

Keeping Up with Web Technologies:
A Review of **Presenting XML**



Presenting XML, Richard Light (with forward by Tim Bray) Sams.net Publishing. Copyright 1997 (First Edition). \$24.99, softbound, 414 pages plus companion Web site. ISBN 1-57521-334-6. Category: Internet-Web Publishing.

You've struggled with HTML; you've heard of SGML. What is XML and why should you care? Will you have to throw away all your knowledge and investments in HTML? Will knowing XML be a requirement for developing Web sites? Does XML have any value *outside* web sites? This book will explain the meaning of this new acronym and help you decide the answers to these questions.

An SGML Refresher

To understand XML, you need some familiarity with SGML and HTML. In 1986, SGML became International Standards Organization standard ISO 8879:1986. This standard described not a document tagging language but a meta-language, a set of rules for creating tagging languages. Following the SGML rules, you could define a set of document tags that -- together-- become a model describing any type of document's structure. Define a document model (or "Document Type Definition"), and then give me your DTD and a document tagged using it (called a "Document Instance") and look and feel, attributes like font, point size, page dimensions become irrelevant to my understanding your document. This powerful insight — keeping content and appearance separate — made it possible to write a document once and publish it on many different media: CD-ROM, HELP files, computer screens, and traditional paper (normal, digest, and large-print editions). SGML also provided some independence from changing vendors and technological winds — definitely attractive. SGML also provided a way for companies supplying documents within industries or to the U.S. Government a way to do so uniformly.

SGML is a good but complex standard. A subset was needed to allow a large number of authors to publish documents on the World Wide Web.

HTML works fine for simple Web documents, and HTML is based on SGML. However, because it was defined to be easy, it was designed so that "one size fits all." As long as you were satisfied with the tags HTML had to offer, you could (and will always be able to) use HTML and forget about other tagging schemes. The web grew like Henry Ford's original auto business where you could have any color Ford you wanted as long as it was black. As in the early auto business, people soon wanted more. To satisfy differing requests, HTML grew in size and complexity and became decreasingly adequate.

In 1996, a Web compromise was developed by a group of 80 SGML experts over a tense 11-week period applying the 80/20 rule: Remove all but about 30 pages of a several hundred page SGML standard. Keep the ability to develop unique document models, but remove most of SGML's complexity. The result was XML.

Presenting XML

For a subject as esoteric and potentially mind-numbing as XML, Richard Light does an incredibly good job of both organizing and presenting the material understandably. He subdivides his book into four parts, starting with Part I, "Introducing XML." By the end of Chapter 1, only 18 pages long, you understand both document markup concepts and why you should read further. He asserts that XML will be used by many diverse organizations, ranging from technical book publishers to those dealing with medical records and even EDI financial transactions.

Part II explains the details of XML, and comprises 6 chapters, nearly 150 pages. This is not a quick read-- but written as clearly as the subject allows -- Light explains the similarities between XML and SGML.

Part III, Using XML, includes an exercise transforming an HTML 2.0 Document Type Definition to make it XML-compatible. You can even check your work by running an XML validator using freeware. The point of this exercise isn't to make you fluent in XML, or to suggest that this conversion exercise is something that will be commonplace. Rather, the exercise demonstrates the differences between the two standards and what can go wrong should you be required to convert HTML-tagged documents to XML.

Part IV, The Future of SGML, presents not only some enticing possibilities for XML but also provides an excellent list of resources, including XML freeware. Examples of enriched Web applications that XML could provide include richer searching, horizontal (hierarchical) links, and more focused subscriptions to information via push agents and the Channel Definition

Format.

Shortcomings?

Light readily admits the only shortcoming in this book, timeliness, and this is caused by the dynamic nature of XML and document technologies. And who would have predicted Bill Gate's ringing endorsement of XML in the San Francisco Seybold Seminar '97 conference? The book is only as complete as it could be using the March 1997 draft of the XML language standard. However, Light remedies this by providing the URL where you can find updates. And Sams.net has also provided a web site supporting the book and containing HTML, SGML and XML samples. I wish the book had included a searchable electronic medium, but with the book's low price you don't mind visiting the suggested web sites to get what you need.

Light's book focuses on the value of XML as a web standard, and does not explore other SGML-like areas of application. Web content is increasingly being integrated with other (such as CD-ROM media) and it is likely XML will find application beyond the web.

Summary

Like the XML standard, Light wrote this book in "Internet time." Even with its admission that some parts of the book are subject to change, this is the best treatment of XML I've seen. If you are involved with web site development or design, re-purposing of document content, or are simply interested in what is likely to be a key document technology, buy this book.

©Bob Boeri. Bob Boeri works by day as a Document Technologies Consultant for Factory Mutual. After hours, he is also a columnist and frequent contributor to *EMedia Professional* and the *Boston Software Newspaper*. bboeri@world.std.com