Internet Standards for the Web: Part II

Larry Masinter

April 1998
Outline of tutorial

- Part 1: Current State
  - Standards organizations & process
  - Overview of web-related standards
- Part 2: Recent activities
  - What’s happening with web standards?
  - What are the hard problems
What was covered in Part I?

- IETF, W3C organizations and process
- Why standards?
- Survey of basic web standards for
  - content
  - identification
  - protocols

Readings, pp. 1-21
Purpose of Part II

- Highlight some recent events
- Explain some controversial issues
- Encourage you to get involved
Categories for Web Standards

- Content (e.g., HTML)
  - kinds of objects we’re moving around?
- References (e.g., URLs)
  - how to talk about something not in hand?
- Protocols (e.g., HTTP)
  - how do things move around the net?
Content standards: highlights

- HTML, XML, and style sheets
- Active content vs protocol
- Character sets
- Page layout
- Streaming media
- MHTML
- Metadata: PICS, RDF, dSig, Dublin Core
HTML and style sheets

- ~1995-6: addition of inline presentation markup by browser vendor
  - font, size, positioning, color
- recently: introduction of standard for style sheets
  - retain structural markup (HTML) or other semantic markup (XML)
- Deployment still an issue
The debate over inline style

✓ People want it
✘ They’ll misuse it
✓ Inline style displays faster incrementally
✘ Precomputed styles
✓ It’s easier to enter inline markup
✘ Automated tools make styles just as easy
✓ “Give them rope”
✘ “They’ll hang themselves”
Style sheets

- Separate presentation information
  - `<H1>` should be bold, TimesRoman, 36 point

- Multiple styles for single document
  - print, display, handheld

- Developments
  - Cascading Style Sheets (designed for web)
  - Document Style Semantics and Specification Language (designed for SGML)
  - eXtensible Style Language (new development)
XML: SGML simplified

- Primarily: simplify SGML
- Fix up ‘naming’
- Tools just now being deployed
“Active Content”

It’s a program! It’s a script! It’s a document format!

- Create documents that embed computation that control the document’s display
  - Pros and cons for this approach
  - *Postscript does this, PDF doesn’t*

- Dynamic HTML
  - Cascading Style Sheet… plus ...
  - JavaScript (ECMAScript)
  - control points for Document Object Model (DOM)

- Java applets as a document format
Charsets: Moving to Unicode

- non-European languages
- Some issues resolved:
  - The “document character set” vs. the document’s charset
  - Internationalization of HTML
- Some still open:
  - URLs and domain names
  - deployment
    - efficiency (localization uses fewer bytes)
    - politics (issues with Korean Unicode, Vietnamese)
Page layout on the web

- Postscript
- PDF
- Challenges:
  - compressed image formats
  - XML + XSL
Streaming media

- RealAudio
- Combined protocol & content
- Multiple Codecs
MHTML

- How to send HTML in email?
  - Include the images without changing URLs
- created new “multipart/related”
  - works for more than HTML
  - doesn’t require rewrite
MetaData standards

- Dublin Core and RDF
- Ratings: PICS
- Signatures, copyright and digital property rights
Dublin Core

- How to “Catalog” the web?
- 15 common resource description elements
  - title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights
- Expression in Resource Definition Format
- Authored using WebDAV
Platform for Internet Content Selection

- **Self-rating:**
  - content providers voluntarily label

- **Third-party rating:**
  - multiple, independent labeling services
    - Services may devise their own labeling systems
    - same content may receive different labels from different services.

- **Ease-of-use:**
  - for parents and teachers; labels from multiple sources
Signatures, copyright

- Another kind of metadata
- Another kind of ‘rating’
- object-based security requires key management
Other content activities

- WebObjects:
  - Merge User & program interface
- Tuning for specific applications
  - Handheld Device Markup Language
Web References: highlights

- Unsolved problems
- URLs vs. URNs
- top-level domains
- URL guidelines
Some unsolved problems with URIs

- things go away
  - Material behind URLs disappears
- pimples.com
  - vanity domains for billboard use
- Apple Computer and Apple Music
  - conflicts over short names
- urn:hdl:MTV/I_quit
  - how does authority migrate?
- http://www.métro.paris.fr/métro
  - Non-ASCII names
URLs vs URNs

- Some URLs aren’t really “locators”
  - data:, mid:, news:
- Does the URL syntax constrain the URN syntax?
- Does the URL syntax constrain all URIs?
- Will URNs actually work?
The top-level domain issue

- vanity domains in .com
- .au? .com?
- Hierarchy is lost
- Trademark disputes
- attempt to add new ones politically sensitive
  - monopoly
  - fairness
Web protocols: highlights

- HTTP/1.1 draft standard
- HTTP-NG
- Content negotiation
- WebDAV
- Push
HTTP/1.1 Draft Standard

- Resolved over 100 “issues” with RFC 2068
- Revised ‘digest authentication’
- Newer cookies, too!
HTTP is *not* a good protocol

- HTTP/1.0 didn’t work well as web evolved
- HTTP/1.1 fixed some problems
  - backward compatibility was more important
- It still has lots of problems!
  - Don’t copy it for new protocols
  - Session Initiation Protocol, Real Time Streaming Protocol do
  - See RFC 2324: HTCP CP
HTTP-NG

- “Next Generation” design
- Not required to be compatible
- Design goals:
  - simple
  - performance
  - asynchronous operation
- use distributed object technology
Distributed objects and the web

- CORBA, DCOM designed for LANs
  - Global scaling?
  - Extensibility?
  - Caching, redirection?

- HTTP-NG aspires to bridge the gap between HTTP and distributed object protocols
Content Negotiation

- Different recipients have different capabilities
  - Cellphone
  - Reading machine
  - Print vs. display
- How to tune content for recipient?
- How to describe recipients
HTTP Content Negotiation

- Language (Accept-Language)
- Character set (Accept-Charset)
- Capabilities to handle media (Accept)
- Brand of software (User-Agent)

need more
WebDAV:
Distributed Authoring and Versioning

- Locking
- Compound objects
- Version management
- Directory management
- *WebDA finished, versioning, search language in progress*
Other web-related protocol work

- Transaction Layer Security (TLS)
  - derived from Secure Sockets Layer (SSL)
- Internet Payment
  - no clear standards yet
- Content Rating (PICS)
Other (less) related activities:

- Internet Fax
- Internet Printing (IPP)
- Directories
- Calendaring & Scheduling
- Messaging
- Chat
Tutorial Review

- Increasing Number of Organizations
  - Common goal: improve the net
- Evolution along many fronts
- Standards come after innovation
- Lots happening in many areas:
  - content
  - references
  - protocols
How to get involved?

- Inform yourself
  - All specifications are available for review
  - Standards work when everyone participates
- IETF is open
  - Contributors from all over the world
- W3C invites contributions
  - Members are vendors and implementors of the software you use
- You’re here...
Internet Standards for the Web: Part II

Larry Masinter
masinter@parc.xerox.com
lmm@acm.org

for copies of the tutorial notes or slides
please give your mailing address if you want a paper copy

April 1998