XML: An Evolving Technologies Committee Report

What Is Happening In The World of Internet Languages?

There is a rising move to standardize and develop protocols as the prospects for massive data exchange emerge throughout the world. To explain the current movement, you must first have some familiarity with Web page scripts. If you have never actually seen HTML just pull down “View” from your browser menu, and highlight “Page Source”. You will see the “script” behind the page you are viewing and each time you move to a new page and view the source, you will gain a clearer understanding of the scripting language that is used to present a Web page.

HTML is borrowed from SGML, a markup language using (less than – greater than) <> tags for presenting documents on the Web. The key in recognizing the tremendous change in global information exchange about to take place is in understanding the presentation of information via HTML. HTML essentially is about the look and feel, or more specifically, about the presentation of documents on the Internet. Until recently, integrating the data contained in HTML documents has been accomplished only through compatible systems and languages. In some instance, some open, easy script writing in languages such as Perl or Java, have been used across systems to handle some actions on a Web page, such as sending an e-mail or counting page hits. Unless there is a concomitant application/program written in Java or another language on the page, no programming activity takes place.

The Initial Use of the Internet

Most of us can remember the first time we logged onto the Internet using Gopher or Mosaic. Until this time in computing history, we had rarely been able to successfully exchange documents or print e-mails. Initially, the idea of receiving an e-mail and being able to see a document on the screen that could be retyped was a big deal, since it took only a few hours or less to transmit. This was overwhelmingly useful to researchers and university colleagues around the world who previously had to mail paper copies of their work and edits back and forth.

Within a few months of news on the capability of what was being called the World Wide Web, computing center gurus typically had found some old box to use as a server and allow people to post a few important documents along with what was dubbed vanity pages. Beyond the initial chaos, disorganized goals and objectives, and in many cases, total lack of vision, the new capability allowed people to be creative and energized. This clearly monumental and new capability for users, i.e., logging on to an address (that was usually sent to you via e-mail or over the phone since exchanges were rarely randomly selected through searches at this time) and reading through an entire document that could be cut and pasted was pretty amazing stuff. How we have moved from initial postings on Gopher, to the use of Mosaic--then Netscape, search engines, and now, to the breathtaking choices in access, so too, will the content of Internet information be continuously cleansed and converged. To demonstrate one of the new evolutions, let’s look at XML, a way to make the aspirations to easily exchange data over the Internet a reality.

XML is a simplified subset of SGML, but with many additional characteristics and is used for data exchange on the Web. It is not used to show columns or to size fonts as does HTML. On the other hand, HTML is not used to process information. XML is used to structure a document, not present it. It is frequently defined as self-describing which accommodates flexibility, collaboration, and eases reuse. This definition is obscure for
many people outside the application development arena, but it has to do with allowing the Web page developer to create tags that do not already have a fixed role.

In other words, if you look at the below sample collection of objects in self-describing tags in XML, you will notice that it looks a lot like HTML, but also that it looks a lot like a typical root-styled directory or tree organization. What is different is that in HTML, the definitions within the tags are predefined. Example: in HTML if you want a word to present itself as bold, you would use the <b>bold</b> open and close tags. The presentation would show up on the browser as: bold.

On the other hand, below are self-defining XML styled data in a sweetheart tree.

```xml
<sales>
  <sheet music>
    <titles>
      <with colors>
        <Love Is Blue>
        <When Sunny Gets Blue>
        <Song Sung Blue>
        <Evergreen>
        <White Christmas>
      </with colors>
      <with names>
        <Maria>
        <What Will Mary Say>
        <Gina>
        <When Sunny Gets Blue>
      </with names>
    </titles>
  </sheet music>
</sales>
```

This example demonstrates that the words like sales and titles do not present information. Instead, it shows words can have a data table-like conceptual role. You may think of potential uses with additional data, i.e., “How many sales, of sheet music from the fifties were posted in 1995?”

This example shows a structure that is a result of data organized between elements. If your data is not structured, it cannot be exchanged easily between applications. Unstructured data does not include rules or repeating elements. It is disorganized. XML is a method for putting structured data in a text file. In part, XML is a set of rules, guidelines, conventions, whatever you want to call them, for designing text formats for exchanging data. It is accomplished in a way that produces files that are easy to generate and read (by a machine), that are obvious and clear, and that transcend current deficiencies in Internet data exchange.¹ This is accomplished with the help of a parser on a server.²

XML functionality is similar to that provided by EDI (electronic data interchange) applications. They too used structured data to exchange data between entities. There must be exchanges of data by all parties involved to be electronic data interchange. In the past, only those entities, such as banks, that could see a high return on their investment in EDI technologies. Those were the days, of course, when it was still possible to effectively use standards among many entities, not to mention, afford to install elaborate EDI systems.³ XML creates EDI-like interchange, but is not as complex nor as costly as the traditional EDI of the early 90s, at least until recently.⁴ Now EDI technology is evolving with XML as its survival may depend on both methods ability to work better together. XML’s self-defining extensibility can support international language exchanges, multiple distinctions in business processes, such as banking processes, scientific markings, to name a few, and is hardware and software independent. Be aware, however, that just because the process is open and extensible, does not guarantee that business-to-business enterprises will agree on the use of words or symbols needed to share data, i.e., you may say tomato and I might want to use tmt for the tomato tag. If we don’t agree, my purchases of your tomatoes will not communicate with your invoices. These differences are staggering even within data resource management groups in
universities. Once suggested protocols and standards are to be used throughout the global network—well, the chaos may be beyond calculation.

Nevertheless, a current goal of XML is provide a framework for more meaningful searches – across databases accessible on the Internet. It also enables (through the structure and ultimately accepted standards) a way for multiple databases to be combined.\(^5\)

This may be one of the most important applications for university researchers, regardless of discipline or interests. This is because almost all research includes excessive amounts of supporting data that must be updated regularly and sometimes from multiple sources. To date, sharing a database through the current capabilities of Internet exchange has been expensive and time consuming. With XML, or something very much like it, data sharing capabilities will be limitless.

Will XML survive the future and can you predict a future that includes XML?

In a recent article published by xmlhack.com, there are notes from a presentation by Elliotte Rusty Harold called the “Future of XML”. His last prediction was that browsers wouldn't support XML until 2002 and that, when broad support comes, Mozilla would regain market share from Internet Explorer. (Only recently are none IE browsers supporting XML directories.) In spite of this delay, he did see browser support as “an inevitability spurred by fragmentation of the Web by non-PC devices such as cell phones and hand holds.” These, he thought, would force the adoption of XML as a browser-independent format.\(^6\)

It may be that XML is the Gopher of the past, and may transform itself several times as well, but it is surely a burgeoning of standard for data exchange on the Internet that allows one to envision the full ramifications of Wireless and Internet communications. These incremental developments tell us that we are moving into new territory where research, teaching, business, entertainment, and administration will be advanced by the ability to share database activity. This promise of a new world of information is personified by the new word ‘datafication’ used to describe it.\(^7\)

If you are interested in seeing how early adaptations of XML will look, a Web site called, “Directory of Content – an Exhibition in XML”\(^8\) is available for you to experience. The first thing you will notice is a Home Page that lists topics that appear to be distinct and unrelated. Practice the data access by selecting a subject area such as, “Recreation.” The site allows you to refine your search further by subjects such as, Autos with 33 additional links and Aviation with 19 additional links or, for example, Travel with 199 links. If you were to click on Travel, you will be given options to refine your search again by subjects such as rail, budgets, and health. If you click on health, one option is currently available. MedServ Medical News, which “provides medical information and news to medical professionals worldwide….“ Various options to do something with this page are available. What would you want to add? Say you work for the Centers for Disease Control, hypothetically, you might want to add links with specific information about inoculations or recent warnings and alerts for travelers. You can add a keyword or a new XML directory that will be integrated into this one. The vision becomes one of a universal knowledge bank with a much more reasonable way to find data that is of specific interest.

Examine how the classification works so that you may see the potential for XML. Under the Keywords, you will see all the trees of information within the data that includes the subject Health. I followed the path, Subject > Recreation > Travel > Health. But there is also the Subject > Health > Consumer Support Groups

> All Health PORTALS;
> Men's Health; Subject
> Women's Health Issues, to name a few.

Try, Subject > Recreation > Travel > Lodging. This path will take you to the Portugal Travel Guide. You will discover that this travel guide is also classified a resource under the following Keywords:
Recreation > Food > Dining Guides
Language > English
Language > Portuguese
Location > Europe > Portugal

The Open Directory Project touts, "As the Web grows, automated search engines and directories with small editorial staffs will be unable to cope with the volume of sites. The Open Directory Project’s goal is to produce the most comprehensive directory of the Web, by relying on a vast army of volunteer editors." You can volunteer to be an editor, and help organize information according to your expertise.

Why Is XML Important To Universities?

This one evolving technology nested within the vast array of changes taking place in Wireless communication and other Internet technologies, is a hugely important feature of Internet development, going beyond the current boundaries of just sharing information. Says Ira Fuchs, "We are institutions of higher education. Our open, collaborative environments are essential to our nature." Today, there are hundreds, maybe even thousands of people in higher education contributing to XML development, its testing, and standards. (Not to the exclusion of hundreds of other new technologies such as PDAs, Internet2, e-books, to name a few.) Only by browsing through the recent work of many committees and consortiums on upgrading Internet information, can one really appreciate the future waiting to unfold.

"One should not conclude that the Internet has now finished changing. The Internet, “will …continue to change and evolve at the speed of the computer industry if it is to remain relevant. It is now changing to provide such new services as real time transport, in order to support, for example, audio, and video streams. The availability of pervasive networking along with powerful affordable computing and communications in portable form (i.e., laptop computers, two-way pagers, PDAs, cellular phones), is making possible a new paradigm of nomadic computing and communications.” Undoubtedly, there will be a growing extension of early integrations of technology into teaching. The first and probably most popularly use of conferencing tools were quickly adopted in many classrooms. It is likely that these precedents will require on-going advanced additional integrations with wireless and PDA devices.

Nor should we conclude that XML or any one standard will succeed easily. The financial interests involved in the Internet now go way beyond the less fractious development among early pioneers. However, there are sure signs of success hanging in the wings. The emerging style sheet standards in XML fills gaps in Internet publishing that have long been a thorn in the side of publishers. XML is poised to become the one method for exchanging information among all sorts of devices besides typical computers. For example, a hand held device such as a cell phone or PDA may use a Global Position System (GPS) device to receive Internet enabled information. Currently, such devices do not have much memory. This example illustrates the diversity of purpose for converting data to an XML standard. There are a growing number of reasons for using a standard for exchanging valid data on the Internet. WML (Wireless Markup Language) is just one of hundreds of many new specialized markup language protocols under construction within the rubric of XML.
We now can be confident that something equivalent to the current architecture of XML, i.e., some evolutions of it, and with the similar chaos that has been associated with Internet development, will surely develop in the same way, and will be the standard for making many types of devices work together and improve data sharing over the Internet.

For EDUCAUSE 2001 – Evolving Technologies Committee

By Laura Joyce Moriarty
Emory University ITD
Ljm@emory.edu – 404.727.7663
This and the sites below contributed to the explanation of XML technology.

Cafe con Leche XML News and Resources
http://www.ibiblio.org/xml/

Late Night News Software Ltd. The Application

XML
http://www.xmlsoft.org/#XML

http://www.w3.org/TR/xmlschema-1/

Ralf I. Pfeiffer, IBM XML Technology Group XML Tutorials for Programmers

Advanced XML
http://www.ibiblio.org/xml/slides/sd2000east/advancedxml/

Wen-chieh Chuang, Introduction of XML
http://128.255.23.177/xml/xml_lecture.htm

Search the XMLphant resource database
http://www.xmlphant.com/

XML and JUMBO at WWW6
http://www.vsms.nottingham.ac.uk/vsms/java/jumbo/w3_blurb.html

W3XML-Data: W3C Note 05 Jan 1998

XML.org – The XML Industry Porta
http://www.xml.org/xmllrg_resources/whitepapers.shtml#resources

Welcome to the Apache XML Project: The goals of the Apache XML Project are:
http://xml.apache.org/

W3C Schools: Free Web Building Tutorials
http://www.w3schools.com/default.asp

---


3 Electronic Data Interchange (EDI) Standards
http://www.diffuse.org/edi.html

4 Lenny Liebmann, *XML’s Tower Of Babel - Industry-specific dialects may sound good, but they can create more confusion than communication*. InternetWeek, April 24, 2001.

   Note: This is an interesting perspective on the monumental task of sorting out the hundreds, if not thousands of standards, permutations, and combinations that are infiltrating XML development work.
http://update.internetweek.com/cgi-bin4/flo?y=eDV60BeHeo0V30NFW0Ac


6 Elliotte Rusty Harold on the Future of XML, 16:52, 13 Nov 2000 UTC.
http://www.xmlhack.com/read.php?item=891

7 John Seely Brown and Paul Duguid, FIRST MONDAY, *The Social Life of Information*
http://www.firstmonday.dk/issues/issue5_4/brown_chapter1.html

8 XML.TREE, *Directory of Content*
http://www.xmltree.com/index.html

9 The Directory Project
http://dmoz.org/
See above list of XML sites and especially W3C Consortium links for committees and memberships. Also See: EDUCAUSE Comittees and Memberships at: http://www.EDUCAUSE.edu/memdir/member_info.html
