

# Introduction to XML

---

Qiang Yang

Thanks: Ethan Cerami, New York University, Ian Graham, Univ of Toronto, Francis Lau, HK Univ.

1/30/01 Introduction to XML 1

## Road Map

- What is XML?
  - A Brief Overview
  - Origins of XML
- Creating XML Documents
  - Basic Rules
- Example XML Documents
- Case Studies

1/30/01 Introduction to XML 2

## Tim Berners Lee: the Semantic Web

“In communicating between people using the Web, computers and networks have as their job to enable the information space, and otherwise get out of the way. But doesn't it make sense to also bring computers more into the action? – part two of the dream!

“... This creates what I call a Semantic Web – a web of data that can be processed directly or indirectly by machines.”

<<Weaving the Web>> pp 177.

1/30/01 Introduction to XML 3

## What is XML?

- XML: eXtensible Markup Language**
- Subset of SGML, but more efficient and lightweight**
- XML:**
  - Extensible:** tags can be defined; can be extended to lots of different applications.
  - Markup language:** language used to mark up data.
  - Meta Language:** Language used to create other languages.

1/30/01 Introduction to XML 4

## XML v. HTML

- contrast with HTML: XML is Extensible:
  - HTML:** restricted set of tags, e.g. <TABLE>, <H1>, <B>, etc.
  - XML:** you can create your own tags
- Example: Put a library catalog on the web.
  - HTML:** You are stuck with regular HTML tags, e.g. H1, H3, etc.
  - XML:** You can create your own set of tags: TITLE, AUTHOR, DATE, PUBLISHER, etc.

1/30/01 Introduction to XML 5

## Book Catalog in HTML

```

<HTML>
<BODY>
<H1>Harry Potter</H1>
<H2>J. K. Rowling</H2>
<H3>1999</H3>
<H3>Scholastic</H3>
</BODY>
</HTML>

```

HTML conveys the “look and feel” of your page.

As a human, it is easy to pick out the publisher.

But, how would a computer pick out the publisher?

Answer: XML

1/30/01 Introduction to XML 6

## Book Catalog in XML

```
<BOOK>
  <TITLE>Harry Potter</TITLE>
  <AUTHOR>J. K. Rowling</AUTHOR>
  <DATE>1999</DATE>
  <PUBLISHER>Scholastic</PUBLISHER>
</BOOK>
```

Look at the new tags!  
A Human and a computer can now easily  
extract the publisher data.

1/30/01

Introduction to XML

7

## XML v. HTML

- General Structure:
  - Both have Start tags and end tags.
- Tag Sets:
  - HTML has set tags
  - XML lets you create your own tags.
- General Purposes:
  - HTML focuses on structure and presentation -- "look and feel"
  - XML focuses on the structure of the data, not presentation
- XML is *not* meant to be a replacement for HTML. In fact, they are usually used together.

1/30/01

Introduction to XML

8

## Origins of XML

1/30/01

Introduction to XML

9

## Origins of XML

- XML is based on SGML: Standard Generalized Markup Language
- SGML:
  - Developed in the 1970s
  - Used by big organizations: IRS, IBM, Department of Defense
  - Focuses on content structure, not look and feel
  - Good for creating catalogs, manuals.
  - Very complex

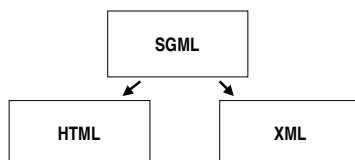
1/30/01

Introduction to XML

10

## Origins of XML

- XML: SGML-Lite: 20% of SGML's complexity, 80% of its capacity.
- HTML and XML are both based on SGML.



1/30/01

Introduction to XML

11

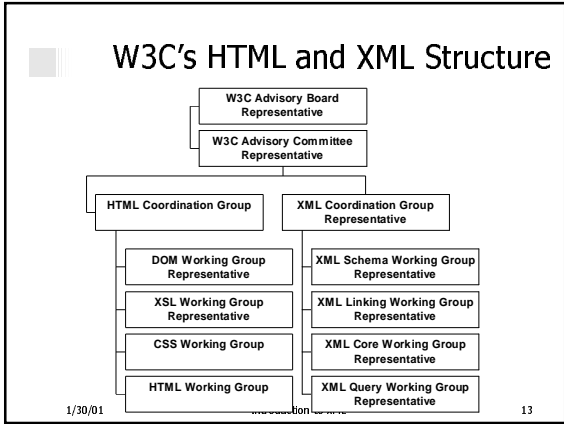
## XML and the W3C

- XML is an official standard of the World Wide Web Consortium (W3C)
- The Official Version is 1.0
- Official information is available at:
  - <http://www.w3.org/XML/>
- The Official spec is available at:
  - <http://www.w3.org/TR/1998/REC-xml-19980210>
- The Official XML FAQ:
  - <http://www.ucc.ie/xml/>
- W3C sponsors many projects which seek to enhance and improve on XML.

1/30/01

Introduction to XML

12



## Creating XML Documents Basic Rules

---

1/30/01 Introduction to XML 14

## Basic Definitions

- **Tag:** a piece of markup
  - Example: <P>, <H1>, <TABLE>, etc.
- **Element:** a start and an end tag
  - Example: <H1>Hello</H1>
- **HTML Code:**
  - <P>This is a <B>sample</B> paragraph.
  - This code contains:
    - 3 tags, <P>, <B>, and </B>
  - However, it only contains one element: <B>...</B>

1/30/01 Introduction to XML 15

## Rule 1: Well-Formedness

- XML is much more strict than HTML.
- XML requires that documents be well-formed:
  - every start tag must have an end tag
  - all tags must be properly nested.
- **XML Code:**
  - <P>This is a <B>sample</B> paragraph.</P>

Note the end </P>

1/30/01 Introduction to XML 16

## Rule 1: Well-Formedness

- **Another HTML Example:**
  - <b><i>This text is bold and italic</i></b></i>
- This will render in a browser, but contains a nesting error.
- **XML Code (with proper nesting)**
  - <b><i>This text is bold and italic</i></b>

1/30/01 Introduction to XML 17

## Rule 2: XML is Case Sensitive

- XML is Case Sensitive.
- HTML is not.
- The following is valid in HTML:
  - <H1>Hello World</h1>
- This will not work in XML. Would result in a well-formedness error:
  - H1 does not have a matching end H1 tag.

1/30/01 Introduction to XML 18

### Rule 3: Attributes must be quoted.

- In HTML you can get away with doing the following:
  - <FONT FACE=ARIAL SIZE=2>
- In XML, you must put quotes around all your attributes:
  - <BOOK ID="894329">Harry Potter</BOOK>

1/30/01

Introduction to XML

19

### Examples

- To get a feel for XML, let's take a look at several examples:
  - An XML Memo
  - CD Catalog
  - Plant Catalog
  - Restaurant Menu

1/30/01

Introduction to XML

20

### Example 1: A Memo

```
<?xml version="1.0" encoding="ISO8859-1" ?>
<note>
  <to>Class 470/882 </to>
  <from>Qiang Yang</from>
  <heading>Introduction</heading>
  <body>This is an XML document!</body>
</note>
```

This XML Note could be part of a message board application.

1/30/01

Introduction to XML

21

### Example 2: CD Collection

```
<?xml version="1.0" encoding="ISO8859-1" ?>
<CATALOG>
  <CD>
    <TITLE>Eyes on Me</TITLE>
    <ARTIST>Faye Wong</ARTIST>
    <COUNTRY>Hong Kong</COUNTRY>
    <COMPANY>EMI</COMPANY>
    <PRICE>10.90</PRICE>
    <YEAR>1998</YEAR>
  </CD>
```

Continued...

1/30/01

Introduction to XML

22

```
<CD>
  <TITLE>Hide your heart</TITLE>
  <ARTIST>Bonnie Tylor</ARTIST>
  <COUNTRY>UK</COUNTRY>
  <COMPANY>CBS Records</COMPANY>
  <PRICE>9.90</PRICE>
  <YEAR>1988</YEAR>
</CD>
<CD>
  <TITLE>Unchain my heart</TITLE>
  <ARTIST>Joe Cocker</ARTIST>
  <COUNTRY>USA</COUNTRY>
  <COMPANY>EMI</COMPANY>
  <PRICE>8.20</PRICE>
  <YEAR>1987</YEAR>
</CD>
</CATALOG>
```

Note that indentation helps you follow the flow of the document, Although they are not required

1/30/01

Introduction to XML

23

### Example 3: A Plant Catalog

```
<?xml version="1.0" encoding="ISO8859-1" ?>
<CATALOG>
  <PLANT>
    <COMMON>Bloodroot</COMMON>
    <BOTANICAL>Sanguinaria canadensis</BOTANICAL>
    <ZONE>4</ZONE>
    <LIGHT>Mostly Shady</LIGHT>
    <PRICE>$2.44</PRICE>
    <AVAILABILITY>031599</AVAILABILITY>
  </PLANT>
```

Continued...

1/30/01

Introduction to XML

24

```

<PLANT>
<COMMON>Columbine</COMMON>
<BOTANICAL>Aquilegia canadensis</BOTANICAL>
<ZONE>3</ZONE>
<LIGHT>Mostly Shady</LIGHT>
<PRICE>$9.37</PRICE>
<AVAILABILITY>030699</AVAILABILITY>
</PLANT>

```

```

<PLANT>
<COMMON>Marsh Marigold</COMMON>
<BOTANICAL>Caltha palustris</BOTANICAL>
<ZONE>4</ZONE>
<LIGHT>Mostly Sunny</LIGHT>
<PRICE>$6.81</PRICE>
<AVAILABILITY>051799</AVAILABILITY>
</PLANT>
</CATALOG>

```

1/30/01

Introduction to XML

25

## Example 4: Restaurant Menu

```

<?xml version="1.0" encoding="ISO8859-1" ?>
<breakfast-menu>
<food>
  <name>Belgian Waffles</name>
  <price>$5.95</price>
  <description>two of our famous Belgian Waffles with plenty
of real maple syrup</description>
  <calories>650</calories>
</food>

```

Continued...

1/30/01

Introduction to XML

26

```

<food>
  <name>Strawberry Belgian Waffles</name>
  <price>$7.95</price>
  <description>light Belgian waffles covered with
strawberrys and whipped cream
</description>
  <calories>900</calories>
</food>
<food>
  <name>Berry-Berry Belgian Waffles</name>
  <price>$8.95</price>
  <description>light Belgian waffles covered with
an assortment of fresh berries and
whipped cream
</description>
  <calories>900</calories>
</food>

```

Continued...

1/30/01

Introduction to XML

27

```

<food>
  <name>French Toast</name>
  <price>$4.50</price>
  <description>thick slices made
from our homemade sourdough bread
</description>
  <calories>600</calories>
</food>
<food>
  <name>Homestyle Breakfast</name>
  <price>$6.95</price>
  <description>two eggs, bacon or sausage, toast, and our
ever-popular hash browns</description>
  <calories>950</calories>
</food>
</breakfast-menu>

```

1/30/01

Introduction to XML

28

## XML: what do you want to do?

- Parsing
  - To check well formedness
- Porting to your own Databases
  - Need to know the document meaning and structure (DOM)
- Printing for viewing by humans
  - Style sheet and HTML transformation
- Query answering

1/30/01

Introduction to XML

29

## XML Tools

- XML is a family of technologies
  - **Xlink** for adding hyperlinks to an XML file
  - **XPointer** for pointing to parts of an XML document
  - **CSS** is applicable to XML as it is to HTML
  - **XSL** is an advanced language for expressing style sheets
    - **XSLT** for transforming XML to other formats
  - **DOM** for manipulating XML (and HTML) file from a programming language
  - **Namespaces** for differentiating elements of different XML documents
  - **Schemas** for developers to precisely define their own XML-based formats

1/30/01

Introduction to XML

30

## ■ Displaying XML

- More complicated than HTML
  - XML represents data only, **not** how it looks
  - Need **extra instructions** (a “style sheet” document) to define how things should look

1/30/01

Introduction to XML

31

## ■ What Browsers Do Now?

- Navigator 4, Internet Explorer 4
  - can't handle XML at all
- Internet Explorer 5 -- shows a **tree** of elements
  - See examples
- Mozilla/Netscape 6 -- **ignores** the tags

1/30/01

Introduction to XML

32

## ■ Case Studies

1/30/01

Introduction to XML

33

## ■ Applications of XML

- Applications: Business to Business !!
  - Search Engines
    - that query other search engines
    - They need to have understandable format
  - News Distribution
    - To multiple news portals
  - E-Commerce
    - Business transactions
    - Company merger: how to merge their data?
    - Comparison shopping

1/30/01

Introduction to XML

34

## ■ Case Study 1: Web Search

- Scenario:
  - You want to offer a meta-web search functionality for your site.
  - You want control over the look and feel of the search results.
  - You do not want to support your own database of millions of web sites.

1/30/01

Introduction to XML

35

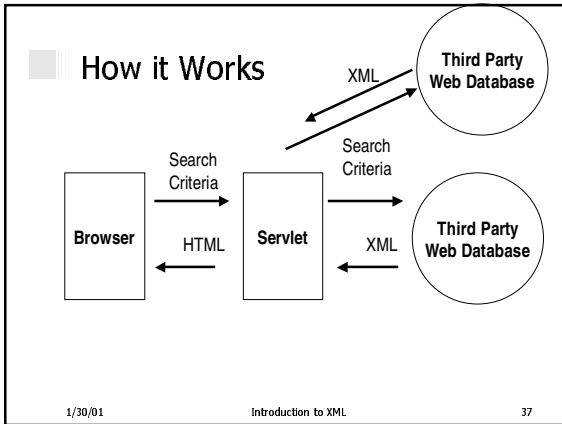
## ■ Case Study 1: Web Search

- XML to the Rescue...
  - Several companies provide XML Access to their Web Search Databases.
  - For example:
    - Open a network connection and send search criteria.
    - Third Party returns results in XML.

1/30/01

Introduction to XML

36



### Case Study 2: Price Comparison

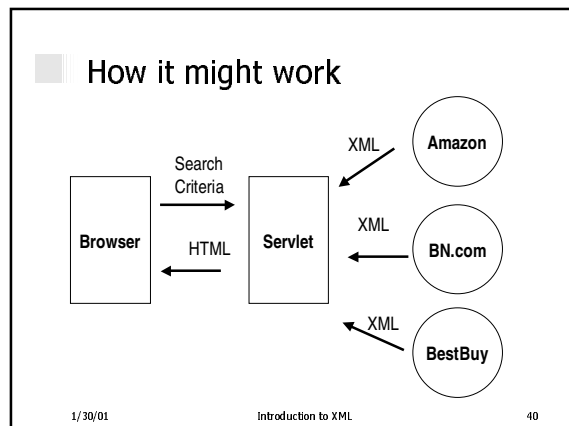
- Scenario:
  - You want to create a site that compares prices of books.
  - For example, a user enters a book title, and your page displays the price at bn.com, amazon.com, bestbuy.com, etc.
  - User can choose the cheapest price.

1/30/01 Introduction to XML 38

### How it might work

- How it works
  - User sends book title
  - Servlet makes three concurrent connections and queries the bookstores:
    - Amazon, bn.com, bestbuy.com
  - Each Bookstore returns results in a standard XML.
  - Servlet parses XML and creates a small price comparison table.

1/30/01 Introduction to XML 39



### A Hypothetical Example of B2B Application

- Wing runs a music wholesaling business
  - He buys CDs from publishers using EDI or by fax
  - He sells CDs to shops, taking orders by mail, phone, fax, or over the Web
- His ordering using EDI is actually worse than fax (Big business -- the record company -- benefits, the small guy Wing.com suffers)
  - His suppliers all use different EDI standards
  - Arthur has to use four PCs, one for each supplier, running some expensive software to produce EDI orders and accept EDI invoices
- Worse, none of systems links to his accounting system

1/30/01 Introduction to XML 41

### EDI

- EDI (Electronic Data Interchange)
  - To eliminate the use of paper for business data exchange
  - Single point of information capture, electronic delivery, low storage and retrieval costs
- Statistics show that only the top 10,000 companies on a global scale are using EDI
- The rest of the business world: only 5% using EDI, all others, paper
- The success of any new way of data exchange must depend on adoption by SMEs

1/30/01 Introduction to XML 42

## What's Wrong with EDI?

- Information coded in EDI might look like any of
  - "Wing Discspinner Music", "Music distributor", "Wing Discspinner", Wing@discspinner.co.uk
- This is not self-describing!
- Imagine what happens when adding a new field
  - "Arther Discspinner Music", "Music distributor", "Arthur Discspinner", "0118 912 3456", Arthur@discspinner.co.uk
- In general, the EDI system will report an error
- EDI systems must be 100% compatible in the message structures they understand
- So companies need to band together to define standards

1/30/01

Introduction to XML

43

## With XML

- Using XML, the same info will be coded as

```
<Company>Wing Discspinner Music</Company>
<MarketSector>Music distributor</MarketSector>
<Contact>
  <Name>Wing Discspinner</Name>
  <Phone>0118 912 3456</Phone>
  <E mail>wing@discspinner.co.uk</E mail>
</Contact>
```

- The software will access the data by element name

1/30/01

Introduction to XML

44

## New Vocabularies for E-Business

- What Wing is/we are looking for
  - A system that can link accounting systems over the Web or by email
    - A "many-to-many" solution
    - "Flexible interoperability"
- XML can achieve all this
- To use XML to define vocabularies for business relationships and transactions
- An example: [ebXML](http://www.ebxml.org/) (http://www.ebxml.org/)

1/30/01

Introduction to XML

45

## RDF

- The Resource Description Framework (RDF) is a specification currently under development within the W3C Metadata activity
- Metadata** is "data about data" or "data describing web resources"
- With RDF, W3C hopes to elevate the status of the web from machine-readable to machine-understandable
- RDF uses XML as the syntax in order to leverage other tools and code bases being built around XML

1/30/01

Introduction to XML

46

## RDF (cont'd)

- A quote from a W3C note:

"XML is becoming increasingly adopted as a common syntax for expressing structure in data. Now the *Resource Description Framework (RDF)*, a layer on top of XML, provides a common basis for expressing semantics. Applications which allow programs to combine data logically will be built using RDF (and therefore XML) and this will enhance the modularity and extensibility of the Web. This is essential to its rapid future growth, multiplying together the strengths of new, independently developed, applications."

1/30/01

Introduction to XML

47

## RDF (cont'd)

- RDF metadata can be used in a variety of application areas such as:
  - in *resource discovery* to provide better search engine capabilities
  - in *cataloging* for describing the content and content relationships available at a particular Web site, page, or digital library
  - by *intelligent software agents* to facilitate knowledge sharing and exchange
  - in *content rating* for child protection and privacy protection
  - in describing *collections* of pages that represent a single logical "document"
  - for describing *intellectual property rights* of Web pages

1/30/01

Introduction to XML

48

## RDF (cont'd)

- An example of RDF using the Dublin Core language

```
<? xml version="1.0" ?>
<RDF xmlns = http://w3.org/TR/1999/PR-rdf-syntax-19990105#
xmlns:DC = "http://purl.org/DC#" >
  <Description about = "http://dstc.com.au/report.html" >
    <DC:Title> The Future of Metadata </DC:Title>
    <DC:Creator> Jacky Crystal </DC:Creator>
    <DC:Date> 1998-01-01 </DC:Date>
    <DC:Subject> Metadata, RDF, Dublin Core </DC:Subject>
  </Description>
</RDF>
```

1/30/01

Introduction to XML

49

## SMIL

- The Synchronized Multimedia Integration Language (SMIL, pronounced "smile")
- A W3C recommendation
- To enable simple authoring of TV-like multimedia presentations such as training courses on the Web
- An XML language
  - Thus, SMIL presentations can be written using a simple text-editor
- A SMIL presentation can be composed of streaming audio, streaming video, images, text or any other media type

1/30/01

Introduction to XML

50

## SMIL (cont'd)

- You can also
  - Control the timeline of a document
  - Describe the layout of a document
- Structure of an SMIL document:

```
<smil>
<head>
... head of the document
</head>
<body>
... body of the document
</body>
</smil>
```

1/30/01

51

## SMIL (cont'd)

- Neither Netscape and Microsoft currently support SMIL in their browsers
- Have to use a "player", such as RealPlayer

```
<smil>
<head>
<layout>
  <root-layout background-color="black" width="252" height="168"/>
  <region id="picture" z-index="1" left="0" top="0"
width="252" height="168"/>
</layout>
</head>
<body>
.....
```

1/30/01

Introduction to XML

52

## SMIL (cont'd)

```
<par>
  <audio id="a" src="ms.wav" repeat="5" />
  <seq repeat="5" >
    
    
    
    
    
    
    
    
  </seq>
</par>
</body>
</smil>
```

Try it

(<http://www.csis.hku.hk/~fcmlau/flower.smil>)

1/30/01

Introduction to XML

53

## SMIL (cont'd)

- Examples/demos
  - "What I did last summer..." by Larry Bouthillier; he also supplies the source code (<http://www.people.hbs.edu/lbouthillier/smil>)
  - "Star Wars- The Phantom Menace" from LucasFilms ([http://www.broadcast.com/g2/video/Starwars\\_broadband.ram](http://www.broadcast.com/g2/video/Starwars_broadband.ram))
  - "Filming for the Internet" by the Real Networks Media Lab (<http://ramhurl.real.com/smildemohurl.html?file=tools/producerpro/demo1.smil>)

1/30/01

Introduction to XML

54