

Why XML Sucks...

And what *you* can't do about it!

Kendall Grant Clark, <kendall@monkeyfist.com>

Faculty Research Associate, MIND

Managing Editor, XML.com

XML

- A fact of technical life for the foreseeable future
- Integral to “both” visions of the Web’s future: the Semantic Web and Web Services
- Lingua franca in the publishing industry

Documents v. Data

- Different needs, one format -- a mixed blessing:
 - programmers have one “thing” to learn, but only one thing to “use”
 - systematic fracture: the curse of mixed content
- How did it happen? SGML + B2B

Syntax, not Semantics

- The myth of the “self-describing” meta-language
- If you want semantics, you have to specify them, per application, or use RDF (where they’ve already been specified for you!)
- What does `<foo bar=“0”><bap/></foo>` mean exactly?

The Toolset

- Universal Syntactic Interop: wf, valid XML, Infoset, Unicode, Namespaces, RDDL
- Addressing: URI, XPointer, XLink
- CRUD: SAX, DOM, XSLT, XQuery, XPath
- Validation & Constraint: XML Schema, RelaxNG, Schematron, DTD
- Display, UI: XHTML, XForms, SVG, CSS, XSL-FO

Syntax

- It's freaking text! :)
- Well-formedness & validity
- `int`: Unicode
- Namespaces & RDDDL
- Binary Storms are Gathering :(

Addressing

- URI (plus HTTP)
- XPointer
- XLink (but, really, RDF)
- XLink is the biggest unfinished piece of business in “core XML”, if we ignore XQuery

Create, Read, Update, Delete

- Low-level APIs: DOM (standardized & language-specific), SAX, endless variations thereof
- Mid-to-high level: XSLT, XQuery, XPath
- Data binding APIs

Validation & Constraint

- W3C XML Schema: Please make it go away!
- RelaxNG: from the cognoscenti to the W3C?
"Many constraints, especially those which are on the fringe of non-deterministic models, can be expressed by RELAX NG and not by W3C XML Schema. Some combinations in document structures forbidden by W3C XML Schema can be described by RELAX NG."
- Schematron: declarative, rule-based.
- DTD: In the old days, I'd walk 5 miles to find a good DTD mode for emacs... :)

Validation, etc

- Do you really really really really need it?
- Let 1,000 flowers bloom:
 - not one schema
 - not one schema language
- Read <http://www.xfront.com/BestPracticesHomepage.html> & <http://www.xml.com/pub/a/2004/10/27/extend.html>

Display & UI

- XHTML (secret weapon: XHTML + class="foo" & id="bar")
- SVG
- XForms
- Device Independence