

Learning XML

Creating Self-Describing Data

book review by Sandy McCourt

Learning XML by Erik T. Ray

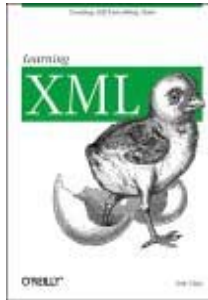
Publisher: O'Reilly & Associates - January 2001
ISBN 0-596-00046-4, 368 pages \$34.95 - from
Amazon.co.uk (via UK-BUG site) £19.96 (-20%)

Publisher's blurb:

XML (Extensible Markup Language) is a flexible way to create "self-describing data"—and to share both the format and the data on the World Wide Web, intranets, and elsewhere. In Learning XML, the author explains XML and its capabilities succinctly and professionally, with references to real-life projects and other cogent examples. Learning XML shows the purpose of XML markup itself, the CSS and XSL styling languages, and the XLink and XPointer specifications for creating rich link structures.

O'Reilly books have a good record. Well-known for their ...in a Nutshell series and their zoological cover illustrations, I've always found them highly satisfactory. There cannot be many works that can be referred to familiarly as, for instance, *The Camel Book*. Well now we have *The Newly-Hatched Chick* book.

From the title of this one I had somehow (wrongly) expected an XML primer covering only basics, but this is a thorough treatment of the subject. The book's general layout conforms to what we have become familiar with in O'Reilly.



The first chapter is a very readable coverage of: what XML is; its terminology; features; development history and tools.

Chapter 2 is perhaps the most important chapter covering the fundamental building blocks, concepts and syntax – in fact the anatomy of an XML document. Still very digestible, there are copious commented examples based on real situations as there are throughout the book. One of these, returned to and expanded on in later chapters, is an illustration of a simplified version of DocBook - XML used for a variety of technical documentation tasks; a standard which enjoys a large base of users including developers and publishers. Here, too, we are introduced to the Document Type Declaration (DTD) which specifies a document model and provides a set of rules against which a document can be checked to establish whether it is valid.

Later chapters deal with resources and links and clarify the relationship between XML and HTML and introduces XHTML – an application of XML very similar to HTML4 which makes HTML compatible with XML rules and, therefore, with XML tools and processors. (The book assumes a basic knowledge of HTML.)

The distinction made in XML between content and presentation (unlike, say, word processors such as MSWord) is usefully explored as a topic to introduce stylesheets – keeping the appearance out of the document and thus allowing different presentations of the same content – on paper or on the Web, for example. Cascading Style Sheets (CSS) are covered with examples, culminating in a CSS for an XHTML document.

Chapter 5 is an exhaustive treatment of the syntax of the DTD plus a mention of an alternative – the XML Schema.

Much of the remainder of the book deals with transformation – repackaging the information in one XML document into a different document type using the Extensible Stylesheet Language for Transformation (XSLT) which parses the structure of a document and rebuilds it according to a set of instructions. An example here shows how to handle a bank account, payments and deposits, with a statement produced in HTML format. It's followed by the further development of the Barebones DocBook from chapter 2.

Other topics include multi-lingual XML documents and XML processors including how to write your own.

The index is excellent and there are appendices giving a glossary and lists of standards, tools, books and resources and where to find them.



Learning XML, very readable as well as being a good reference, is strongly recommended.