Extensible Markup Language
Eric William Burger
Centigram Chief Scientist
GMU DACCE Student
November 16, 1999
Roadmap

- Introduction to XML, The Language
- Applications and XML
- Just What is a Document?
- Futures
- References
What is HTML?  

- HTML Defines Presentation
  
  `<P><B>Prof. Sanjeev Setia</B></P>
  <BR>Department of Computer Science
  <BR>George Mason University
  <BR>4400 University Dr.
  <BR>Fairfax, VA  22030-4444

- Rendered as
  
  Prof. Sanjeev Setia  
  Department of Computer Science  
  George Mason University  
  4400 University Dr.  
  Fairfax, VA  22030-4444

- Note no Meaning, Only Layout
What is XML?

- XML Defines Content

```xml
<address>
  <name>
    <title>Prof.</title>
    <first-name>Sanjeev</first-name>
    <last-name>Setia</last-name>
  </name>
  <department>Department of Computer Science</department>
  <organization>George Mason University</organization>
  <street>4400 University Dr.</street>
  <city>Fairfax</city><state>VA</state>
  <zipcode>22030-4444</zipcode>
</address>
```
Rendering XML

- With a Given Style Sheet, Can Render Same as HTML

```xml
<address>
  <name>
    <title>Prof.</title>
    <first-name>Sanjeev</first-name>
    <last-name>Setia</last-name>
  </name>
  <department>Department of Computer Science</department>
  <organization>George Mason University</organization>
  <street>4400 University Dr.</street>
  <city>Fairfax</city><state>VA</state>
  <zipcode>22030-4444</zipcode>
</address>

- Rendered as

  Prof. Sanjeev Setia
  Department of Computer Science
  George Mason University
  4400 University Dr.
  Fairfax, VA  22030-4444
Alternate Renderings

◆ Same XML, Different Stylesheet

```
<address>
  <name>
    <title>Prof.</title>
    <first-name>Sanjeev</first-name>
    <last-name>Setia</last-name>
  </name>
  <department>Department of Computer Science</department>
  <organization>George Mason University</organization>
  <street>4400 University Dr.</street>
  <city>Fairfax</city><state>VA</state>
  <zipcode>22030-4444</zipcode>
</address>
```
How Can I See XML?

◆ In Your Browser (Experimental - IE5)
◆ Through XML Enabler (IBM)
HTML vs. XML Tags

- HTML Tags Are Fixed
  - H1, H2, P, B, I, FONT, etc.
- XML Tags Undefined
  - Defined in a Document Type Definition (DTD)
- DTD Specifies
  - Legal Tag Names
  - Nesting Rules for Tags (Grammar)
  - Entity Declarations
- Tag Meanings
  - HTML Tags Specify Rendering Format
  - XML Tags Specify Information
  - XSL Map XML Document to Rendering Format
Power of XML As Markup Language

- troff on Steroids
  - Conditional Inclusion
  - Hierarchical Include With Override
- Escape to Interpreter / OS From Document
  - Security Issues “Mentioned” 😊
Kinds of XML Documents

- **Invalid Document**
  - Doesn’t Follow XML Tag Rules
    - All Start Tags Have End Tags `<P>foo</P>`
    - XML Tags May Not Overlap `<B><I>foo</B></I>`
    - Empty Tag Shorthand `<entity parameter="value" />`

- **Well-Formed Document**
  - Follows Tag Rules
  - Doesn’t Have or Follow DTD

- **Valid Document**
  - Has DTD
  - Meets Grammar Specified in DTD
Validating & Non-Validating

- Well-Formed Documents Can Be Useful
- Class of Non-Validating Parsers
  - Not Necessarily Portable
- Validating Parser Takes More Time
- BTW: XML is 100% SGML
  - Only Valid, Well-Formed Documents are SGML
Roadmap

- Introduction to XML, The Language
- Applications and XML
- Just What is a Document?
- Futures
- References
How Does XML Get to Me?

- **IIOP**
  - Not Firewall Friendly
  - Binary Transformation
- **MOM**
  - proprietary
- **HTTP/HTTPS**
  - Firewall Friendly
  - Not Binary Friendly
    - XML ONLY SUPPORTS PRINTABLE CHARACTERS
    - Binary Data Must Be uuencoded Or (Better Yet) base64 Encoded (No [<\[]&])
Why All This Java Stuff?

- Promise of Java Was Distributed Applications on Any Browser
- Problem Is How to Represent Data
- Given Tagged Data
  - User-Driven Views Easy
  - View Computations at Client, not Server
- Nice Synergy With HTTP for Transport
- “XML Gives Java Something to Do.”
Application Access to XML

- Document Object Model (DOM)
  - Interprets XML and Returns Object Tree

- Simple API for XML (SAX)
  - Register Callbacks to Interpreter
  - Interpreter Calls Callbacks When Object Found
  - Akin to bison or yacc

- Availability
  - Sun, IBM, Microstar, Microsoft, Oracle, FSF
    - All Free; These Are Only the Ones I Know About!
  - Java, Perl, C++, C, Tcl, Delphi, Python
Rendering is Nice But Not New

- XML-Tagged Data is System Independent
- Schema Travels With Data
- Relatively Free-Form DTD
- Impact
  - Easy EDI
  - Smart Agents Understanding Web Data
  - Smart Searches: Search for “Chip”
Roadmap

- Introduction to XML, The Language
- Applications and XML
- Just What is a Document?
- Futures
- References
What Is a Document?

- Voice Markup Language
- Defines a DTD For Telephone User Interface Interaction

```xml
<?xml version="1.0"?>
<vxml>
  <var name="hi" expr="'Hello World!'"/>
  <form>
    <block><value name="hi"/>Goodbye!</block>
  </form>
</vxml>
```
<form id="weather_info">

  <block> Welcome to the weather information service. </block>

  <field name="state">
    <prompt> What state? </prompt>
    <grammar src="weather.gram#state"/>
    <catch event="help">
      Please speak the state for which you want the weather.
    </catch>
  </field>

  <field name="city">
    <prompt> What city? </prompt>
    <grammar src="weather.gram#city"/>
    <catch event="help">
      Please speak the city for which you want the weather.
    </catch>
  </field>

  <block>
    <goto next="http://www.clouds.example/" submit="city state"/>
  </block>

</form>
But There’s More

◆ So, We Have a New Forms Concept
◆ Could It Work Visually?
◆ Absolutely
  ● MobileADK (Motorolla)
  ● VoxML Goes In
  ● Voice TUI Comes Out
  ● WAP Comes Out
Interapplication Communication

◆ **RDB**
  - Use Entity for Table Row
    - XML Entity Attributes for Columns
    - XML Sub-Entities for Columns
  - Quasi-Native XML I/O In Oracle 8i

◆ **CORBA**
  - Use CORBA & IDL As Is
  - Use XML/HTTP for IIOP
Roadmap

- Introduction to XML, The Language
- Applications and XML
- Just What is a Document?
- Futures
- References
Expected Impacts

- Transition Pain of HTML → XML
  - Only Will Happen For People That Need It, E.g.,
    - Web Sites That Want to be Found
    - Open E-Commerce Sites
- Business Opportunity for Being DTD Repository
- Replacement of Flat File Systems for Object Database File Systems (Poet Wish)
XML Certainly More Verbose Than HTML

Don’t Need “Click to Sort By Date, Name, Cost, etc.”

Can Cache DTD’s

Most Likely More Bandwidth Will Be Used to Deliver More Functionality
Roadmap

- Introduction to XML, The Language
- Applications and XML
- Just What is a Document?
- Futures
- References
References - Papers

◆ Annotated XML Definition

◆ Tidwell Tutorial Introduction

◆ “Media-Independent Publishing: Four Myths about XML” By John Bosak
  ● http://metalab.unc.edu/pub/sun-info/standards/xml/why/4myths.htm


◆ Simple Object Access Protocol (SOAP)
References - Organizations

- Voice Markup Language Forum
  - [http://www.voxml.com/voxml](http://www.voxml.com/voxml)
- XML Architecture / Standards
  - [http://www.xml.org](http://www.xml.org)
- XML DTD Standards / Promotion
  - [http://www.xml.com](http://www.xml.com)
- W3C XML Site (Architecture Group)
  - [http://www.w3.org/XML](http://www.w3.org/XML)
References - Corporate

- IBM DeveloperWorks Site
- Sun XML Tools & Information
- Oracle XML Tools & Information
  - http://technet.oracle.com/tech/xml/
- Microsoft XML Tools & Information
Discussion