World-Wide Web Software
An Insider’s View

Roy T. Fielding

<fielding@ics.uci.edu>
University of California, Irvine

Bay Area Roundtable (BART)
January 1996

Overview

What does "World-Wide Web" mean?

- **The Project**
- **The Protocols**
- **The Applications**
- **The Information**

<break>

How the Web Evolved (by Design)

Application Architectures: Current vs Future

What does "World-Wide Web" mean?

The Project

- **Collaboration on a Global Scale**

The Protocols

- **URI, HTML, HTTP**

The Applications

- **Clients and Servers**
- **Browsers, Agents, Proxies, Gateways**

The Information

- **For every topic, there is at least one fanatic**

The Project

I don’t want the Web to constrain what people do: the Web is not there to constrain society. It’s there to model society in its completeness, in its entirety.

Tim Berners-Lee (GNN Interview)

Mar 1989: CERN (Geneva)

- **Tim Berners-Lee submits first proposal for the World-Wide Web**

Oct 1990: CERN

- **Tim Berners-Lee and Robert Cailliau submit revised Web proposal**

The Project

Society is a web of social relationships

- **people, products, tasks, hobbies**

If we associate people with their work...

- **we learn more about both**

If we associate people with other people and their work...

- **we establish a web of commonalities...**
- **which forms the basis for a community**

Communities build rich information spaces

The Project

Nov 1990: CERN

- **First prototype Web browser runs on NeXT**

Jan 1991: Stanford SLAC and NIKHEF

- **First HTTP servers outside CERN**

Mar 1991: CERN

- **libwww and line-mode browser released**

Aug 1991: Netnews

- **WWW announced on hypertext groups**

Oct 1991: E-mail

- **www-talk archived mailing list started**

Early Web Technology (pre-1992)

Web Addresses
(a.k.a. Universal Document Identifiers)

- now called Uniform Resource Locators
- schemes: file, news, http, telnet, gopher
- no significant change in 5 years

Hypertext Mark-up Language: HTML Level 0

- structure elements (P , H1-H6, DL, UL, PLAINTEXT)
- hypertext anchors (A, BASE)

Hypertext Transfer Protocol: HTTP/0.9

- only allows retrieval (GET /path)
- always returns HTML

The Project

Jan 1992: Berkeley

- **Pei W ei demos V iola scriptable browser**

Jul 1992: Berkeley , MIT

- **New V iola, tkWWW and Erwise browsers**

Nov 1992: Stanford SLAC

- **MidasWWW released**

Jan 1993: NCSA 50 servers

- **First alpha release of Mosaic for X**

Feb 1993: NCSA

- **Marc Andreessen proposes the IMG tag**

The Project

Mar 1993: U Indiana, U Kansas

- **Marc V anHeyningen finishes Perl httpd**
- **Lou Montulli webifies L ynx text browser**

Apr 1993: NCSA

- **XMosaic 1.0, with IMG support**
- **Rob McCool announces NCSA httpd 0.3**

May 1993: CERN, BSDI, Honolulu

- **CERN makes libwww public domain**
- **Tony Sanders releases Plexus httpd**
- **Kevin Hughes produces hypermedia CWIS**

Early Web Technology (1992 - mid'93)

Uniform Resource Locators: URL

- added schemes: wais, ftp

Hypertext Mark-up Language: HTML Level 1

- Dan Connolly restores SGML rules
- Character highlighting added
- Marc Andreessen adds IMG tag

Hypertext Transfer Protocol: pre HTTP/1.0

- proposal for HTTP/1.0 to use MIME format/names
- Rob McCool begins work on NCSA httpd

The Project

Jun 1993: Cornell **134 Servers**

- **Thomas Bruce announces Cello browser**

Aug 1993: Cambridge, Mass.

- **WWW Wizards Workshop, 40 attendees**

Sep 1993: NCSA **500 Servers**

- **NCSA httpd 1.0 released**

Nov 1993: NCSA, UC Irvine

- **Mosaic 2.0 for X, Mac, and Windows**
- **MOMspider project begins**

Dec 1993: NY Times Business Front Page

Early Web Technology

mid 1993: HTTP/1.0 Defined

- messages now include metainformation
- enables data independence and protocol extensibility

late 1993: HTML Level 2

- FORMs capability added
- allows restricted interaction with remote services

end 1993: HTML+ proposed

- Dave Raggett's proposal to extend HTML
- provides better structuring, tables, math
- lost in the rush to commercialization

The Project

Jan 1994: UC Irvine

- **wwwstat log analyzer released**

Mar 1994: Mountain V iew 2000 Servers

- **Mosaic (later Netscape) Communications founded by Clarke and Andreessen**
- **WWW byte traffic surpasses Gopher , doubles every four months for over a year**

May 1994: CERN

- **First International WWW Conference with 380 attendees**
- **libwww-perl project genesis**

The Project

Jun 1994: UC Irvine **3500 Servers**

- **libwww-perl released**

Oct 1994: Chicago **7000 Servers**

- **Second International WWW Conference with over 1500 attendees**

Dec 1994: Mountain View, MIT **12000 Servers**

- **Netscape Navigator 1.0 released**
- **First meeting of W3 Consortium**

Feb 1995: Europe

- **CERN drops WWW Project to Inria**

The Protocols

Uniform Resource Identifiers (URI)

Hypertext Mark-up Language (HTML)

Hypertext T ransfer Protocol (HTTP)

URI

Uniform Resource Identifiers

- **Uniform Syntax**
  - scheme://site/path
  - http://www .ics.uci.edu:8000/T est/hello.html
- **Identifies**
  - separates identification from action
- **Resources**
  - documents
  - services
- **Relative Addressing**
  - /Test/hello2.html
  - hello2.html
  - ../IRUS/

HTML

Hypertext Mark-up Language

- Based on SGML (well, almost)
- Simple

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

- Hypertext

```
<TITLE>Hello</TITLE>
Hello out <A href="there.html">
there!</A>
```

- Extensible (at first)

```
<TITLE>Hello</TITLE>
<SND src="hello.au">
Hello out <A href="there.html">
there!</A>
```

HTTP

Hypertext T ransfer Protocol

Design goals:

- **Speed**
- **Simplicity**
- **Extensibility**
- **Data Independence**

HTTP Example

```
GET /Test/hello.html HTTP/1.0
Accept: */*
User-Agent: GET/0.5 libwww-perl/0.40

HTTP/1.0 200 OK
Date: Fri, 12 Jan 1996 01:02:49 GMT
Server: Apache/1.0.0
Content-type: text/html
Content-length: 38
Last-modified: Fri, 12 Jan 1996 00:40:56 GMT

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

HTTP Speed

Emphasis is on immediacy of traversal response across a global network

- **stateless**
- **use response caches**
- **send important info first**
- **reduce round-trips**
- ***but without any transport hacks***

HTTP Simplicity

- **ASCII control structure**
  - **Readability via TELNET is a feature!**
- **Mail entity reuse**
  - **header containing metainformation**
  - **body containing data**
- **Request/response sequence**
  - **no asynchronous or interleaved responses**
  - **pipelined requests are okay**
- **Few requirements for minimal compliance**
  - **it should still be possible to implement a minimal server in a single day’s work**

HTTP Extensibility

- **Methods**
  - **each method defines an operation to be applied by a resource**
- **Uniform Resource Identifiers (URI)**
  - **resources are identified via the Request-URI**
- **Header Fields**
  - **general header fields give information about the message**
  - **request header fields assign parameters to the request**
  - **response header fields characterize the response**
  - **entity header fields provide metainformation**
- **Response Status Codes**
  - **3 digits indicate status independent of language**

HTTP Data Independence

- **Transfer Codings**
  - **define encodings applied to the message entity for transfer**
- **Content Codings**
  - **define encodings applied by the resource over media type**
- **Media T ypes**
  - **define the format/purpose of the enclosed data entity**
- **Character Sets**
  - **define the character encoding for textual media types**
- **Language T ags**
  - **define the natural language(s) for which the entity is intended**

The Applications

Clients and Servers

Browsers

Agents

Proxies

Gateways

Origin Servers

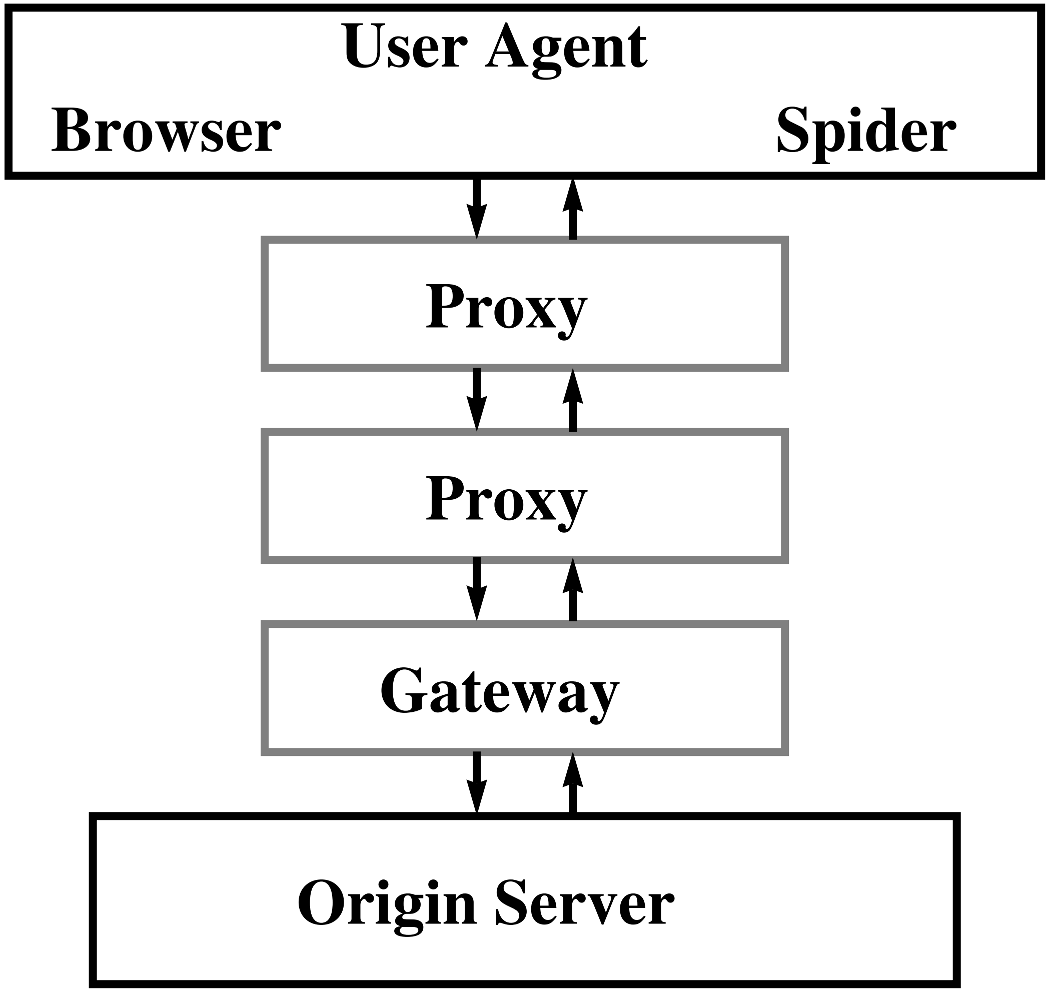
The Information

It's as if the Library of Congress had exploded in midair. There's all kinds of information strewn over the countryside, approximately laid out according to its logical relationship.

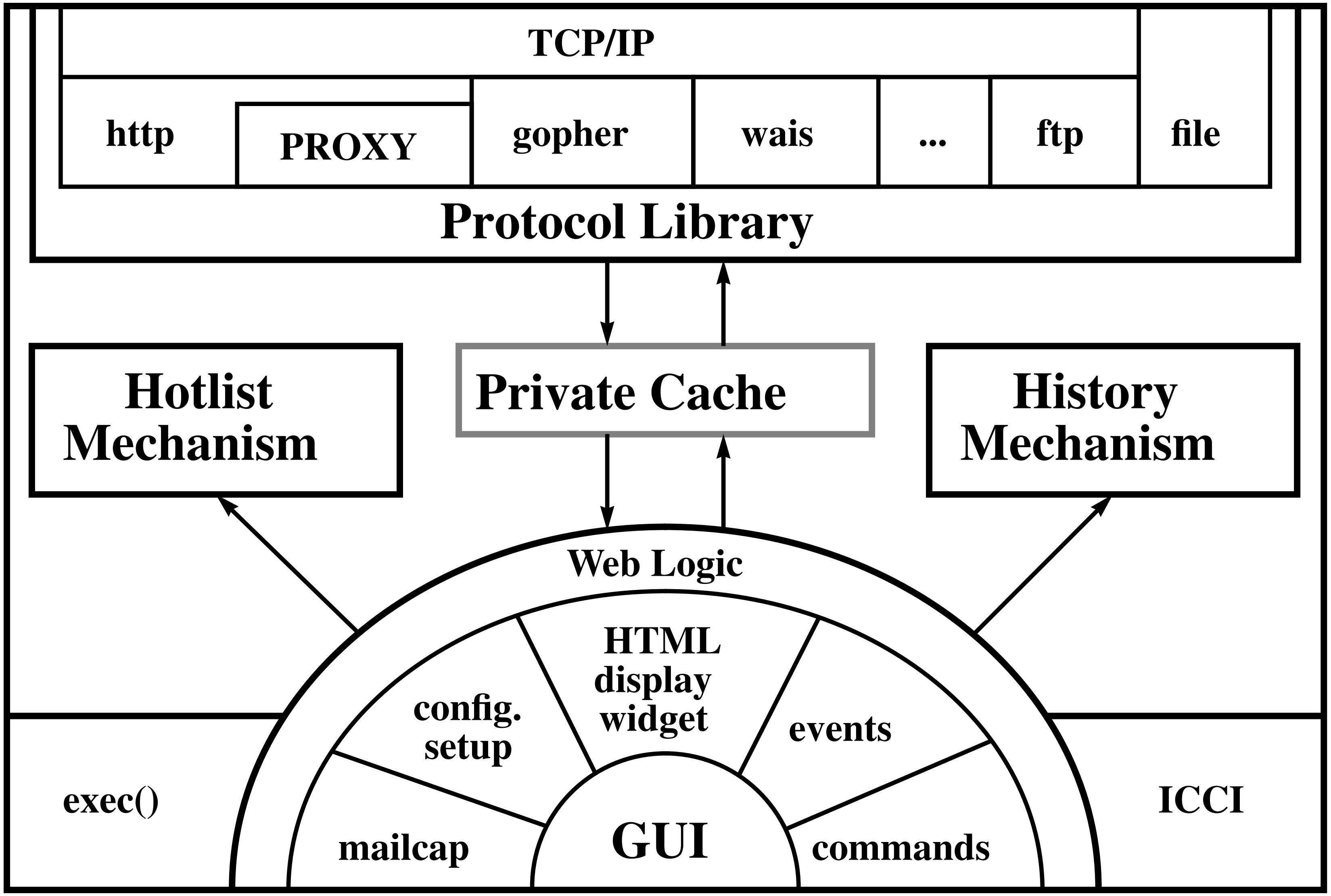
David Brooks (comp.infosystems.www.misc, 2/95)

For every topic, no matter how obscure, there exists at least one fanatic with the undying need to tell you all about it

Application Architectures



Current Browser Architecture



Future Browser Architecture