org/XML/>

The XML home page at <URL: http://www.xml.com/>

Frequently Asked Questions about the Extensible Markup Language at <URL: http://www.ucc.ie/xml/>

The XML Specification, W3C. Proposed version December 1997 at <URL: http://www.w3.org/TR/PR-xml.html>

The Annotated XML specification at <URL: http://www.xml.com/axml/axml.html>

The SGML / XML web pages at <URL: http://www.sil.org/sgml/xml.html>

James Tauber's XML page at <URL: http://www.jtauber.com/xml/>

Tim Bray's XML tutorial at <URL: http://www.textuality.com/WWW7/>

BUILDER.COM's 20 Questions on XML at <URL: http://www.cnet.com/Content/Builder/Authoring/Xml20/>

References

1. *MathML*, W3C, <URL: http://www.w3.org/TR/REC-MathML>

2. *An Extensible Rendering Engine for XML and HTML*, Ciancarini, Rizzi and Vitali, <URL: http://www7.conf.au/programme/fullpapers/1926/com1926.htm>

3. *CDF*, W3C, <URL: http://www.w3.org/TR/NOTE-CDFsubmit.html>

4. *CML*, Virtual School of Molecular Sciences, <URL: http://www.venus.co.uk/omf/cml/>

5. *PGML*, W3C <URL: http://www.w3.org/Submission/1998/06/>

6. *RDF*, W3C <URL: http://www.w3.org/RDF/>

7. *OSD*, W3C, <URL: http://www.w3.org/TR/NOTE-OSD.html>

8. *SMIL*, W3C, <URL: http://www.w3.org/TR/1998/PR-smil-19980409/>

9. *Tidy*, Dave Raggett, <URL: http://www.w3.org/People/Raggett/tidy>

10. *XSL*, Summer Institute of Linguistics, <URL: http://www.sil.org/sgml/xsl.html>

11. *XLink*, Summer Institute of Linguistics, <URL: http://www.sil.org/sgml/xll.html>

12. *XPointer*, W3C, <URL: http://www.w3.org/TR/1998/WD-xptr-19980303>