

XML Introduction

Grahame Grieve

Outline

- What's XML?
- Related Standards
- Tools overview

What's XML?

SGML -> HTML -> XML

```
<?xml version="1.0" encoding="UTF-8"?>  
<rootElement>  
  <element attributeName="value">  
    <childElement>Text</childElement>  
  </element>  
</rootElement>
```

What does XML achieve?

- Semantics not presentation
- Accepted document/data exchange
- Unicode
- Flexible / Robust document markup
- Focus for tooling

Other XML Standards

- DOM
- SAX
- XPath
- Namespaces
- XSL/XSLT
- XLink
- DTD
- Schema
- Schematron
- Canonical XML
- XML-DSig
- SOAP / WSDL

DOM

- Document Object Model for XML
- Load document into a set of Objects

```
<?xml version="1.0" encoding="UTF-8"?>  
<rootElement>  
  <element attributeName="value">  
    <childElement>Text</childElement>  
  </element>  
</rootElement>
```

SAX

- Serial API for XML
- Report events as document loads

```
• XML processing Instruction
• Element "rootElement"
• Element "element"
• Attribute "attributeName" = "value"
• Element "childElement"
• Text "Text"
```

XPath

- Query language for XML

```
/rootElement/element/childElement = "Text"
/rootElement/element/@attributeName = "value"
```

```
<?xml version="1.0" encoding="UTF-8"?>
<rootElement>
  <element attributeName="value">
    <childElement>Text</childElement>
  </element>
</rootElement>
```

Namespaces

- Resolve naming conflicts

```
<?xml version="1.0" encoding="UTF-8"?>
<rootElement xmlns:a="ns1" xmlns:b="ns2">
  <element attributeName="value">
    <a:childElement>Text</a:childElement>
    <b:childElement>Text</b:childElement>
  </element>
</rootElement>
```

XSLT

- Transform documents
- Validation

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0"
  xmlns:xsl="http://www.w3.org/1999/XSL/..."
  <xsl:template match="childElement"/>
  <xsl:template match="/" | @* | * | comment...
  <xsl:copy>
  <xsl:apply-templates select="@* | * | ...
  </xsl:copy>
  </xsl:template>
  </xsl:stylesheet>

<?xml version="1.0" encoding="UTF-8"?>
<rootElement>
  <element attributeName="value">
  <childElement>Text</childElement>
  </element>
  </rootElement>
```

XLink

- url references for XML files

```
<?xml version="1.0" encoding="UTF-8"?>
<rootElement>
  <element attributeName="value">
    <childElement
      xmlns:xlink="http://www.w3.org/1999/xlink"
      xlink:type="simple"
      xlink:href="sometext.xml"
      xlink:title="Text for this Node"
      xlink:show="new"
      xlink:actuate="onRequest"/>
    </element>
  </rootElement>
```

DTD

- Document Type Definition
- Rules for document structure

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT childElement (#PCDATA)>
<!ELEMENT element (childElement)>
<!ATTLIST element
  attributeName CDATA #REQUIRED
>
<!ELEMENT rootElement (element)>
```

Schema

- Rules for document structure
- More control over types and cardinality

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">
  <xs:element name="childElement" type="xs:string"/>
  <xs:element name="element">
    <xs:sequence>
      <xs:element name="childElement" ref="childElement"/>
    </xs:sequence>
    <xs:attribute name="attributeName" type="xs:string" use="required"/>
  </xs:complexType>
  </xs:element>
  <xs:element name="rootElement">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="element" ref="element"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Schematron

- XPath rules that must be true

```
/rootElement/element/childElement
/rootElement/element/@attributeName
```

```
<?xml version="1.0" encoding="UTF-8"?>
<rootElement>
  <element attributeName="value">
    <childElement>Text</childElement>
  </element>
</rootElement>
```

Canonical XML

- Stylised conceptual format
- Resolves variability

```
<?xml version="1.0" encoding="UTF-8"?>
<rootElement>
  <element attributeName="value">
    <childElement>Text & more</childElement>
    <childElement>Text &#38; more</childElement>
  </element>
</rootElement>
```

XML D-Sig

- Digital signatures for XML
- Specifies what to sign, and how to encode a signature

```
<?xml version="1.0" encoding="UTF-8"?>
<rootElement>
  <element attributeName="value">
    <childElement>Text</childElement>
  </element>
</rootElement>
```

SOAP / WSDL

- Simple Object Access Protocol
- Web Services Description Language

```
POST /worker HTTP/1.1
Host: www.stockquotesserver.com
Content-Type: text/xml; charset="utf-8"
Content-Length: 234
SOAPAction: "process"

<SOAP-ENV:Envelope
  xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
  SOAP-ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/">
  <SOAP-ENV:Body>
    <a:rootElement xmlns:a="ns1">
      <a:element attributeName="value">
        <a:childElement>Text</a:childElement>
      </a:element>
    </a:rootElement>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

What can you do with XML?

- Nothing
- You have to define a schema that you will follow
- Shared understanding of document

XML Tools

- Interactive Tools
 - XMLSpy/XMLCooktop (IE 6.0)
 - IDE's, etc
- Programmers Tools
 - Msxml, Xerces
 - others
- Integration Tools
 - Database XML support