XML Tutorial

Thomas G. Habing Grainger Engineering Library Information Center

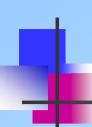
E-Text Working Group 2004-01-15



Tom Habing

Research Programmer, Grainger Engineering Library Information Center

thabing@uiuc.edu



Agenda - XML Tutorial

- Well-formed XML (basic markup)
- DTDs and schemas
- 3. Stylesheets
- XML and databases

Topics include illustrations, demos, and exercises.



Well-formed XML

XML 1.0 Specification

- Markup
 - Prolog & Document Type Declaration
 - Elements
 - Attributes
- Content
 - Entities
 - Encoded (Unicode) characters

Pro

Prolog & Document Type Definition

- XML documents should begin with an XML Declaration which specifies version
 - Optionally may also include:
 - Encoding (recommended)
 - Stand-alone declaration
- Document Type Definition is typically next

- <?xml version="1.0" encoding='UTF-8' standalone='no' ?>
- <!DOCTYPE root SYSTEM "myDocs.dtd" >



Elements

Elements are markup that enclose content

- <element_name>...</element_name> or <element_name />
- Content models
 - Parsed Character Data Only
 - Child Elements Only
 - Mixed
 - Empty

<author>Cole, T</author>

- Attributes

Associate a name-value pair with an element

- <tag name1="value1" name2='value2'>...</tag>
 - Can be used to embellish content...
 - or to associate added content to an element
- Attributes beginning XML are reserved

```
<author order='1'>Cole, T</author>
<author name='Habing, T' />
```

Entities

- Placeholders for internal or external content
 - Placeholder for a single character...
 - or string of text...
 - or external content (images, audio, etc.)
- Implementation specifics may vary

```
<!ENTITY copyright "&#xA9;" >
&copyright; is replaced by ©
<!ENTITY pic SYSTEM "mugshot.gif" NDATA gif >
&pic; is replaced by graphic image
```

Character Encoding Issues

- XML Parsers must accept UTF-8 & UTF-16
- Also must accept &#nnnn; or &#xhhhh;
- MARC-8 encodings <u>must</u> be converted to Unicode for use in XML
- Special rules for EOL and control characters
 - Changes will be effective version 1.1

http://lcweb.loc.gov/marc/specifications/specchartables.html

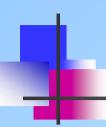


<u>Namespaces</u>

- Qualify element and attribute names
- Allows modularization of schemas
 - Mix and match elements from multiple schemas in document instances
 - Import or include from one XML Schema into another



http://www.zvon.org/xxl/XMLTutorial/General/contents.html



Simple Illustrations

A POEM

Unqualified Dublin Core



Simple Authoring Tools

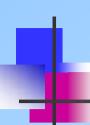
MS Notepad (Plain Text Editor)

MS XML Notepad Beta 1.5



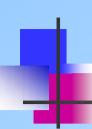
- Markup one or more of the sample Bib records
- Use <BibRecord> for your root element
- Use Dublin Core element names where applicable:

title	creator	subject
description	publisher	contributor
type	format	date
coverage	relation	identifier
rights	language	source



Advanced Examples

- MARC
- MODS
- OAI (version 2.0)
- DLI/DLIB Test Suite Journal Article
- RDF Qualified Dublin Core
 - Guidelines for <u>DC</u> & <u>RDF DC</u> in XML
- SOAP (Primer)



Advanced Authoring Tool Demos

YAWC - Microsoft Word

Extensibility TurboXML

2. DTDs and Schemas

- Formal descriptions of document structure
 - Set expectations
 - Maximize reusability
 - Enforce business rules
- Validation ensures that documents conform
 - Could be multiple schemas for different purposes or different points in a documents lifecycle
- Primary: DTD or XML Schema Language
- Others: Schematron or Relax NG

Document Type Definitions (DTD)

Legacy from SGML; part of XML standard

```
<!DOCTYPE Book SYSTEM 'http://...'>
<!ELEMENT Book (Front, Chapter+, Back?)>
<!ATTLIST Book
    type (series | monograph) #REQUIRED>
```



http://www.zvon.org/xxl/DTDTutorial/General/contents.html



A simple DTD for poems

XHTML 1.0 Strict

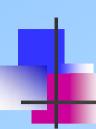
- How to validate
 - In Internet Explorer
 - Using XML Notepad

DTD Exercise

- Create a DTD for the previous XML Bib Files
 - Name your DTD file BibClass.dtd
- Insert at the top of each of your XML Bib Files

- <?xml version="1.0" encoding="UTF-8"?>
- <!DOCTYPE BibRecord SYSTEM "BibClass.dtd" >

Validate using XML Notepad or Internet Explorer



More Examples of DTDs

- OASIS DocBook
- TEI
- Gutenberg



XML schema language

- New in XML
 - Uses XML syntax
 - Supports datatyping
 - Richer and more complex



- BibClass.xsd
- Simple Dublin Core

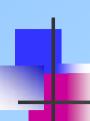
- Datatypes and Facets
 - Unions
 - Enumerations
 - Lists



More examples of XML Schemas

Qualified Dublin Core

OAI-PMH



Alternatives: Schematron & RelaxNG

- Schematron based on XPath (XSLT)
 - Doesn't support datatyping as well
 - Supports additional content models
 - May become an ISO standard

RelaxNG

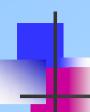
- Returns some of the power of SGML DTDs back to XML (mixed and unordered content)
- Uses datatyping from the XML Schema spec
- Does not support inheritance
- Developed by an OASIS Technical Committee chaired by James Clark



Extensibility TurboXML

Corel/SoftQuad XMetaL

XSV for validating



XML & Cascading Style Sheets

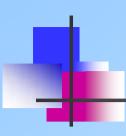
- Attach styling instructions directly to XML files
 - <?xml-stylesheet href="http:..." type="text/css" ?>
 - Supported by newest browsers: IE5+, Mozilla, Opera
- Can style but not rearrange elements
 - Block or inline style
 - Bold, italic, underline, font, color, etc.
 - Margins, positioning
 - Generated content (browser support not good)

```
front author {color:red; font-weight:bold; font-
    family:serif;}
```



Language for transforming XML documents

- Into HTML, Text, or other XML documents
- Supported in new browsers (IE5+, Mozilla; not Opera)
- Usually applied on the server or in batch mode
- Valuable for interoperability or reusability



CSS & XSLT Examples

A CSS for poems in XML

- XSLT for Dublin Core XML to XHTML
- XSLT for MARC21 XML to HTML

XSLT for MARC21 XML to DC XML



http://www.zvon.org/xxl/XPathTutorial/General/examples.html

http://www.zvon.org/xxl/XSLTutorial/Output/contents.html

XML Tutorial

XSLT Exercise

- Create a simple XSLT to transform your previous
 Bib XML files into HTML
 - Name the file simple.xsl
- Either attach the XSLT to your files using <?xml-stylesheet href="simple.xsl" type="text/xsl"?> And display it in IE
- Or use MSXSL from the command line to create the HTML file to display msxsl.exe source.xml simple.xsl –o source.htm



- TIBCO XMLTransform
- MSXSL
- EXSLT

2004-01-15

XSL Formatter

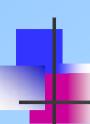
33



XML & databases

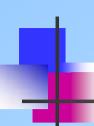
Discussion Points

- Suitability of XML for databases
- XML Query Language
- XML & relational databases
- Demonstrations
 - XML and MS SQL Server
 - A simple XML search using XSLT



Suitability for databases

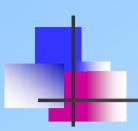
- Documents are well-structured (wellformed syntax requirement)
- Content is well-labeled (markup semantics)
- Ubiquitous, cross-application, crosscommunity
 - Open, non-proprietary standard
- Validated XML potentially "machineunderstandable"



XML Query Language

Intended to facilitate "interaction between the web world and the database world."

- Follow-on to Quilt (Don Chamberlin, et al.)
- Another special-purpose programming language for processing XML (& XML-like database structures)
 - Expressions, data types, functions, etc.
- Access (search) functionality e.g., not for update
- Like XSLT, it relies on XPath
- Borrows data types from XML schema language
- Template Processor (like XSLT, PHP, JSP, ASP)



E.g., List books published by Addison-Wesley after 1991:

XQuery:

</book>

</bib>

```
<br/>
```

37



XML Query Implementations

- W3C Grammar Test Page
- SoftwareAG Tamino / QuiP
 - Queries XML files directly or XML in Tamino
- Microsoft's <u>XQuery Demo</u>
 - Online demo queries <u>W3C use cases documents</u>
- Oracle prototype XQuery implementation
 - Queries XML files
 - Same site has Oracle SQL/XML implementation
- Xindice (apache.org)
 - Uses XPath instead of XQuery

XML & Relational DBMS

- SQLX
 - Mapping data types, character sets, ... from SQL to XML
 - Searching XML using XPath-based SQL style queries
- Import / Export of XML from relational database management systems now common
 - Specifics vary by implementation
- Microsoft implementation allows:
 - Query relational DB using XPath
 - Query with SQL, return XML
 - Load XML directly into Microsoft SQL
 - OLEDB provider for XML files



- Using XSLT to query multiple XML docs
 - XSLT to search multiple DC XML files

Fetching XML from Microsoft SQL



Contact information

- Presentation
 - http://dli.grainger.uiuc.edu/ETEXT/
- Presenters
 - thabing@uiuc.edu

41