

INTRODUCTION TO XML

Introduction

XML (eXtensible Markup Language) is a useful tool for creating uniform information formats. It is related to the more complex SGML (Standard Generalized Markup Language). XML is especially useful for groups of people who wish to share information in a standardized format. To a certain extent, XML resembles HTML, because it uses markup symbols such as <xml> in order to describe the document. But unlike HTML, which merely determines how text and graphics are displayed and utilized, XML describes the *kind* of information within the document. One of the great advantages of XML is that tags may be self-defined by the user. As a result, the language is quite flexible and unlimited, unlike HTML. A helpful explanation of XML is provided by www.whatis.com, that states, "XML describes the content in terms of what data is being described. For example, a <PHONENUM> could indicate that the data that followed it was a phone number. This means that an XML file can be processed purely as data by a program or it can be stored with similar data on another computer or, like an HTML file, that it can be displayed." Walsh gives an excellent reason for the usefulness of XML, stating, "XML was created so that richly structured documents could be used over the web. The only viable alternatives, HTML and SGML, are not practical for this purpose."

Significance of the Topic

It is anticipated that XML will grow during the next few years. Because of its content-oriented structure, XML should serve as an aid in Web navigation, by making it simpler for a user to find information on the Web. This is possible due to the fact that XML enables the user to search for *concepts* instead of words. Ryan, stating, "A Web-based search for Columbia will yield results for U.S. cities, the space shuttle, the University, the River, British Columbia, the record company and so forth", outlines an example. However, XML allows users to search within a context, reducing the amount of irrelevant results from a search engine. XML can also be embedded within HTML forms. This is especially useful for large companies in e-commerce environments that exchange information on a rapid, ongoing basis, because it provides a consistent format for the information that may also be summarized and categorized. This discussion also attempts to show the *variety* of possible uses of XML.

PART I: BASIC BUILDING BLOCKS OF XML: THE SELF-DEFINING ASPECT

Examples of some of the building blocks of XML are these five assignment types listed as follows: DOCTYPE, ELEMENT, ATTLIST, ENTITY, NOTATION. The DOCTYPE and ELEMENT types will be discussed in this paper.

According to Stanek, the document type and the element definitions make up the document type definition (DTD). Whatis.com defines the DTD as, "a specification that accompanies a document and identifies what the funny little codes (or markup) are that separate paragraphs, identify topic headings, and so forth and how each is to be processed. By mailing a DTD with a document, any location that has a DTD *reader* (or *SGML compiler*) will be able to process the document and display or print it as intended."

A person composing an XML document can use an *element* to determine and define the types of markup tags that will be used in the document. In order to do this, the programmer simply creates an element declaration, which Stanek defines as "the element name followed by a description of the element's contents." Here is an example of an element declaration:

```
<!ELEMENT main (transaction|contactperson)>
```

In the above example, *transaction* and *contactperson* are *main* elements, as specified in the declaration. Because these two concepts are prevalent in this sample document, they are designated as main elements in the DTD. This would enable a programmer to create separate *transaction* or *contactperson* elements such as the example below.

```
<!ELEMENT transaction (date,idnumber?,product+)>
```

As a result of the above coding, the following markup tags could be used in the body.

```
<date></date><idnumber></idnumber><product></product>
```

The *transaction* and *contactperson* elements are main elements, and thus they contain other elements that are lower in the hierarchy. The elements that are at the *lowest* level of the document hierarchy will not be "parents" to other elements. Therefore, their data type has to be defined, so that the data will translate in the XML compiler. According to Stanek, the item PCDATA will serve this purpose. Stanek defines PCDATA as "a reserved name that describes basic elements that contain parsed-character or raw data, that is the actual text of the document and can include letters, numbers, punctuation, and special characters." In most cases, the lowest level elements are simple concepts, which merely need to be defined as raw data, which the compiler can then translate as words, numbers, or special characters. An example of a tag using this code is below:

```
<!ELEMENT date (#PCDATA)>
```

XML has a vast number of features not discussed above. However, the basic theoretical concepts and possible uses of the code, as outlined above, may provide the reader with a basic understanding for further exploration. The next section illustrates how companies, organizations, and individuals are using XML.

PART II: How XML is Being Used

Enhanced Multimedia Delivery Over the Web

XML is the basis for other technologies, such as SMIL (Synchronized Multimedia Integration Language). SMIL enables the presentation of various media such as text, links, audio, or animation whose content can be varied according to the user's available bandwidth. SMIL supports streaming technology. Therefore, a user does not have to wait for a SMIL presentation to be downloaded in its entirety before viewing the file. In order to view content generated by SMIL, the user must have a player, which is available Windows 95, 98, and NT or Macintosh from www.real.com/g2. Delivery of SMIL content also depends upon a SMIL server. There are two basic player and server types described by Stanek. The second type of player not only supports SMIL technology, but it also includes certain proprietary extensions. According to Stanek, RealPlayer G2 is the most effective player to date, and a free version is available. As for servers, the RealServer is also currently the most popular for delivering SMIL content. There is a version available for ISPs, which may enable a user to upgrade their ISP account to support delivery of SMIL content.

Online Web Course Delivery

One current use of XML for online course delivery is illustrated in an article by Bota. The author states, "We believe that on-line courses should be flexible enough to adapt to a number of different user profiles, that have different entry levels in terms of already acquired concepts and skills, and different learning goals." According to Bota's plan, a teacher with little knowledge of Web courses can build XML documents by using an XML editor. According to Bota, XML editors are becoming more graphic and user-friendly. Once the teacher generates the necessary XML documents, s/he simply assigns score values to every study or test question and makes all of the information available to the system administrator. The Web-based course described by Bota is quite self-contained with various modules for presenting course material, examples, study questions, and tests, with evaluations. The course developers use XML documents firstly to identify the user and to create a user profile. In addition, XML is used to generate test modules. Many aspects of the course are defined as XML elements, such as *question* and *answer* in the case of a test question or study question. Through the use of XML element tags, any student input, such as the answer to a study question, can be passed into a CGI script to determine the student's understanding of a particular topic. If a student shows mastery of the particular topic, the course module will present *only* advanced material and omit all basic information on the given topic. Because so many aspects of the course input are defined as elements in XML, such interactivity is made possible. Therefore, the course truly adapts to an individual's unique skills and mastery of the knowledge.

Document Warehousing and Management

Ishikawa focuses upon the usage of XML for providing easier access to large amounts of data such as email, stored WWW pages, and word processing documents. He describes the need for a document warehouse system to organize and manage the massive

amounts of documents generated particularly in large companies. Through the use of XML, users have access to documents through queries, rather than having to browse large amounts of data. In addition, the document warehouses enable users to analyze document data and to extract keywords from each document for indexing. Another important feature is bi-directional access to documents through the use of both keyword and content-based searching. In short, this type of search provides users with exact document matches as well as related documents. Data warehousing should also allow documents to be stored under multiple categories for easier user access, yet this task can be accomplished without some time-consuming work for individuals. Ishikawa describes the heart of the use of XML for this project in the following quote. "XML data scattered over various Web sites can be distributed databases in general terms. Therefore, we must provide a query facility over distributed XML data. We devise an XML data query language and an XML data view facility based on the query language for customizing XML data and packaging business rules."

Summary

As is readily apparent, XML can be utilized for a wide variety of purposes, due to the flexibility of the language. Its hierarchical structure has many benefits, such as increased online searching for more precise results. In addition to this, it supports personalized, adaptable Web environments according to user input. Perhaps most importantly, it is an excellent way to standardize data that is being shared between large organizations and companies. Since XML is designed especially for Web delivery and has been approved by the W3C, it will most likely be used more and more in the coming years.

References

Bota, Florin; Farinetti, Laura; Rarau, Anca. (2000). An Educational-Oriented Framework for Building On-Line Courses Using XML. IEEE International Conference on Multimedia and Expo, 2000. Volume: 1, 2000. Page(s): 19 -22 vol.1.

Ishikawa, Hiroshi; Kubota, Kazumi; Noguchi, Yasuo. (1999). Document Warehousing Based on a Multimedia Database System. Proceedings., 15th International Conference on Data Engineering, 1999. Page(s): 168 -173.

Ryan, Tracy. (June 2000) XML Promises Better Search, Navigation Techniques. Retrieved October 24, 2000 from the World Wide Web:
http://www.unisysworld.com/un0600/xml_search.htm

Stanek, William Robert. (February 9, 1999). The New Web Format for Multimedia. (Synchronized Multimedia Integration Language). *PC Magazine*, 233-238.

Stanek, William Robert. (2000). Structuring Data With XML. Retrieved October 24, 2000 from the World Wide Web:
<http://www.zdnet.com/zdhelp/stories/main/0,5594,2396941,00.html>.

Walsh, Norman. (Oct. 3, 1998) What is XML? Retrieved October 25, 2000 from the World Wide Web: <http://www.xml.com/pub/98/10/guide1.html>.

Related Links on the Web

Apache XML. <http://xml.apache.org>. Apache is developing multiple projects for various aspects of XML authoring. These products shall follow W3C specifications and address various implementation concerns.

Ask the XML Pro. www.inquiry.com/techtips/xml_pro/. This site includes a long index of answers as well as a search engine. Many answers address XML's integration with other products and tools.

Creating Custom XML Multimedia Elements.

<http://msdn.microsoft.com/workshop/imedia/directx/docs/da/xml.asp>. This article uses one example of animation and shows the reader how to code this in XML. By using XML, developers may reuse scripts for multimedia elements, rather than writing complex script for every multimedia event.

Dublin Core MetaData Initiative. <http://purl.org/dc/documents/rec-dces-19990702.htm>. This group has been working for a couple of years toward a better way to classify Web documents, whether graphic, text, or some other media. The Dublin Core has defined several attributes that shall be set up in XML. In this way, Web documents may be searchable not merely by body content, but also by the type of document, graphic, author, format, etc.

Frequently Asked Questions about Extensible Markup Language. www.ucc.ie/xml/. This is an excellent site for a beginner. It includes simple discussions, such as "What is XML?" and "What XML software can I use today?" Answers often include diagrams and source code examples.

Intranet Design Magazine: Web Development XML. <http://idm.internet.com/XML/>. Intranet Design Magazine has a page devoted solely to XML articles. Currently, several links to articles are displayed with topics such as "Using XML/XSL for Web Publication."

Microsoft -- XML. msdn.microsoft.com/xml/default.asp. This site is geared more toward programmer and developers. It includes information about the latest XML compilers, plus links to XML newsgroups chat, and other resources.

POET XML Resources. www.poet.com/products/cms/xml_library/. This site has links to white papers and XML software. In addition it has links to other XML sites and information about the latest proposed standards.

Robin Cover's SGML/XML Resources. www.oasis-open.org/cover/xml.html. This nonprofit site has the latest news, technical information and product information for XML. Opens with a comprehensive index leading to relevant articles, press releases, and discussion groups.

XML.com. This is the definitive XML resource. It includes a vast amount of articles and updated information posted by XML users. It is provided by O'Reilly & Associates and Seybold Publications.

XML Cover Pages www.oasis-open.org/cover/sgml-xml.html. Oasis sponsors this comprehensive page, full of news, standard papers, and links. A caveat is a section on upcoming XML conferences and contact information for some XML experts.

The XML Files/Web Developer.com. www.webdeveloper.com/xml/. This page, part of Web Developer.com, has a list of articles and descriptions. Resources include links to XML BBS, examples of XML code, and tutorials.

XML Magazine. www.xmlmag.com/. This quarterly online magazine, devoted entirely to XML, includes practical articles to aid in tasks such as choosing an XML server. Includes a wide variety of facets and usages of XML, such as XML's integration with JavaServer pages.

XML.ORG - The XML Industry Portal. Another Oasis-sponsored page, this site focuses on specifications, schemas and vocabularies being developed for XML. This site has a unique registry in which participants can submit relevant material regarding specifications.

XML Spy 3.0. www.xmlspy.com. The site is the home of the first Integrated Development Environment for XML. Provides latest product information and tools for streamlining the XML markup process.

XML Tree. www.xmltree.com. This site has excellent navigation options, such as a browse search, a search engine, or using the Dewey System. Users can also type in a schema in order to find valuable code and other solutions for XML development.

XML Tutorial. www.w3schools.com/xml/default.asp. This site describes itself as an XML school, taking the user from the basics to advanced techniques. Articles cover a wide array of XML topics and tasks. Many articles discuss XML with other tools such as HTML or Cascading Style Sheets.

XML Zone www.xml-zone.com. This is an excellent site that includes a search engine, an introductory article for beginners, and other resources. It also includes message boards, links to the top XML sites, and a code library.

W3C: Extensible Markup Language. <http://www.w3.org>. This page also includes updates and news involving the latest XML standards. Most material is from the World-Wide

Web Consortium. It includes the official W3C recommendation document, dated February 10, 1998.

WDVL: XML: Extensible Markup Language.

<http://wdvl.com/Authoring/Languages/XML/>. This side provides basic information with an overview and some links to other related sites. It also has a link to XML conferences and trade shows.

webreview.com - XML. <http://webreview.com/pub/XML>. Webreview provides a brief page with links to other XML sites and articles. Each link has a brief overview and date of submission.

*This paper was written by Michelle Minto for the course EDC385G Interactive Multimedia Design & Production at the University of Texas-Austin.