World Wide Web Infrastructure and Evolution

Roy T. Fielding

University of California, Irvine

http://www.ics.uci.edu/~fielding/

World Wide Web Project

- Goal
  - Universal access to the world of information
- Started at CERN
  - March 1989: Tim Berners-Lee
  - October 1990: Tim Berners-Lee and Robert Cailliau
  - November 1990: First prototype on NeXT
- Ended
  - March 1991: libwww and line-mode browser
- Revival
  - October 1991: www-talk mailing list initiated

Simplicity

- How do we define it?
  - The ability to comprehend the technology
  - Applies to both development and use
- Why is it significant?
  - Enables a Low Entry Barrier
  - Reduces fear of change

Extensibility

- How do we define it?
  - The ability to extend functionality when the technology is already in use
  - Applies to both development and use
- Why is it significant?
  - Prevents us from being stuck at the Low Entry Barrier
  - Enables collaborative work on divergent projects/goals

Conclusions

- Designing for Simplicity and Extensibility allowed the Web to succeed and evolve
  – for both developers and users
  – for collaboration on a global scale
- Can be seen in the Web Infrastructure
  – Protocols: URI, HTML, HTTP
  – Applications: Clients and Servers
- but not always in the Implementations

Uniform Resource Identifiers

- Web Addresses
  - Universal Document Identifiers
  - Universal Resource Identifiers
  - Uniform Resource Identifiers
    - URL: Locators
    - URN: Names

  http://www.ics.uci.edu/~fielding/talks/

- No significant change in 5 years!

URI Simplicity

- Just a string of common characters
  - – chosen for transcribability
- Scheme defines interpretation and structure

  <scheme>:<scheme-specific-part>
- Hierarchical when desired [Engelbart]

  <scheme>://<site>/<pathname>
- Name components may be meaningful

  http://www.ics.uci.edu/~fielding/talks/

URI Extensibility

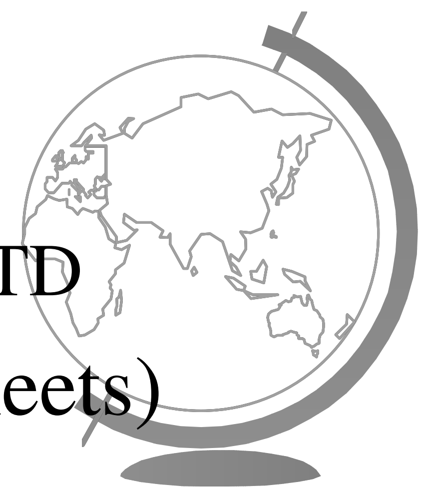
- URI scheme allows new types to be defined without affecting old uses
  - file, news, http, telnet, gopher, wais, ftp, …
- Naturally lends itself to table-driven implementations
  - browser uses table to select handler
  - handler can be dynamically loaded
- Unfortunately,
  - some people haven’t figured that out yet

Hypertext Mark-up Language

- Simple mechanism to structure text for presentation and hyperlinks
- Based on SGML, but
  - with an assumed DTD (grammar rules)
  - with unrecognized tags ignored
  - without entity declarations
  - without marked sections
- HTML sacrificed extensibility for simplicity

Why not just SGML?

- SGML sacrificed simplicity for extensibility
  - poor (and expensive) documentation
  - designed by committee, written by a lawyer
- Huge Entry Barrier
  - takes a year to develop a complete parser
  - less than 10 people truly understand it
- Not designed for on-line rendering
  - equivalent to running yacc for each DTD
- Doesn’t define presentation (stylesheets)

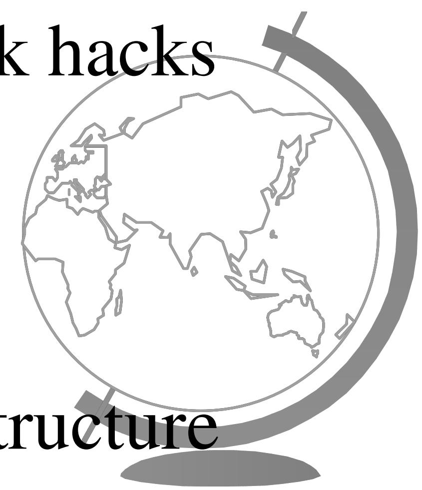


HTML Evolution

- Level 0: pre-1992
  - structure (P, H1-6, DL, UL, PLAINTEXT)
  - hypertext (A, BASE)
- Level 1: 1992-mid`93
  - SGML rules restored (Dan Connolly)
  - Character highlighting added
  - IMG tag added (Marc Andreessen)
- Level 2: late 1993
  - FORM tags added (Eric Bina, Marc A.)

HTML Evolution

- Level 3: end of 1993 (HTML+)
  - Generational redesign (Dave Raggett)
  - Better structure, tables, math
  - Lost in the rush to commercialize
- Netscapisms: Fall 94 - today (MS-isms)
  - Low-level presentation and other quick hacks
- Version 3.2: last week
- Stylesheets: this summer
  - presentation separated from element structure



Hypertext Transfer Protocol

- A protocol (syntax and semantics) for transferring information
  - usually across the Internet using TCP
- Design goals
  - speed (stateless, cachable, few round-trips)
  - simplicity
  - extensibility
  - data (payload) independence
- Has evolved considerably

HTTP/0.9 (pre-1993)

- Absolute Simplicity

```
GET /url-path
<TITLE>Hello World</TITLE>
Hello World
```

- No Extensibility
  - only one method (GET)
  - no request modifiers
  - no response metainformation

HTTP/1.0 (1993-present)

- Simple and (mostly) Extensible

```
GET /Test/hello.html HTTP/1.0
Accept: text/html
User-Agent: GET/5 libwww-perl/0.40

HTTP/1.0 200 OK
Date: Fri, 12 Jan 1996 01:02:49 GMT
Server: Apache/1.0.5
Content-type: text/html
Content-length: 38
Last-modified: Wed, 10 Jan 1996 01:

<TITLE>Hello</TITLE>
Hello out there!
```

HTTP Future

- HTTP/1.1 (June 1996)
  - persistent connections
  - control over caching, proxies
- HTTP/1.2 (late 1996)
  - content negotiation
  - extensible security framework
- HTTP/2.0 (late 1996)
  - multiplexing requests and responses

Lessons Learned

- URI
  - Simple and Extensible from the start
  - Immediate and lasting success
- HTML
  - Simple, but lost its Extensibility
  - Subject to hacking wars, fate unknown
- HTTP
  - Originally Simple, but not Extensible
  - WWW took off when extensibility was added