20

XML and LINQ to XML



20.3 W3C XML Schema Documents

- Unlike DTDs, schemas use XML syntax and are actually XML documents that programs can manipulate.
- Like DTDs, schemas are used by validating parsers to validate documents.
- In this section, we focus on the W3C's XML Schema vocabulary.
- XML Schema enables schema authors to specify that element quantity's data must be numeric or, even more specifically, an integer.
- An XML document that conforms to a schema document is schema valid, and one that does not conform is schema invalid.
- Schemas are XML documents and therefore must themselves be valid.

Validating Against an XML Schema Document

Outline

- By convention, schemas use the **xsd** extension.
- Figure 20.3 contains markup describing several books.

book. xml

```
<?xml version = "1.0"?>
                                                                                      (1 \text{ of } 2)
   <!-- Fig. 20.3: book. xml -->
   <!-- Book list marked up as XML -->
                                                                              The books element has the
                                                                              namespace prefix dei tel, indicating
   <deitel: books xmlns: deitel = "http://www.deitel.com/booklist">
                                                                              that the books element is a part of
      <book>
6
                                                                              the namespace
7
         <title>Visual Basic 2008 How to Program</title>
                                                                              http://www.deitel.com/boo
      </book>
8
                                                                               klist.
10
      <book>
11
         <title>Visual C# 2008 How to Program, 3/e</title>
      </book>
12
13
      <book>
14
         <title>Java How to Program, 7/e</title>
15
      </book>
16
```

Fig. 20.3 | Schema-valid XML document describing a list of books. (Part 1 of 2.)

book. xml

```
17
                                                                                      (2 \text{ of } 2)
18
      <book>
         <title>C++ How to Program, 6/e</title>
19
20
      </book>
21
      <book>
22
23
         <title>Internet and World Wide Web How to Program, 4/e</title>
      </book>
24
25 </deitel:books>
```

Fig. 20.3 | Schema-valid XML document describing a list of books. (Part 2 of 2.)

Creating an XML Schema Document

• Figure 20.4 presents the XML Schema document that specifies the structure of **book**. **xml**.

book. xsd

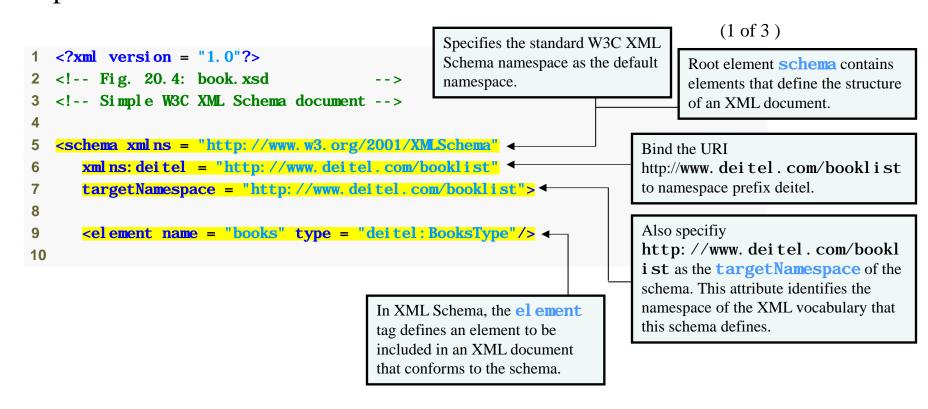


Fig. 20.4 | XML Schema document for book. xml. (Part 1 of 3.)



book, xsd

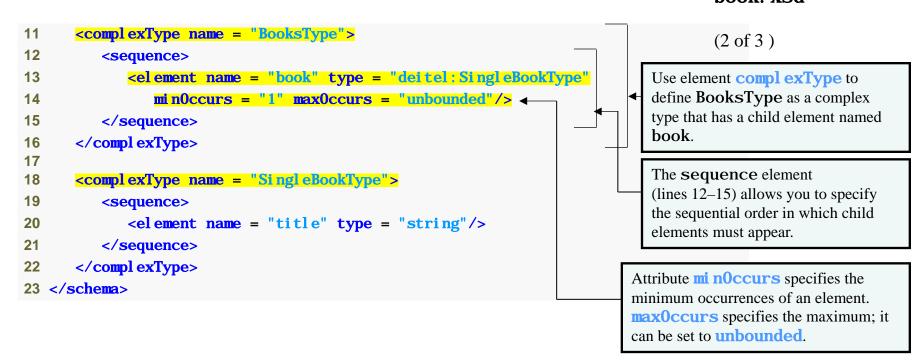


Fig. 20.4 | XML Schema document for book. xml. (Part 2 of 3.)



book, xsd

(3 of 3)

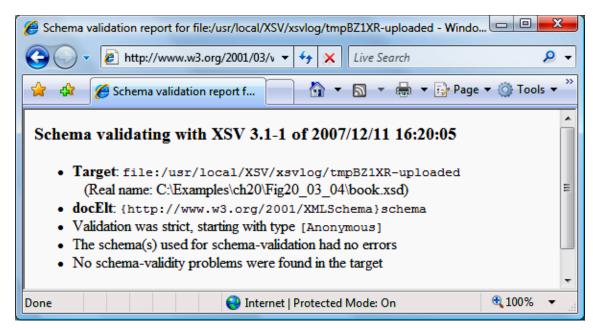


Fig. 20.4 | XML Schema document for book. xml. (Part 3 of 3.)



- This schema defines an XML-based language (i.e., a vocabulary) for writing XML documents about collections of books.
- The schema defines the elements, attributes and parent-child relationships that such a document can (or must) include.
- The schema also specifies the type of data that these elements and attributes may contain.
- Root element schema contains elements that define the structure of an XML document.

Defining an Element in XML Schema

- In XML Schema, the **element** tag defines an element to be included in an XML document that conforms to the schema.
- An element's **type** attribute indicates the data type that the element may contain.
- Possible types include XML Schema—defined types and user-defined types.

• Figure 20.5 lists several of XML Schema's many built-in types.

Туре	Description	Ranges or structures	Examples
string	A character string.		"hello"
bool ean	True or false.	true, fal se	true
deci mal	A decimal numeral.	$i * (10^n)$, where i is an integer and n is an integer that is less than or equal to zero.	5, - 12, - 45. 78
float	A floating-point number.	m* (2°), where m is an integer whose absolute value is less than 2 ²⁴ and e is an integer in the range - 149 to 104. Plus three additional numbers: positive infinity (I NF), negative infinity (- I NF) and not-a-number (NaN).	

Fig. 20.5 | Some XML Schema types. (Part 1 of 3.)



Туре	Description	Ranges or structures	Examples
doubl e	A floating-point number.	m* (2°), where m is an integer whose absolute value is less than 2 ⁵³ and e is an integer in the range - 1075 to 970. Plus three additional numbers: positive infinity, negative infinity and not-a-number (NaN).	
l ong	A whole number.	- 9223372036854775808 to 9223372036854775807, inclusive.	1234567890, - 1234567890
i nt	A whole number.	- 2147483648 to 2147483647, inclusive.	1234567890, - 1234567890

Fig. 20.5 | Some XML Schema types. (Part 2 of 3.)

Туре	Description	Ranges or structures	Examples
short	A whole number.	- 32768 to 32767, inclusive.	12, - 345
date	A date consisting of a year, month and day.	yyyy-mm with an optional dd and an optional time zone, where yyyy is four digits long and mm and dd are two digits long. The time zone is specified as +hh: mm or - hh: mm, giving an offset in hours and minutes.	2008-07-25+01:00
time	A time consisting of hours, minutes and seconds.	hh: mm: ss with an optional time zone, where hh, mm and ss are two digits long.	16: 30: 25-05: 00

Fig. 20.5 | Some XML Schema types. (Part 3 of 3.)

- Two categories of types exist in XML Schema—simple types and complex types.
- They differ only in that simple types cannot contain attributes or child elements and complex types can.
- A user-defined type that contains attributes or child elements must be defined as a complex type.
- The **sequence** allows you to specify the sequential order in which child elements must appear.
- Attribute mi noccurs specifies the minimum occurrences of an element. maxoccurs specifies the maximum; it can be set to unbounded.
- Both of these attributes have default values of 1.

A Closer Look at Types in XML Schema

- Every element in XML Schema has a type that is either built-in or user-defined.
- Every simple type defines a **restriction** on a previously defined type.
- Complex types are divided into two groups—those with simple content and those with complex content.
 - Both can contain attributes.
 - Only complex content can contain child elements.
- Complex types with simple content must extend or restrict some other existing type.
- Complex types with complex content do not have this limitation.



• The schema document in Fig. 20.6 creates both simple types and complex types.

computer. xsd

```
<?xml version = "1.0"?>
                                                                                          (1 \text{ of } 2)
   <!-- Fig. 20.6: computer.xsd -->
   <!-- W3C XML Schema document -->
   <schema xml ns = "http://www.w3.org/2001/XMLSchema"</pre>
      xml ns: computer = "http://www.deitel.com/computer"
6
7
      targetNamespace = "http://www.deitel.com/computer">
8
      <simpleType name = "gigahertz">
9
                                                                                  Specify the base type as deci mal.
          <restriction base = "decimal"> <</pre>
10
             <minInclusive value = "2.1"/> ←
11
                                                                                  Restrict the value to be at least 2. 1.
12
          </restriction>
13
      </simpleType>
                                                                         Create a simple type using the
14
                                                                         si mpl eType element.
      <complexType name = "CPU">
15
          <si mpl eContent>
16
             <extension base = "string"> 
17
                                                                                 Declare a compl exType named
                                                                                 CPU that has simpleContent and
                                 The extensi on element with
                                                                                 a base type of string.
                                 attribute base sets the base type.
```

Fig. 20.6 | XML Schema document defining simple and complex types. (Part 1 of 2.)



© 2009 Pearson Education, Inc. All rights reserved.

computer. xsd

```
18
                 <attribute name = "model" type = "string"/> ←
                                                                                         (2 of 2)
19
             </extension>
                                                                                 The attri bute element specifies
20
          </simpleContent>
                                                                                 that an element of type CPU and may
21
      </complexType>
                                                                                 contain a model attribute that is of
22
                                                                                 type string.
23
      <complexType name = "portable">
24
          <all>
25
             <element name = "processor" type = "computer: CPU"/>
                                                                                   The element all encloses elements
             <element name = "monitor" type = "int"/>
                                                                                   that must each be included once in
26
                                                                                   the document. These elements can
             <el ement name = "CPUSpeed" type = "computer: gi gahertz"/>
27
                                                                                   be included in any order.
             <element name = "RAM" type = "int"/>
28
29
          </all>
                                                                                portable contains an attribute of type
          <attribute name = "manufacturer" type = "string"/>
30
                                                                                string named manufacturer.
      </complexType>
31
32
                                                                               Declare the actual element, called
      <element name = "laptop" type = "computer: portable"/> 
33
                                                                               laptop, that uses the three types
34 </schema>
                                                                               defined in the schema.
```

Fig. 20.6 | XML Schema document defining simple and complex types. (Part 2 of 2.)



- A document that conforms to a schema is known as an XML instance document.
- Create a simple type using the **simple Type** element.
- Simple types are restrictions of a type typically called a base type.
- The **extensi** on element with attribute **base** sets the base type.
- The element all encloses elements that must each be included once in the document. These elements can be included in any order.

• Figure 20.7 uses the laptop element defined in the computer. xsd schema.

l aptop. xml

```
<?xml version = "1.0"?>
  <!-- Fig. 20.7: laptop.xml
  <!-- Laptop components marked up as XML -->
  <computer: laptop xml ns: computer = "http://www.deitel.com/computer"</pre>
     manufacturer = "IBM">
6
7
8
      cprocessor model = "Centrino">Intel
      <monitor>17</monitor>
9
10
      <CPUSpeed>2. 4</CPUSpeed>
      <RAM>256</RAM>
11
12 </computer:laptop>
```

Fig. 20.7 | XML document using the laptop element defined in computer. xsd.

Automatically Creating Schemas using Visual Studio

- Visual Studio includes a tool that allows you to create a schema from an existing XML document, using the document as a template.
- With an XML document open, select **XML** > **Create Schema** to use this feature.

Good Programming Practice 20.1

The schema generated by Visual Studio is a good starting point, but you should refine the restrictions and types it specifies so they are appropriate for your XML documents.

- Extensible Stylesheet Language (XSL) documents specify how programs are to render XML document data.
- XSL is a group of three technologies—XSL-FO (XSL Formatting Objects), XPath (XML Path Language) and XSLT (XSL Transformations).
 - XSL-FO is a vocabulary for specifying formatting.
 - XPath is a string-based language used by XML to locate structures and data in XML documents.
 - XSL Transformations (XSLT) is a technology for transforming XML documents into other documents.
- XSLT allows you to convert an XML document to an XHTML (Extensible HyperText Markup Language) document for display in a web browser.

- Transforming an XML document using XSLT involves two tree structures—the source tree and the result tree.
- XPath is used to locate parts of the source-tree document that match templates defined in an XSL style sheet.
- When a match occurs, the matching template executes and adds its result to the result tree.
- The XSLT does not analyze every node of the source tree; it selectively navigates the source tree using XSLT's select and match attributes.
- For XSLT to function, the source tree must be properly structured.

<u>Outline</u>

A Simple XSL Example

• Figure 20.8 lists an XML document that describes various sports.

sports. xsl

```
<?xml version = "1.0"?>
                                                                                         (1 \text{ of } 2)
  <?xml-stylesheet type = "text/xsl" href = "sports. xsl"?> 
3
                                                                               This processing instruction (PI)
   <!-- Fig. 20.8: sports.xml -->
                                                                               specifies the location of the XSL style
   <!-- Sports Database -->
                                                                               sheet sports. xsl, which will be
6
                                                                               used to transform the XML document.
   <sports>
      < game id = "783" >
8
          <name>Cri cket</name>
9
10
          <paragraph>
11
             More popular among Commonwealth nations.
12
13
          </paragraph>
14
      </game>
15
      < game id = "239" >
16
          <name>Baseball</name>
17
18
```

Fig. 20.8 | XML document that describes various sports. (Part 1 of 2.)



```
19
          <paragraph>
20
              More popular in America.
          </paragraph>
21
                                                                                              sports. xsl
22
      </game>
23
                                                                                              (2 \text{ of } 2)
      < game id = "418" >
24
25
          <name>Soccer (Futbol)</name>
26
27
          <paragraph>
              Most popular sport in the world.
28
          </paragraph>
29
30
       </game>
31 </sports>
                                                                  _ _ _ X
                        Sports - Windows Internet Explorer
                                  Live Search
                                                        Sports
                                                    Information
                          ID
                                 Sport
                                         More popular among commonwealth nations.
                          783 Cricket
                          239 Baseball
                                         More popular in America.
                         418 Soccer (Futbol) Most popular sport in the world.
                                                                  4 100% ▼
                                  Protected Mode: Off
```

Fig. 20.8 | XML document that describes various sports. (Part 2 of 2.)



- The output shows the result of the transformation (specified in the XSLT template of Fig. 20.9) rendered by Internet Explorer 7.
- To perform transformations, an XSLT processor is required.
- The XML document shown in Fig. 20.8 is transformed into an XHTML document by MSXML when the document is loaded in Internet Explorer.
- MSXML is both an XML parser and an XSLT processor.
- The characters <? and ?> delimit a processing instruction, which consists of a PI target and a PI value.

Software Engineering Observation 20.4

XSL enables document authors to separate data presentation (specified in XSL documents) from data description (specified in XML documents).

• Figure 20.9 shows the XSL document for transforming the structured data of the XML document of Fig. 20.8 into an XHTML document for presentation.

Outline

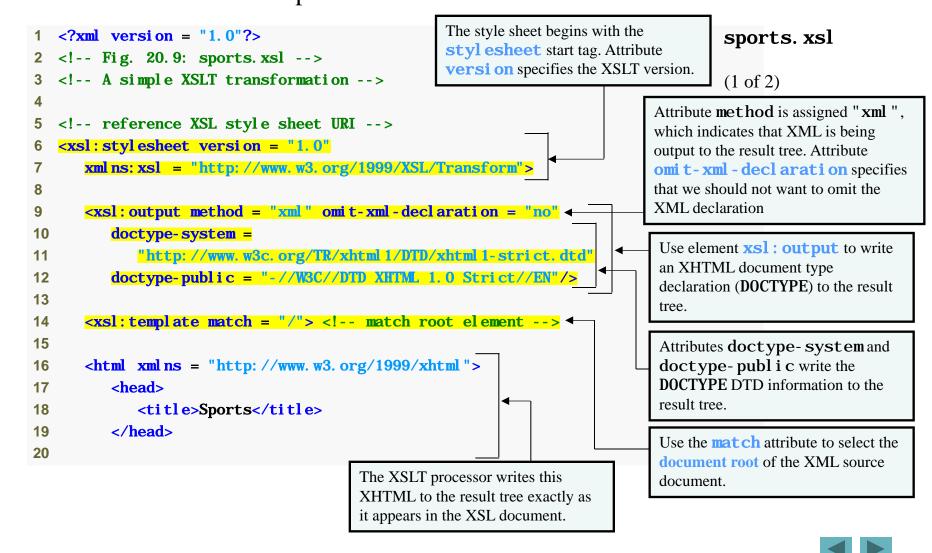


Fig. 20.9 | XSLT that creates elements and attributes in an XHTML document. (Part 1 of 2.)

© 2009 Pearson Education, Inc. All rights reserved.

```
<body>
21
           22
23
             <thead>
                24
                                                                         sports. xsl
                   I D
25
26
                   Sport
                                                                         (2 \text{ of } 2)
                   Information
27
28
                </thead>
29
                                                                      Use element val ue- of to
30
                                                                      retrieve an attribute's value
             <!-- insert each name and paragraph element value -->
31
                                                                      using the XPath symbol @.
             <!-- into a table row. -->
32
33
             <xsl:for-each select = "/sports/game">
34
                <xsl: value- of select = "@id"/> <
                                                                      Element xsl: for-each
35
                   <xsl: value- of select = "name"/>
                                                                      iterates through the source XML
36
                                                                      document, searching for the
37
                   <xsl:value-of select = "paragraph"/>
                                                                      element specified by the
38
                sel ect attribute.
             </xsl:for-each>
39
           40
        </body>
41
     </html>
42
43
     </xsl:template>
44
45 </xsl:stylesheet>
```

Fig. 20.9 | XSLT that creates elements and attributes in an XHTML document. (Part 2 of 2.)



- The style sheet begins with the **stylesheet** start tag. Attribute **version** specifies the XSLT version.
- The **DOCTYPE** identifies XHTML as the type of the resulting document.
- Attributes doctype-system and doctype-public write the DOCTYPE DTD information to the result tree.
- XSLT uses templates (i.e., xsl: template elements) to describe how to transform particular nodes from the source tree to the result tree.
- A template is applied to nodes that are specified in the match attribute.

- The XPath character / (a forward slash) is used as a separator between element names.
- In XPath, a leading forward slash specifies that we are using absolute addressing (i.e., we are starting from the root).
- Element xsl: for-each iterates through the source XML document, searching for the element specified by the select attribute.
- Use element val ue- of to retrieve an attribute's value using the XPath symbol @.
- When an XPath expression has no beginning forward slash, the expression uses **relative addressing**.

Using XSLT to Sort and Format Data

• Figure 20.10 presents an XML document (**sorting. xml**) that marks up information about a book.

sorting. xml

```
(1 \text{ of } 2)
1 <?xml version = "1.0"?>
   <!-- Fig. 20.10: sorting.xml -->
   <!-- XML document containing book information -->
   <?xml - styl esheet type = "text/xsl" href = "sorting. xsl"?>
6
   <book isbn = "999-99999-9-X">
8
      <title>Deitel&apos; s XML Primer</title>
      <author>
10
11
         <firstName>Jane</firstName>
         <lastName>Blue</lastName>
12
      </author>
13
14
15
      <chapters>
```

Fig. 20.10 | XML document containing book information. (Part 1 of 2.)

```
16
         <frontMatter>
                                                                                   sorting. xml
            cpreface pages = "2" />
17
            <contents pages = "5" />
18
                                                                                   (2 \text{ of } 2)
19
            <illustrations pages = "4" />
20
         </frontMatter>
21
         <chapter number = "3" pages = "44">Advanced XML</chapter>
22
23
         <chapter number = "2" pages = "35">Intermediate XML</chapter>
         <appendi x number = "B" pages = "26">Parsers and Tool s</appendi x>
24
         <appendi x number = "A" pages = "7">Entities</appendi x>
25
26
         <chapter number = "1" pages = "28">XML Fundamental s</chapter>
      </chapters>
27
28
      <media type = "CD" />
29
30 </book>
```

Fig. 20.10 | XML document containing book information. (Part 2 of 2.)

• XSL style sheet can sort an XML file's data for presentation purposes.

• Figure 20.11 presents an XSL document (**sorting. xsl**) for transforming **sorting. xml** (Fig. 20.10) to XHTML.

```
sorting. xsl
1 <?xml version = "1.0"?>
  <!-- Fig. 20.11: sorting.xsl -->
                                                                                         (1 \text{ of } 5)
   <!-- Transformation of book information into XHTML -->
4
   <xsl:stylesheet version = "1.0" xmlns = "http://www.w3.org/1999/xhtml"</pre>
      xml ns: xsl = "http://www.w3.org/1999/XSL/Transform">
6
7
      <!-- write XML declaration and DOCTYPE DTD information -->
8
      <xsl: output method = "xml" omit-xml-declaration = "no"</pre>
9
         doctype-system = "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd"
10
11
         doctype-public = "-//W3C//DTD XHTML 1.1//EN"/>
12
13
      <!-- match document root -->
                                                                                  The <xsl: appl y-
      <xsl:template match = "/">
14
                                                                                  templates/> element specifies
         <html>
15
                                                                                  that the XSLT processor is to apply
16
             \langle xsl: appl y-templ ates/ \rangle \leftarrow
                                                                                  the xsl: templates defined in
         </html>
17
                                                                                  this XSL document to the current
18
      </xsl:template>
                                                                                  node's children.
19
```

Fig. 20.11 | XSL document that transforms sorting. xml into XHTML. (Part 1 of 5.)



```
<!-- match book -->
20
21
     <xsl:template match = "book">
         <head>
22
            <title>ISBN <xsl: value-of select = "@isbn"/> -
23
                                                                                sorting. xsl
               <xsl:value-of select = "title"/></title>
24
         </head>
25
                                                                                 (2 \text{ of } 5)
26
                                                                              Use the book's ISBN (from
         <body>
27
                                                                              attribute i sbn) and the contents
            <h1 style = "color: blue"><xsl:value-of select = "title"/></h1>
28
                                                                              of element title to create the
            <h2 style = "color: blue">by
29
                                                                              string that appears in the
30
               <xsl:value-of select = "author/firstName"/>
                                                                              browser window's title bar.
31
               <xsl:text> </xsl:text> <
32
               <xsl:value-of select = "author/lastName"/>
                                                                              The xsl: text element is used
                                                                              to insert literal text.
33
            </h2>
34
35
            36
                                                                              Select each element (indicated
               <xsl:for-each select = "chapters/frontMatter/*">
                                                                              by an asterisk) that is a child of
37
                                                                              element front Matter.
                  38
                     39
                                                                              Node-set function name
                        <xsl:value-of select = "name()"/> <</pre>
40
                                                                              retrieves the current node's
41
                     element name.
```

Fig. 20.11 | XSL document that transforms sorting. xml into XHTML. (Part 2 of 5.)



```
42
43
                     ( <xsl: value-of select = "@pages"/> pages )
44
                                                                                sorting. xsl
                     45
                  46
                                                                                (3 \text{ of } 5)
               </xsl:for-each>
47
48
               <xsl:for-each select = "chapters/chapter">
49
                                                                              Use element xsl: sort to sort
                  <xsl:sort select = "@number" data-type = "number'</pre>
50
                                                                              the selected elements by the
51
                       order = "ascending"/>
                                                                              value given by its select
52
                  attribute.
                     53
54
                        Chapter <xsl: value-of select = "@number"/>
55
                     56
57
                     Use node-set function text to
                        <xsl:value-of select = "text()"/> 
58
                                                                              obtain the text between the
59
                        ( <xsl: value-of select = "@pages"/> pages )
                                                                              chapter start and end tags.
                     60
                  61
               </xsl:for-each>
62
```

Fig. 20.11 | XSL document that transforms sorting. xml into XHTML. (Part 3 of 5.)



```
63
64
             <xsl: for-each select = "chapters/appendix">
                                                                       sorting. xsl
65
                <xsl:sort select = "@number" data-type = "text"</pre>
                  order = "ascending"/>
66
                                                                       (4 \text{ of } 5)
67
                68
69
                     Appendix <xsl: value-of select = "@number"/>
                  70
71
72
                  <xsl:value-of select = "text()"/>
73
74
                     ( <xsl: value- of select = "@pages"/> pages )
75
                   76
             </xsl:for-each>
77
          78
79
```

Fig. 20.11 | XSL document that transforms sorting. xml into XHTML. (Part 4 of 5.)

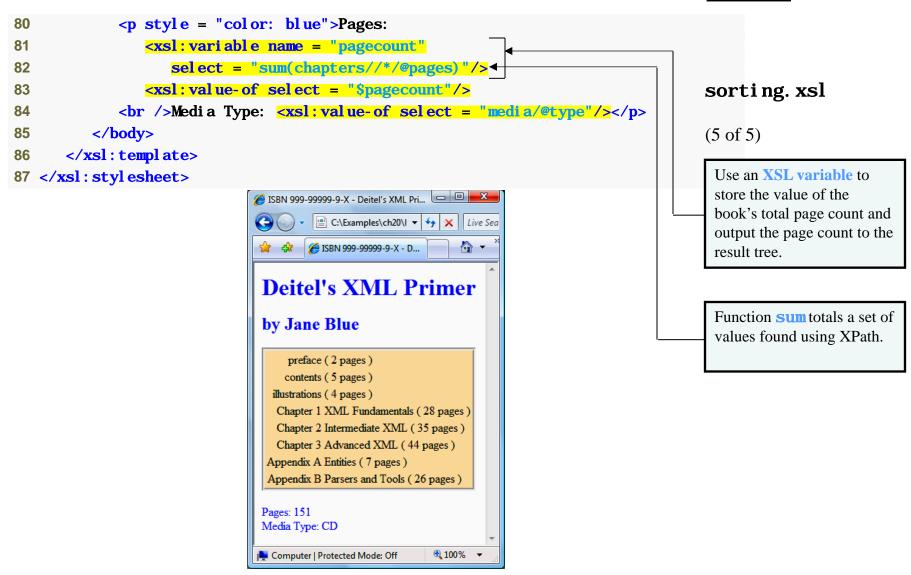


Fig. 20.11 | XSL document that transforms sorting. xml into XHTML.



- The <xsl: apply-templates/> element specifies that the XSLT processor is to apply the xsl: templates defined in this XSL document to the current node's children.
- The xsl: text element is used to insert literal text.
- Node-set function name retrieves the current node's element name.
- Use element xsl: sort to sort the selected elements by the value given by its select attribute.
- Attribute data-type, with value "number", specifies a numeric sort. This attribute also accepts the value "text".

- Attribute order can be set to either "ascending" or "descending".
- XSL variables cannot be modified after they are initialized.
 - Attribute **name** specifies the variable's name
 - Attribute sel ect assigns a value to the variable.
- Function **sum** totals a set of values found using XPath.
- Two slashes in an XPath expression indicate recursive descent.

Performance Tip 20.1

Selecting all nodes in a document when it is not necessary slows XSLT processing.



Summary of XSL Style-Sheet Elements

• Figure 20.12 lists commonly used XSL elements.

Element	Description
<pre><xsl: appl="" ates="" templ="" y-=""></xsl:></pre>	Applies the templates of the XSL document to the children of the current node.
<pre><xsl: appl="" ates="" match="expression" templ="" y-=""></xsl:></pre>	Applies the templates of the XSL document to the children of the nodes matching <i>expression</i> . The value of the attribute match (i.e., <i>expression</i>) must be an XPath expression that specifies elements.
<xsl:template></xsl:template>	Contains rules to apply when a specified node is matched.
<pre><xsl:value-of select="expression"></xsl:value-of></pre>	Selects the value of an XML element and adds it to the output tree of the transformation. The required select attribute contains an XPath expression.

Fig. 20.12 | XSL style-sheet elements. (Part 1 of 2.)



Element	Description
<pre><xsl: for-each="" select="expression"></xsl:></pre>	Applies a template to every node selected by the XPath specified by the select attribute.
<pre><xsl:sort select="expression"></xsl:sort></pre>	Used as a child element of an <xsl: appl="" y-<br="">templ ates> or <xsl: for-each=""> element. Sorts the nodes selected by the <xsl: appl="" ate="" y-templ=""> or <xsl: for-each=""> element so that the nodes are processed in sorted order.</xsl:></xsl:></xsl:></xsl:>
<xsl:output></xsl:output>	Has various attributes to define the format (e.g., XML, XHTML), version (e.g., 1.0, 2.0), document type and MIME type of the output document. MIME types are discussed in Section 22.2. This tag is a top-level element—it can be used only as a child element of an xsl: styl esheet.
<xsl :="" copy=""></xsl>	Adds the current node to the output tree.

Fig. 20.12 | XSL style-sheet elements. (Part 2 of 2.)

