



Capstone Courseware, LLC

33 Boylston Street
Jamaica Plain, MA 02130

877-227-2477
capstonecourseware.com

Introduction to XML

Will Provost

Instructor's Guide

Revision 1.1.1



Revision Notes

Revision 1.1 upgrades the coursebook to current technology and Capstone standards.

- Covers XML 1.1 and Namespaces 1.1.
- New chapter breakdown to pull related-technology topics to a new final chapter instead of covering them up front.
- Focus on Chapter 3, which now covers less on DTDs and gives about equal weight to Schema.
- More depth on namespaces.
- Detailed table of contents.
- Code listings are visually clearer.
- Diagrams have been redone to be more visible in classroom projection.

Revision 1.0.4 implements minor fixes – mostly book typos.

Revision 1.0 is the initial public release.





Course Overview and Philosophy

This one-day module provides an introduction to XML, solid grounding in reading and writing XML documents, and an introduction to DTDs and XML Schema. It is intended for a broad audience, including relatively non-technical students who need only to understand the role of XML in their work, and possibly to develop some fluency in reading well-formed documents.

All of Capstone Courseware's Core XML modules present what might be called "pure XML", by which we mean two things. Firstly, everything in these modules is based strictly on W3C specifications, without any vendor-specific extensions. Secondly, no knowledge of any particular programming language or other external technology is required to participate fully in the training. As with almost any XML training, the instructor will have to be prepared for a diverse student body. We have tried to make as few assumptions as necessary in writing each module, keeping terminology simple and introducing parallels to other architectures and paradigms only as asides in the text.

Much has been made of the difference between XML proper and the "information set". While we believe that the information set is the more important body of knowledge for the intermediate-to-advanced XML author or developer, we have also concluded that the language itself is the best approach to that knowledge, which is to say that it's best to learn the character representations and grammar first. So, throughout the module, you will find that our presentations favor the XML representation of a component over the abstract definitions of that component, with occasional reminders about underlying concepts, the independence of a particular information set from its representation, or validation rules or constraints that can't be represented in XML proper. We advise instructors to listen to student questions and discussions to assure that the split between representation and information is clear in the students' minds, and to reinforce this as necessary.

We've made a similar choice regarding DTDs. Although at the time of this revision they are fading from common use, DTDs are still much easier to learn, and as such make it easier to learn about XML models themselves. So we cover DTDs first, then segue into XML Schema.

More specific discussions of module and chapter contents, and suggestions for effective teaching, appear on the pages following the timeline below.





Timeline

Day 1

Chapter 1	A Brief History of XML
Chapter 2	XML Grammar
Chapter 3	Valid XML
Chapter 4	Using XML in Applications

The first two chapters should be covered in the morning, and hopefully a little head-start on Chapter 3 can be achieved before lunch. Chapter 3 is by far the longest and has the most lab work. Chapter 4 can safely be left off in favor of thorough coverage of Chapter 3, or some of the labs can be shortened or skipped to make Chapter 4 fit.





The XML Validator

As you will see in the coursebook and the course setup guide, Capstone Courseware provides a standard tool to support XML parsing and validation in the classroom. This is a simple wrapper around the Xerces parser from Apache: a Java application implements a command-line interface; **parse** and **validate** scripts further simplify usage; and the pluggable tool configuration integrates this with the Crimson text editor. For most classes you should find this basic setup is most appropriate, but here are a few additional notes on the tools should you want to customize a particular class. Note that Capstone Courseware cannot provide technical support for any of these options; if you are at all uncertain you should stick with the standard setup.

- If you don't prefer to use the Crimson editor, of course you won't need the tool pack, either. Most serious text editors (anything much better than Notepad!) have some ability to trigger external scripts from their Tools menus, and you should find that the **parse** and **validate** scripts work easily from TextPad, UltraEdit, Syn, etc. – or for that matter from more sophisticated IDEs such as Visual Studio or Eclipse, if you don't like the XML tools built into them.
- The standard setup assumes only a Java runtime, and hence includes the Xerces implementation as part of the download. There is an alternate version (also mentioned in the online setup sheet for this tool) that can be used when Xerces is already present thanks to a previous J2EE, JWSDP, or JAXP installation. If you are comfortable with Java class path configuration, you may also choose to modify the **parse** and **validate** scripts to include the Xerces JAR from any other location on a student machine.
- Scripts **parse_pause** and **validate_pause** are also available. These are handy for attaching to the XML file type in Windows as "actions." So you might assign each of these scripts so that they appear as right-button menu items for an XML document in Windows Explorer. These two scripts hold the resulting DOS console on the screen until the user presses a key – just for this purpose.

Introduction to XML

Chapter 1 – A Brief History of XML

This chapter is intended to answer, comprehensively, the many questions students of various backgrounds may have about the reasons for XML's existence and use. The chapter assumes very little about student background, and paints a broad-brush picture of XML and its use in various software tasks. There is no lab for this chapter (one quickie demo); by the end, students should be clear on basic terminology and should have a good sense of what will be expected of an XML document for various purposes. The "XML scenarios" introduced here are sprinkled through the rest of the course, brought in as they are useful in introducing/motivating new technology topics.



Chapter 2 – XML Grammar

Coverage in this chapter and the next is roughly broken into well-formed XML and valid XML. However, until models are understood, there are a few aspects of XML itself that are hard to understand. Thus this chapter is quite brief while the next is much longer. Over this chapter students should get comfortable with hand-writing small XML documents: the structure of the document and the syntax of well-formed elements and attributes, comments, and processing instructions.

Chapter 3 – Valid XML

This is a long one! Ostensibly covering DTDs and Schema, this chapter also develops additional knowledge about XML itself, as features of a vocabulary such as ID/IDREFs or key/keyrefs, and defaulted or fixed attributes, and especially namespaces lend to a sense of how to express information well in XML itself. This chapter will likely take about a half a day by itself, and breaks up into three natural sections: DTDs, namespaces, and XML Schema.. Both for its own sake and to keep pace through this long chapter, emphasize hands-on work in all the labs over extensive discussion: let students absorb the various syntactic constructs of XML as a way of getting to the underlying concepts.

Chapter 4 – Using XML in Applications

Here we provide an overview of the technological context surrounding XML itself: XSLT, parsing/processing, and Web services. Certainly there are plenty of other XML specs and technology out there! These have been chosen as the most popular and currently practiced, and should help to connect the day's work to potential applications in students' own jobs and projects.





Troubleshooting and Tool Tips

First, be sure to be familiar with the lab and environment setup instructions in the Table of Contents. Assure that the instructions in the module Setup Guide have been followed, and that you know the locations of the installed tools. If, having followed those steps, a classroom machine is failing to run demos or lab answers correctly, here is a list of items to consider and to doublecheck:

- Make sure the **c:\Capstone\XMLTools** directory is on the path to use the **parse** and **validate** scripts from a command console.
- If you get errors that certain classes are not found when you run either of these scripts, be sure that you have either downloaded and set up the standard XML Validator (which includes Xerces) or have Xerces set up on your system and have modified the scripts to include the Xerces JAR in your class path. See the earlier section on using the XML Validator.
- Running **validate** on a file with no DTD or schema reference will generate errors; use **parse** instead.
- There is a little kludge in how the tool validates schema documents themselves: it checks the file for the extension **.xsd**. If this is found, the tool goes to the extra trouble of loading the “schema for schema” into the parser before validating. This is unnecessary for ordinary documents. In the normal run of this course, this is not known to be a problem, but it’s certainly possible for the tool to trip up if you point it at a non-schema doc with the **.xsd** extension or a schema doc without it. There are no plans to clean this up – the tool is only meant to support Capstone XML courses, after all, and was never offered as an industrial-strength parsing tool.





Feedback

We truly do welcome feedback, both of a specific nature (pointing out mistakes) and general suggestions. For the former sending email with a numbered list of corrections would be most helpful.

Please send feedback to:

Will Provost
Capstone Courseware
<mailto:provost@capstonecourseware.com>
www.capstonecourseware.com

