XML: More than just a cool acronym?

Michael Mason
DecisionSoft Limited

Who am I?

- Mike Mason, mgm@oxlug.org
- Oxford University, BA Computation
- CompSoc Secretary 1997/98
- Joined DecisionSoft June 1999, after spending 3 vacations as an intern

Who are DecisionSoft?

- Software startup established December 1997, with US venture capital backing
- Based in the Oxford Science Park
- Currently developing an XML tool suite, based around XML Script
- Products include X–Tract (command line utility) and X–Stream (server version)

Outline of today's talk

- What is XML?
- XML Standards
- Web apps and XML
- XML Transformations
- XSL and XML Script
- X–Tract demo

What is XML?

- eXtensible Markup Language
- Subset of SGML (Standardized Generalized Markup Language)
- Like HTML, but with user–defined tags.
- It´s got a cool name ;-)

XML is all about data

- **HTML example:**
- **<heading1> Invoice </heading1>**
- **<bold>From: Joe Bloggs <P>**
- **To: A. Another <P>**
- **Date: 2/1/1999<P>**
- **Amounted: $100 <P>**
- **Tax: 21% <P>**
- **Total $121 </bold>**

Data mixed with presentation

XML is all about data

- **XML example:**
- <Invoice>
  - <From> Joe Bloggs </From>
  - <To> A. Another </To>
  - <Date year=`1999´ month=`2´ day = `1´ />
  - <Amount unit = `$´> 100 </Amount>
  - <TaxRate> 21 </TaxRate>
  - <Total currency = `$´>121 </Total>
- </Invoice>

Human Readable

Comes with Tags

XML is all about data

- Separates data (structure) from presentation
- Same XML data can be used to render to HTML, WML (Wireless Markup Language), LaTeX, etc...
- Data interchange –– any XML–enabled application can talk to any other

Well–Formed XML

- Text must conform to certain rules to be considered ˆwell formed˜
- Matching opening and closing tags, or self–closing (`<empty />`) tags
- Tags are properly nested
- If it´s not well formed, it´s not XML

Well formed XML

Matched Tags

- **XML example:**
- **<Invoice>**
  - **<From> Joe Bloggs </From>**
  - **<To> A. Another </To>**
  - **<Date year=`1999´ month=`2´ day = `1´ />**
  - **<Amount [illegible] 100 </Amount>**
  - **<TaxRate [illegible] ate>**
  - **<Total cu[illegible]ency = `$´>121 </Total>**
- **</Invoice>**

Self closing Tags

Valid XML

- An XML document is ^valid~ if it includes a description of the tags it may contain
- We can throw out invalid XML easily
- E–commerce transactions can be checked easily (e.g. non–negative quantity on an order form)
- XML data from a foreign source can be validated before we accept it

Valid XML

- Document Type Definitions (DTDs) are the current standard
- Specify patterns for the tags contained in the XML document
- Written in a different (non–XML) format
- Weak data–typing, based on string substitution
- Other validation methods possible, with fewer drawbacks: Schema

Valid XML

XML version declaration

- <?xml version=^1.0~ ?>
- <!DOCTYPE Invoice SYSTEM ^inv.dtd~>
- <Invoice>
  - <From> Joe Bloggs </Fro
  - <To> A. Another </To>
  - <Date year=`1999´ month
    />
  - <Amount unit = `$´> 100 </Amount>
  - <TaxRate> 21 </TaxRate>
  - <Total currency = `$´>121 </Total>
- </Invoice>

Document Type Definition

Valid XML

```
■ <!ELEMENT Invoice (From, To, Date,
  Amount, TaxRate?, Total?)>
■ <!ELEMENT From #CDATA>
■ <!ELEMENT To #CDATA>
■ <!ATTLIST Date
■           year    CDATA    #REQUIRED
■           month   CDATA    #REQUIRED
■           day     CDATA    #REQUIRED>
■ <!ATTLIST Amount
■           unit (´$´|´ ´|´ ´) ´
```

Sub–element requirements

Choice list with default

XML Schema

- More functionality than DTDs
- Written in XML
- Not yet a W3C recommendation –– confusion over what a ˆSchema˜ is
- Xschema, XML Schema, XML–Data, ...

XML Standards

- XML is a World Wide Web Consortium (W3C) recommendation, v1.0 approved 10th February 1998
- `http://www.w3.org/TR/REC-xml`
- Other standards available online -- DTD, Xpointer, etc. Draft standards for XML Schema, XSL, etc. also available.
- `http://www.xml.com/xml/pub/listrescat`

Web apps and XML

(or: Why you should care about XML)

- One data source, multiple outputs
- XML   HTML is the obvious one
- XML as a data interchange format
- Inherent extensibility –– e.g. Microsoft´s Channel Definition Format
- XML is *the* format for e–commerce
- Vendor independence

XML Transformations

- The great thing about standards is there´s so many to choose from
- Need to translate to/from XML and other formats
- e.g. HTML rendering of XML in IE5
- e.g. LaTeX rendering of XML for printing
- Need a language to describe these XML transformations

XSL and XML Script

- XSL (XML Style Language) is a W3C draft proposed by Microsoft
- XML Script created by DecisionSoft as a scripting language for regression tests
- The two overlap (but we think XSL´s functionality is a subset of that provided by XML Script)

X–Tract demo

- `http://www.xmlscript.org/`
- Linux and Win32 versions (server version is currently Linux only ;–)
- We´ll convert the XML Invoice seen earlier into browser–ready HTML