

World Wide Web

Instructor: Dmitri A. Gusev

Fall 2007

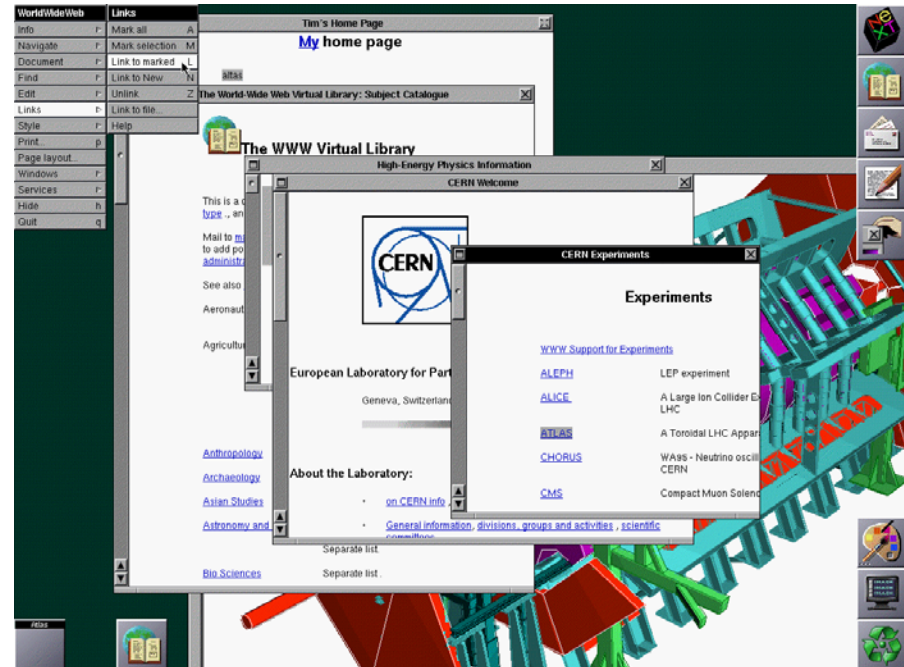
CS 502: Computers and Communications Technology

Lecture 24, December 3, 2007

A Little History

<http://www.w3.org/History.html>

- Mar. 1989: Tim Berners-Lee, "Information Management: A Proposal". "It then summarises my short experience with non-linear text systems known as "**hypertext**", describes what CERN needs from such a system, and what industry may provide. Finally, it suggests steps we should take to involve ourselves with hypertext now, so that individually and collectively we may understand what we are creating. ...The actual observed working structure of the organisation is a multiply connected "**web**" whose interconnections evolve with time."
- Oct. 1990: Tim Berners-Lee starts work on a hypertext GUI browser+editor using the NeXTStep development environment. He makes up "**WorldWideWeb**" as a name for the program. WorldWideWeb was written in Objective-C.
- 1994: WWW consortium (CERN, MIT)
<http://www.w3.org/>



The first browser screenshot
(taken in 1993)

Network Protocols

Internet Protocol (IP)

Transmission Control
Protocol (TCP)

User Datagram
Protocol (UDP)

Simple Mail Transfer
Protocol (SMTP)

File Transfer Protocol (FTP)

Telnet

Hypertext Transfer Protocol (HTTP)

The Client Side

- WEB documents (pages) connected via hyperlinks (hypertext).
- Hyperlinks: highlighted strings in text or images.
- Browsers (text-based or graphical): tools for navigating the WEB.
- Forms and Java applets

Network Addresses

Computers on the Internet are identified by their *IP addresses*. An IP address is displayed a series of four decimal numbers separated by dots.

Example: 192.168.1.101

(looked up by ipconfig)

An IP address is stored in 4 bytes, and each number in the IP address is in the range 0 to 255.

Classes of Networks

Class	Bytes used for the network address	Bytes used for the host number
A	1	3
B	2	2
C	3	1

Domain Name System

Example: The hostname *cadfacstaff3.facstaff.ccsu.edu* corresponds to the IP address *149.152.20.143* (found using nslookup).

Computer name: *cadfacstaff3*

Domain name: *facstaff.ccsu.edu*

ICANN: The Internet Corporation for Assigned Names and Numbers,

<http://www.icann.org/>

IANA: The Internet Assigned Numbers Authority,

<http://www.iana.org/>

Top-Level Domains (TLDs)

- .com, .net, .org, .edu, .int, .mil, .gov;
- .arpa for Internet infrastructure purposes:
 - in-addr.arpa is used for reverse DNS lookup for IP addresses
Advanced Research Projects Agency → Address and Routing Parameter Area (a "backronym")
- In 2000, ICANN introduced seven new gTLDs (generic TLDs): .aero, .biz, .coop, .info, .museum, .name, and .pro.
- .jobs and .travel approved 04/2005
 - katrina.jobs (refers to CareerBuilder.com)
 - <http://www.egypt.travel>
- .cat approved 09/2005
- .eu approved 12/2005
- .tel approved 05/2006, added 03/2007
- .mobi available since 09/2006
- .asia approved 10/2006

See <http://icann.org/faq/>

<http://www.internic.net/whois.html>

<http://www.nic.name/>

Country codes: .au, .br, .ca, etc., see

<http://www.uwhois.com/cgi/domains.cgi?User=NoAds>

Steps of fetching

<http://www.w3.org/History.html>

1. The browser determines the URL
2. The browser asks the local DNS (Domain Name System) server for the IP (Internet Protocol) address
3. DNS replies with 18.23.0.23.
4. The browser makes a TCP (Transmission Control Protocol) connection to port 80 on 18.23.0.23
5. It then sends a *GET /hypertext/WWW/History.html*
6. The *www.w3.org* server send the file *History.html*
7. The TCP connection is released
8. The browser displays all the text in *History.html*
9. The browser displays all the images in *History.html* (new TCP connection for each image)

HTML

- *Hypertext Markup Language (HTML)* is a standardized markup language specifying how the documents are to be formatted and reformatted
- *Tags* indicate how the information should be displayed
- *Hyperlinks:*
 - ` NASA's home page `
 - ` `

HyperText Transfer Protocol (HTTP)

- Simple (GET without the protocol version) and full requests
- Methods (commands)
- GET: request to read a Web page encoded in MIME (Multipurpose Internet Mail Extensions - adding a header to describe the encoding)
- HEAD: request to read a Web page header
- PUT: request to store a Web page (may include authentication headers)
- POST: request to append new data to a Web page (e.g. posting a message to a news group)
- DELETE: request to delete a Web page (may include authentication headers)
- LINK: Connects two existing pages
- UNLINK: breaks an existing connection between two pages

Example of HTTP protocol

C: telnet www.w3.org 80

T: Trying 18.23.0.23 ...

T: Connected to www.w3.org

T: Escape character is '^]

C: GET /hypertext/WWW/TheProject.html HTTP/1.0

HTTP/1.0 200 Document follows

MIME-Version: 1.0

Server: CERN/3.0

Content-Type: text/html

Content-Length: 8247 <HEAD><TITLE> The World Wide Web
Consortium (W3C)</TITLE><HEAD>

<BODY>

<H1><IMG ALIGN=MIDDLE ALT="W3C"
SRC="Icons/WWW/w3c_main.gif">

The World Wide Web Consortium </H><P>

The World Wide Web is the universe of network-accessible
information.

...

</BODY>