CHAPTER 13

Defining Schemas and DTDs

XML ALLOWS YOU TO CREATE your own elements and attributes to define data for your Web application. You can basically create your own vocabulary for describing your data. But how do you communicate that vocabulary? How do you validate that the elements and attributes in the XML document conform to the defined vocabulary? This is the purpose of Document Type Definitions (DTDs) and schemas.

Both DTDs and schemas provide a way for you to define the valid structure and content of your XML document. The DTD or schema is the dictionary for your vocabulary, providing documentation for the elements and attributes used in your XML document.

Because the DTD or schema defines the valid structure and content of an XML document, you can use it to validate your XML. This is particularly useful if you are receiving XML from another source, such as another application, and want to ensure that the data is valid.

The purpose of this chapter is to provide you with an introduction to DTDs and schemas. Because schemas are now recommended by the World Wide Web Consortium (W3C) instead of DTDs, the focus of this chapter is on schemas.

What Will This Cover?

This chapter covers the following key DTD and schema concepts:

- Understanding schemas
- Looking at schemas vs. DTDs
- Examining the anatomy of a schema file
- Creating a schema
- Using a schema to validate your XML

By the end of this chapter, you will know how to use schemas to validate your XML.
Schema Basics

When creating an XML document, you define your own vocabulary based on the requirements of the data within the XML document. Schemas are the current W3C recommendation for documenting that vocabulary. They provide the formal description of the structure and content of your XML documents.

For example, an XML document containing a list of courses could have a vocabulary that includes courses, course, name, date, and location. These words have special meaning within the context of that XML document.

The schema for the course XML document would define a courses element that consists of one or more course elements, a course element that consists of a set of name, date, and location elements, a name element that is a string, a date element that is a string (or a date), and a location element that is a string.

The schema provides the documentation of your XML data structures and includes the following information:

- The complete list of elements used in the XML document
- The complete list of attributes for each element, including default values
- The structure and hierarchy of the XML document in terms of elements and sub-elements
- The order or sequence of elements within the XML document

Because the schema defines the valid set of elements and attributes required for the XML document, you can use the schema to validate the contents of the XML document. After your XML document has been validated using a schema, it is called a valid XML document.

For example, if you produce or receive a course XML document that is missing a name element, the schema can evaluate that XML document and determine it is not valid.

An XML document can be well-formed without being valid. A discussion of well-formed versus valid XML documents is often the first topic in books devoted to XML. A well-formed XML document is one that follows the correct XML syntax as defined in Chapter 11, “Introduction to XML.” A valid XML document is one that contains valid data based on a schema or DTD.

You don't have to have a schema (or DTD) for your XML document. They are optional but recommended if you are sharing your XML document with other applications or other developers. The schema provides the documentation of the vocabulary to aid in understanding and correctly working with your XML document.