XML for Java Developers
G22.3033-002

Session 1 - Main Theme
Markup Language Technologies (Part I)

Dr. Jean-Claude Franchitti
New York University
Computer Science Department
Courant Institute of Mathematical Sciences

Agenda
- XML Generics
- Course Logistics, Structure and Objectives
- History of Meta-Markup Languages
- XML Applications: Markup Languages
- XML Information Modeling Applications
- XML-Based Architectures
- XML and Java
- XML Development Tools
- Summary
- Class Project
- Readings
- Assignment #1a

XML Generics
- XML means eXtensible Markup Language
- XML expresses the structure of information (i.e., document content) separately from its presentation
- XSL style sheets are used to convert documents to a presentation format that can be processed by a target presentation device (e.g., HTML in the case of legacy browsers)
- Document Type Definitions (DTDs) or XML schemas are used to describe and validate content
Course Logistics

- Course Web Site
  - http://cs.nyu.edu/courses/summer01/G22.3033-002/index.htm
  - http://www.nyu.edu/classes/jcf/g22.3033-002/
  - Password: xmljava
  - Review syllabus
- Textbooks
  - XML Development with Java 2
  - Professional Java XML
  - XML and Java, Developing Web Applications

Knowledge Required

- Know how to ftp files, send e-mail, and load URLs into a Web browser
- Reasonable knowledge of HTML
- Be able to write a basic HTML page including links, images, and text using a text editor
- Be able to place a page on a web server
- Advanced Java programming experience

Knowledge Not Required

- SGML
- Programming in Perl, or any language other than Java
- Cascading Style Sheets and other HTML advanced features
Other Useful Knowledge

- Web server configuration and the HTTP protocol
- JavaScript
- C programming
- Database theory (normalization rules)
- SGML (basic knowledge)
- Some knowledge of international character sets

History of Meta-Markup Languages

- SGML (see http://www.w3c.org)
  - Complex and expensive
- HTML
  - Tim Berners-Lee (Cern), 1990
  - Simple replacement to SGML
  - Lacked the flexibility and adaptability of XML
- W3C XML working group
  - Jon Bosak (Sun), 1996
  - SGML strengths + HTML simplicity

History of Meta-Markup Languages (continued)

- XML 1.0 Specification
  - Tim Bray
  - Language simple enough for programmers to implement
  - Language not limited to English
  - Documents easy for search engine
- CML, Jumbo, MathML
Why XML?
- Domain specific markup language
- Common data format
  - Simple, pure Unicode text
- Well-documented
- Data interchange
  - Open Financial Exchange (OFX)
  - XML non copyrighted
- Support of portable structured text data as most electronic information is text-based

Related Technologies
- JavaSoft’s Java API for XML parsing (JAXP)
- Java bindings to SAX, and DOM APIs
- HTML
- CSS, CSS2, and CSS3
- XSL
  - well formed XML documents
- URLs and URIs
  - resource v.s. location
- Linking (XLL)
- Unicode

Applications of XML: Markup Languages
- XML helps define sets of rules for forming semantic tags that break documents into parts
- XML is a Metamarkup language
  - Not a markup language restricted to a fixed set of tags
  - Can make up tags as you go along
- XML Tags can be documented in a DTD
  - i.e., a domain specific syntax and vocabulary (e.g., CML, MathML, etc.)
**HTML Example**

```html
<dt>Java</dt>
<dd>by Sun MicroSystem</dd>
<ul>
<li>Publisher: O’Reilly</li>
<li>Pages: 900</li>
</ul>
<p>The complete documentation.</p>
```

**XML Corresponding Structure**

```xml
<book>
  <title>Java</title>
  <author>Sun Microsystems</author>
  <publisher>O’Reilly</publisher>
  <isbn>0-777-88888-9</isbn>
  <pages>900</pages>
  <blurb>The complete documentation</blurb>
</book>
```

**XML Enables Semantic/Structured Markup Languages**

- XML describes a document’s structure and meaning
- XML does not describe the formatting of the elements
- Formatting is added through style sheets
- XML document only contains tags describing the content (not the appearance) of the document
**XML Information Modeling Applications**  
(Focus on Complex Structured Document Specification)

- Custom and Industry specific Markup Languages with own syntax and vocabulary
- CML (Chemical Markup Language)
- MathML (Mathematical Markup Language)
- CML (Channel Definition Format)
- Classic Literature
- SMIL (Synchronized Multimedia Integration Language)
- OSD (Open Software Description)

---

**XML-Based Architectures**

- Presentation-Oriented Publishing Frameworks
- XML-Based Application Servers
- Message-Oriented Middleware Frameworks
- Web Services Frameworks
  - Peer to Peer (P2P) Computing
  - XML-Based eServices Protocols and Architectures

---

**XML POP Frameworks**  
(Focus on Content Management and Publishing)

- Requirements:
  - XML parser
  - XSL processor
  - Document repository
  - Collection of document DTDs or schema
  - Collection of XSL style sheets
- Parsing & processing can be client or server-based
- Samples server-side XML technology:
  - Perl & CGI ([http://www.perl.com/CPAN](http://www.perl.com/CPAN))
  - ASPs ([http://www.beyondhtml.com/rocket](http://www.beyondhtml.com/rocket))
XML POP Packaged Solutions
- Enterprise Information Portal (EIP) toolkits
  - e.g., DataChannel’s XML framework
- Structured content management and retrieval systems
  - e.g.,
    - Vignette’s StoryServer
    - Gauss Enterprise’s VIP XML-Gateway
    - Poet’s Content Management Suite
    - Chrystal Software’s Astoria

XML MOM Applications
- Component-based modeling
  - e.g., XML Metadata Interchange (XMI)
- Enterprise Application Integration
- Business-to-Customer (B2C) Applications
  - e.g., e-Commerce
- Business-to-Business (B2B) Applications
  - e.g.,
    - WebMethods
    - Bluestone’s XML-Server
    - eXcelon 2.0
    - XMLSolutions’ ExeterXML server

Java-enabled XML Technologies
- XML provides a universal syntax for Java semantics (behavior)
- Portable, reusable data descriptions in XML
- Portable Java code that makes the data behave in various ways
- XML standard extension
  - Basic plumbing that translates XML into Java
  - parser, namespace support in the parser, simple API for XML (SAX), and document object model (DOM)
  - XML data binding standard extension
XML and Java Standards

- XML includes is a family of technologies
  - XSL, XML Schema, XML Query, XPath, XPointer, XLink, DOM, RDF, CSS, XSL, XHTML, XML Signature, MathML, SMIL, SVG, etc.
- Review the current state of the XML standards at http://www.w3c.org/XML
- Review the current state of Java Technology and XML (JAXP) standards at http://java.sun.com/XML
- Review the Java binding to DOM 2.0 at http://www.w3.org/TR/2000/REC-DOM-Level-2-Core-20001113/java-binding.zip

XML Development Tools

- Read and write XML documents
- XML Editors
  - e.g., Wattle Software’s XMLWriter, SoftQuad’s Xmetal, Vervet Logic’s XML Pro, eXcelon Stylus, XML Spy
- XML well-formedness and/or validation Processors
  - e.g., IBM’s XML4J, Apache’s XercesJ, Oracle Parser, Java Project X processor, James Clark’s Expat parser, Microsoft MSXML
- XML Browsers
  - e.g., IE5

Summary

- XML is a metamarkup language
- XML tags describe the structure and semantics of a document’s content (not the format of the content)
- XML addresses SGML’s complexity and HTML’s inadequacies
- Applications of XML for information modeling exist in various domains (chemistry, math, multimedia, etc.)
Summary (continued)

- XML supports MOM and POP applications
- XML supports the Web Services Infrastructure
- XML and Java complements each other
- Java-enabled XML technology is available in the market today
- Standards for integrating XML and Java are being developed

Class Project

- Project Description
  - The project will focus on "multi-channel online community platforms", and will consist of providing custom XML-based services to support the various aspects of a chosen eBusiness application implemented on such platforms (e.g., "community-based shopping"). Examples of useful services to support these platforms may include synchronized multimedia presentation viewing, and "offline" chat capabilities. A sample specification of an online community platform for a virtual university eBusiness application will be provided for illustration purpose.

Readings

- Readings
  - XML Development with Java 2: Chapter 1
  - Professional Java XML: Chapter 1
  - XML and Java: Chapter 1, Appendix C
  - Handouts posted on the course web site
  - Explore brief review of XML at XML in 10 points
  - Read article at XMLJ2EE
  - Review XML 1.0 (Second edition), Namespaces, and the XML activity statement at W3C
- Project Frameworks Setup (ongoing)
  - Apache’s Web Server, TomCat, and Cocoon
  - Apache’s Xerces, Xalan
Assignment

- Assignment #1a:
  - Explore the textbooks’ CDs, and the textbook references to XML tools
  - Install and experiment with XML editors. Settle on a tool you feel comfortable using, and come up with a small XML application that demonstrates the use of that tool.
  - Write a short report that uses your sample application to document your findings and recommendations with respect to selection criteria in support of XML editors

Next Session:
Markup Language Technologies (Part II)

- Current state of the XML standard
- Advanced XML applications
- XML style Specification Languages