

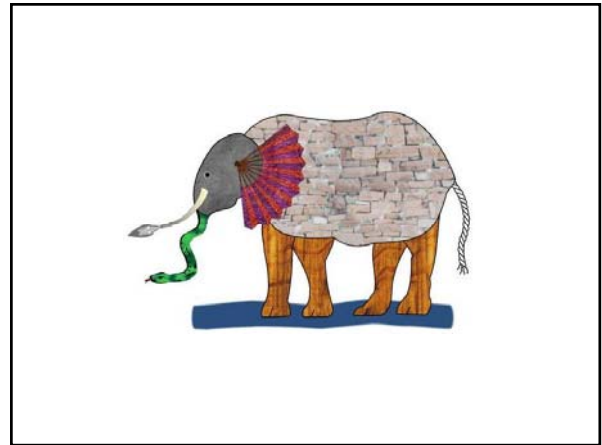
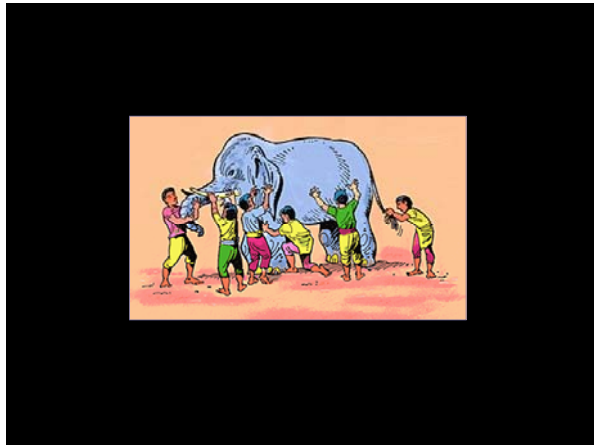
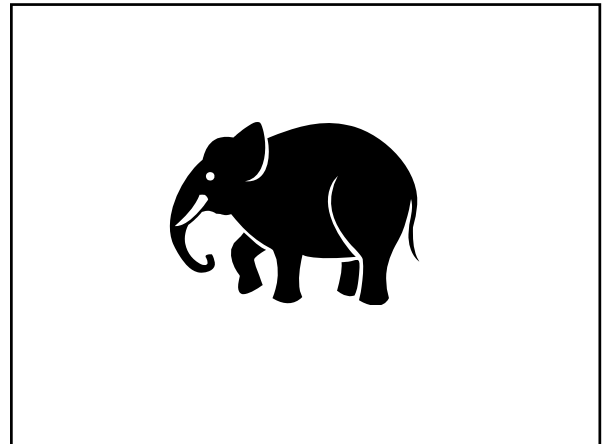
LBSC 690 Session #5
Metadata and XML

Jimmy Lin
The iSchool
University of Maryland

Wednesday, October 1, 2008



This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 United States
See <http://creativecommons.org/licenses/by-nc-sa/3.0/us/> for details



Take-Away Messages

- Metadata makes data useful
- XML is a way to encode data and metadata
- XML allows computers to exchange information in new and interesting ways

The iSchool
University of Maryland

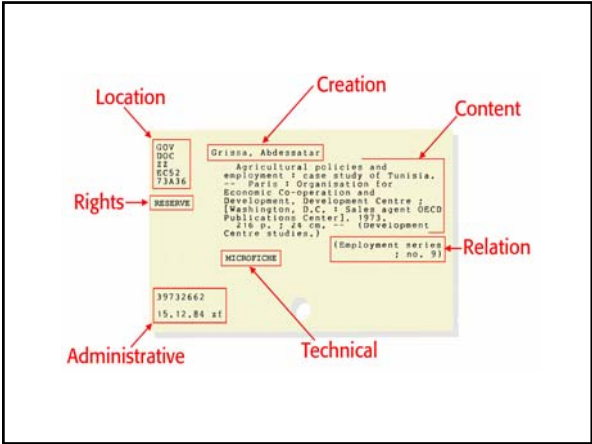


7/1/1999	OL	990	293	13	0.8	-0.1	35.1	27.0	6.3	5.92
7/2/1999	OL	990	242	12.8	1	-0.1	27.8	23.9	3.8	4.96
7/3/1999	OL									
7/4/1999	OL	990	0.4	18.3	0.4	0.2	41	34.8	6.5	15.5
7/5/1999	OL	1005	239	18.9	1.4	0.3	28.9	23.7	6.1	14.22
7/6/1999	OL	1020	323	20.5	1.4	0.3	23.4	18.9	4.5	12.97
7/7/1999	OL	1010	369	24.9	1.7	0.5	18.9	15.9	3.2	13.92
7/8/1999	OL	910	428	22.8	2.5	0.8	20.7	15.9	2.9	15.15
7/9/1999	OL	945	233	24.9	1.7	0.5	27.7	23.9	4.3	12.35
7/10/1999	OL	1030	493	28.2	2.5	0.8	40.3	34	6.3	22.41
7/11/1999	OL	940	448	25.2	2.5	0.8	34	29.2	4.8	16.76
7/12/1999	OL	1010	474	26.9	2.8	0.9	47.2	39.8	7.7	16.13
7/13/1999	OL	948	365	25.2	2.5	0.8	38.7	32.6	4	15.5
7/14/1999	OL	950	191	17.8	1.6	0.4	30.9	26.1	24.9	11.07
7/15/1999	OL	955	317	19.7	1.5	0.4	29.7	25	4.7	9.49
7/16/1999	OL	955	231	18.9	1.6	0.4	34	28.9	2.7	9.14
7/17/1999	OL	1018	237	19.7	1.5	0.4	27.7	24.1	3.7	9.17
7/18/1999	OL	934	321	18.9	1.6	0.4	34	28.9	6.1	9.49
7/19/1999	OL	1010	292	20.4	1.9	0.7	28	22.3	3.7	10.44
7/20/1999	OL	910	441	20.4	1.9	0.7	31.7	27.6	4.2	10.76
7/21/1999	OL	1020	331	20.4	1.9	0.7	34.5	30.1	4.3	12.02
7/22/1999	OL	910	441	20.4	1.9	0.7	31.7	27.6	4.2	10.76
7/23/1999	OL	1020	331	20.4	1.9	0.7	34.5	30.1	4.3	12.02
7/24/1999	OL	923	474	24.9	2	0.6	47.2	39.8	8.4	20.97
7/25/1999	OL	1030	563	28.2	2.8	0.9	40.3	34	6.3	22.41
7/26/1999	OL	950	493	28.2	2.8	0.9	34	29.2	4.8	16.76
7/27/1999	OL	1005	541	30.9	3.3	1.1	34	28.9	11.2	26.96
7/28/1999	OL	1010	465	28.2	2.8	0.9	41	34.4	6.6	16.84
7/29/1999	OL	1010	251	18.9	1.6	0.4	34	28.9	6.1	9.49
7/30/1999	OL	1005	473	17.8	1.6	0.4	18.3	15.9	2.3	4.19
7/31/1999	OL	1010	30	22.8	1.5	0.5	30	25.9	4.2	9.46
8/1/1999	OL	1018	212	8.8	1.1	-0.1	24.7	21.1	3.8	4.81
8/2/1999	OL	1004	231	22.8	2.1	0.7	54	46.9	7.2	9.2
8/3/1999	OL	1011	94	32.8	2.1	0.7	46.8	38.9	6.8	9.49
8/4/1999	OL	990	88	42.1	2.9	1.1	41	35.1	7.9	9.8
8/5/1999	OL	951	553	42.2	3.1	0.8	38	31	7	8.88

Metadata

(literally "data about data")

"a set of data that describes and gives information about other data" — Oxford English Dictionary



- ### What is the Dublin Core?
- o A metadata standard for describing digital resources
 - o An initiative to create a "library card catalog" for the Web
 - o Dublin Core fields:
- | | | |
|-------------|-----------|-------------|
| Title | Creator | Subject |
| Description | Publisher | Contributor |
| Date | Type | Format |
| Identifier | Source | Language |
| Relation | Coverage | Rights |
- The School University of Maryland



- ### Encoding Metadata
- o Language for encoding metadata should be:
 - **Universal** - so all can understand
 - **Flexible** - to incorporate different types
 - **Extensible** - flexible to custom types
 - **Simple** - to encourage adoption
 - **Modular** - so that schemes can be mixed, extended
- The School University of Maryland
- From: Ian Graham, An Introduction to RDF, <http://www.toronto.ca/ian/iake/>

**How do we encode metadata?
To support interoperability?**

January 31, 2001
31 janvier 2001
2001-01-31
01-31-2000
31012000

What is XML?

- XML = eXtensible Markup Language
- XML is a standard for exchanging structured data
 - Provides standardization at the syntactic level
 - Does **not** provide "meaning" for the tags
- XML is a standard recommended by the W3C

Goals of XML

- Easy to use
- Easy to extend and adapt
- Easy to write programs that use XML
- Support a wide variety of applications
- Should be human legible
- Formal and concise

The Basic Rules

- XML is case sensitive
- All start tags must have end tags
- Elements must be properly nested
- XML declaration is the first statement
 - `<?xml version="1.0"?>`
- Every document must contain a root element
- Attribute values must have quotation marks
 - `<item id="33905">`
- Certain characters are reserved for parsing
 - `<` = '`<`'

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://media.example.com/audio/guide.ra">
    <dc:creator>Rose Bush</dc:creator>
    <dc:title>A Guide to Growing Roses</dc:title>
    <dc:description>Describes process for planting and nurturing different kinds of
    rose bushes.</dc:description>
    <dc:date>2001-01-20</dc:date>
  </rdf:Description>
</rdf:RDF>
```

What does XML actually do?
(NOTHING)

Syntax vs. Semantics

(We'll come back to this later...)

How is XML like HTML?
How is HTML like XML?

XML: Historic Perspective

- HTML and the birth of the Web
- HTML is not enough
- Development of XML

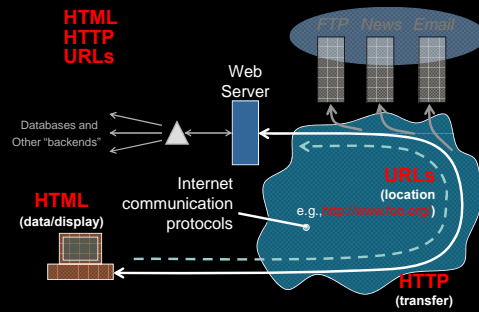
The next few slides are adapted from presentations by Ian Graham:
<http://www.utoronto.ca/ian/talks/>

The iSchool
University of Maryland



In the beginning...

The foundations of the Web:



The iSchool
University of Maryland



Three Core Technologies

- **HTTP** - HyperText Transfer Protocol
 - A protocol for transferring data between machines on the Internet
- **URL** - Uniform Resource Locator
 - A scheme for referencing the specific location of a resource
- **HTML** - HyperText Markup Language
 - A markup language for encoding information to be read by humans

HTTP and URLs have pretty-well stood the test of time.
But by 1996, HTML was already showing signs of age

The iSchool
University of Maryland



HTML

- Started with very few tags ...
- Language evolved as more tags were added:
 - Forms
 - Tables
 - Fonts
 - Frames
 - ...

The iSchool
University of Maryland



Problems with HTML

- I want personalized tags
 - HTML can't be extended
- I want to incorporate other types of data
 - Mathematics, database entries, literary text, poems, purchase orders ...
 - HTML can't accommodate other types of data
- I want to process pages automatically with software
 - HTML is too messy and inconsistent



Back to Basics

- HTML was defined using SGML
 - Standard Generalized Markup Language
 - A meta-language for defining languages
- Complex, sophisticated, powerful
 - ... too difficult to use
- Idea: create a simpler version of SGML
 - The birth of XML!



XML Languages

- XML can be used to define other languages
- Many XML languages, optimized for different roles
 - XHTML: HTML by XML rules
 - MathML: for mathematics
 - SMIL: for synchronized multimedia
 - RSS: for news feeds
 - ...



XHTML: Cleaning up HTML

```
<?xml version="1.0" encoding="iso-8859-1"?>
<html xmlns="http://www.w3.org/TR/xhtml1" >
<head>
  <title> Title of text XHTML Document </title>
</head>
<body>
  <div class="myDiv">
    <h1> Heading of Page </h1>
    <p> here is a paragraph of text. I will include inside this paragraph
      a bunch of wonky text so that it looks fancy. </p>
    <p>Here is another paragraph with <em>inline emphasized</em>
      text, and <b>absolutely no</b> sense of humor. </p>
    <p>And another paragraph, this one with an  image, and a <br /> line break. </p>
  </div>
</body></html>
```



MathML

An XML language for defining mathematic formulas

```
x2 + 4x + 4 = 0
<mrow>
  <mrow>
    <msup><mi>x</mi><mn>2</mn></msup>
    <mo>+</mo>
    <mrow>
      <mn>4</mn>
      <mo>&InvisibleTimes;</mo>
      <mi>x</mi>
    </mrow>
    <mo>+</mo><mn>4</mn>
  </mrow>
  <mo>=</mo><mn>0</mn>
</mrow>
```

See <http://www.mozilla.org/projects/mathml/demo/tester.html>



SMIL

- Synchronized Multimedia Integration Language
- Integration of multimedia with text, audio, video
- Support in RealPlayer

See demo linked from syllabus



RSS

- RSS = Really Simple Syndication or Rich Site Summary
- An XML format for distributing news headlines on the Web

See example at <http://www.nytimes.com/services/xml/rss/>

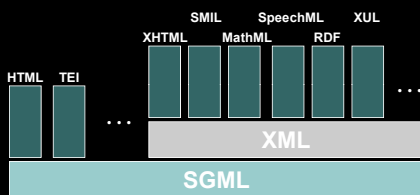


And Others...

- CML – chemical Markup Lang
- CeIIML – biological models
- BSML – bioinformatic sequences
- MAGE-ML – Microarray Gene Expression
- XSTAR – for archaeological research
- MARCXML – MARC in XML
- AML – astronomy markup language
- SportsML – for sharing sports data
- List goes on and on and on...



The XML Family Tree



Mixing XML Dialects

- XML is designed to support the integration of multiple standards
- Allows users to mix elements from different standards
 - Snapping together XML dialects like Lego pieces
 - Based on the notion of "namespaces"



Example

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rss="http://purl.org/rss/1.0/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rss:channel rdf:about="http://www.xml.com/xml/news.rss">
    <rss:title>XML.com</rss:title>
    <rss:link>http://xml.com/pub</rss:link>
    <dc:description>
      XML.com features a rich mix of
      information and services for the XML community.
    </dc:description>
    <dc:subject>XML, RDF, metadata, information
      syndication services</dc:subject>
    <dc:identifier>http://www.xml.com</dc:identifier>
    <dc:publisher>O'Reilly & Associates, Inc.</dc:publisher>
    <dc:rights>Copyright 2000, O'Reilly &
      Associates, Inc.</dc:rights>
  </rss:channel>
</rdf:RDF>
```

Source: <http://www.xml.com/pub/2000/10/25/dublincore>



Another Example

```
<?xml version="1.0" encoding="iso-8859-1"?>
<html xmlns="http://www.w3.org/TR/xhtml1" >
<head>
  <title> Title of XHTML Document </title>
</head><body>
  <div class="myDiv">
    <h1> Heading of Page </h1>
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      ... MathML markup ...
    </math>
    <p> more html stuff goes here </p>
    <smil xmlns="http://www.w3.org/TR/smil1">
      ... SML markup ...
    </smil>
  </div>
</body></html>
```

See demo linked from syllabus



Take-Away Messages

- Metadata makes data useful
- XML is a way to encode data and metadata
- XML allows computers to exchange information in new and interesting ways

The iSchool
University of Maryland



Interoperability

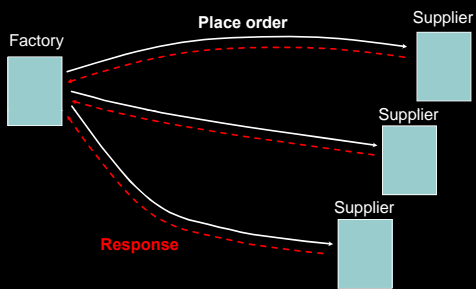
- What does it mean and what's the role of XML?
- XML as a universal format for data interchange
 - Software exchange data as XML-format messages
- Advantages?
 - Eliminates proprietary data formats
 - Promotes interoperability
 - Encourages cooperation
 - Leverages lots of existing XML processing software

The next few slides are adapted from presentations by Ian Graham:
<http://www.utoronto.ca/ian/talks/>

The iSchool
University of Maryland



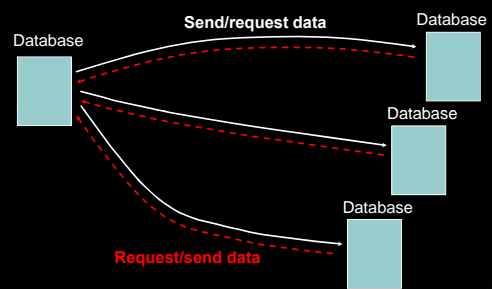
XML Messaging



The iSchool
University of Maryland



XML Messaging



The iSchool
University of Maryland



What's in it for me?

Evolution of Software:

Desktop Applications
Web Forms
Rich Web Applications

Web 2.0 Mashups

Syntax vs. Semantics

How does meaning “come about”?

Take-Away Messages

- Metadata makes data useful
- XML is a way to encode data and metadata
- XML allows computers to exchange information in new and interesting ways

The School
University of Maryland

