1.264 Lecture 21

XML

Quiz 1 review 7pm on either Wednesday next week.
Download and unzip Lecture21Download. Put in Lecture21 Web site.
Next class: ASP.NET book chapter 10, pages 337-348: Web services only.
Exercises due after class

What is XML?

- Extensible Markup Language (XML) is:
  - a World Wide Web Consortium (W3C) standard for
  - a file format to
  - easily and cheaply distribute electronic documents on the World Wide Web
  - extensible, not frozen like HTML
  - supporting rich structure, like objects or hierarchies or relationships
  - supporting validation and well-formed properties
  - avoiding applets, scripts, plug-ins, etc.
  - separating form (how it looks) from content (what it is)

XML Concepts

- XML is self-describing and can be validated:
  - XML document contains the business rules to which its data must conform
  - Rules can be reused in other documents: documents can be more specialized types (inheritance) of a base type
- XML applications
  - Data interchange format between computers
    - Using Web server as data channel between databases
    - Automated processing of documents exchanged
  - Common format for Web, electronic, paper documents, ...
    - XML as a general markup language
    - XML used for manuals, CDs, help and other text documents
    - Handled by standard browsers (IE, Firefox, Chrome, …)
  - Remote procedure call/invocation protocol
    - Executes Web services or processes on other computers

XML Document Structure

- HTML:
  - Head
  - Body
    - Tags are predefined in HTML (or XHTML) standard
    - Tags describe how the page should be displayed
- XML:
  - Prolog
    - XML declaration (defines version)
    - Document type declaration (defines tags): DTD or XSD
  - Body
    - Tags describe data elements, not how to display them
    - Tags are defined in DTD or XSchema document, which anyone can create

Example 1: XML file

```xml
<?xml version='1.0'?>
<!-- The 1.264 reading list -->
<booklist>
    <book topic="software" publicationdate="1996" ISBN="1-55615-900-5">
        <title>Rapid Development</title>
        <author>
            <first-name>Steve</first-name>
            <last-name>McConnell</last-name>
        </author>
        <author>
            <first-name>Bryan</first-name>
            <last-name>Syverson</last-name>
        </author>
        <price>48.99</price>
    </book>
    <book topic="UML" publicationdate="2004" ISBN="0-321-19368-7">
        <title>UML Distilled</title>
        <author>
            <first-name>Martin</first-name>
            <last-name>Fowler</last-name>
        </author>
        <price>34.99</price>
    </book>
</booklist>
```

Books.xml

Exercise 1

- Start VSW
- Create a new Web site in VSW
  - File- New Web Site -> Lecture21 (or other name)
- Download the lecture 21 files from the Web site
  - Move/copy them to the Web site you just created
- In this Web site, write an XML file that describes the 1.264 class, two groups in the class, and the students in the groups
  - File-> New File -> XML File in VSW
  - Create appropriate elements and/or attributes for the class, groups and students
- When done, display it in your browser

Solution 1- XML

```
<classlist class=“1264”>
    <group number=“1”>
        <groupname>Rapid Development</groupname>
        <grouplist>
            <name>Steve Bell</name>
            <name>Jo McDuff</name>
        </grouplist>
    </group>
    <group number=“2”>
        <groupname>UML Forever</groupname>
        <grouplist>
            <name>Steve Smith</name>
            <name>Jo Brown</name>
        </grouplist>
    </group>
</classlist>
```

XML Languages

- Two ways to express business rules/validate:
  - DTD: Document type declaration: business rules
    - Not XML itself (oddly), being replaced by XML Schema
    - You should understand how to write DTDs in 1.264
  - XML Schema or XSD: business rules
    - XSD is XML subset, verbose but powerful
    - You don’t need to understand the details of XSD in 1.264
  - There are user-friendly tools to manage these
    - This would be one more download for you; we will just look at them manually
- Language to format and transform XML document
  - XSLT: extensible stylesheet language/transformation
    - Transforms XML documents
    - Formats XML documents
    - XPath (or XQuery) sublanguage for queries
- And many others for specialized applications
  - XLink, XPointer, RDF, XForms, … see www.w3c.org

Validation 1: XSchema Definition (XSD)

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema  (details omitted) >
  <xs:element name="booklist">
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" name="book">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="title" type="xs:string" />
              <xs:element name="author“ maxOccurs=“unbounded”>
                <xs:complexType>
                  <xs:sequence>
                    <xs:element name="first-name" type="xs:string" />
                    <xs:element name="last-name" type="xs:string" />
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
              <xs:element name="price" type="xs:decimal" />
            </xs:sequence>
            <xs:attribute name="topic" type="xs:string" use="required" />
            <xs:attribute name="publicationdate" type="xs:unsignedShort" use="required"/>
            <xs:attribute name="ISBN" type="xs:string" use="required" />
          </xs:complexType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Books.xsd

Document Type Declaration (DTD)

- DOCTYPE: class (type) of document
  - Placed in XML file, refers to DTD file to be used to validate
- ELEMENT: object in document
  - Either all valid values are given in a list in (), or
  - The element is defined later in the DTD file
  - Symbols: +: 1 or more, *: 0 or more, ?:0 or 1, none: exactly 1
- ATTLIST: valid attribute list for element
  - #CDATA: character data
  - #PCDATA: parsed character data (can't have < > &…)
  - #REQUIRED: element must be present
  - #IMPLIED: element optional, no default value
  - #FIXED: attribute value is fixed
- ENTITY: a constant value
- | means OR

Validation 2: DTD file

```
<?xml version="1.0" encoding="iso-8859-1"?>

<!ELEMENT booklist (book+)>
<!ELEMENT book (title, author+, price)>
<!ATTLIST book
    topic (software|UML|database) "UML"
    publicationdate CDATA #IMPLIED
    ISBN CDATA #IMPLIED>

<!ELEMENT title (#PCDATA)>
<!ELEMENT author (first-name, last-name)>
<!ELEMENT first-name (#PCDATA)>
<!ELEMENT last-name (#PCDATA)>
<!ELEMENT price (#PCDATA)>

<!ENTITY currentyear "2013">
```

Books.dtd

Exercise 2: DTD

- Write a DTD file that states these business rules:
  - The 1.264 class has one to many groups
  - A group has a name, a number, and has one to two (many) students
  - The DTD file specifies the elements and attributes present in the XML file
  - Use the exercise 1 XML file as the basis for your DTD
- There is no DTD file type in VSW any more
  - Create an html file, rename it to Classlist.dtd or similar, go to Source view and delete all the html so you have just a blank file

Show Exercise 1 solution

Solution 2- DTD

```
<?xml version="1.0" encoding="iso-8859-1"?>

<!ELEMENT classlist (group+)>
<!ATTLIST classlist
    class (1264|ESD264) “1264“>
<!ELEMENT group (groupname, grouplist)>
<!ATTLIST group
    number1 CDATA #REQUIRED>
<!ELEMENT grouplist (name+)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT groupname (#PCDATA)>
```

Order of statements may vary

XML Summary

- XML documents can be:
  - Defined by anyone: tags and business rules
  - Sent and received by databases using SQL and HTTP
  - Validated by DTD or XSD files
  - Transformed and styled by XSLT files
  - Placed on a server for clients to attach to, such as blogs
- XML files are little pieces of a database that can be shared. Typically they represent:
  - Rows from a single table, or
  - Rows from two tables in a many-to-one relationship
  - Any arbitrary set of tables/relationships can be sent

Glossary

- W3C: World Wide Web Consortium (standards)
- DTD: Document Type Declaration (XML validation)
- XML Schema or XSD: XML validation
- XSLT: XML Stylesheet Language/Tranformation
  - Modifies or queries XML documents
- XPath: XSLT query sublanguage
- XQuery: XSLT query sublanguage

1.264J / ESD.264J Database, Internet, and Systems Integration Technologies
Fall 2013