

CSC 551: Web Programming

Fall 2001

See online syllabus at:

www.creighton.edu/~csc551

Course goals:

- understand the technology and protocols underlying the World Wide Web
- become familiar with common tools and techniques for developing Web-based applications, both client-side and server-side
- develop a working knowledge of HTML, JavaScript, and Java as languages for developing Web applications

Reasonable questions

What is the World Wide Web?

Is it the same thing as the Internet?

Who invented it?

How old is it?

How does it work?

What kinds of things can it do?

What does it have to do with programming?

Web ≠ Internet

Internet: a physical network connecting millions of computers using the same protocols for sharing/transmitting information (TCP/IP)

- in reality, the Internet is a network of smaller networks

World Wide Web: a collection of interlinked multimedia documents that are stored on the Internet and accessed using a common protocol (HTTP)

Key distinction: Internet is hardware; Web is software

Many other Internet-based applications exist

e.g., email, telnet, ftp, usenet, Instant Messenger, Napster, ...

History of the Internet

the idea of a long-distance computer network traces back to early 60's

- Licklider at M.I.T.
- Baran at Rand
- National Physics Laboratory in U.K.

in particular, the Department of Defense was interested in the development of distributed, decentralized networks

- survivability (i.e., network still functions despite a local attack)
- fault-tolerance (i.e., network still functions despite local failure)

contrast with phone system, electrical system

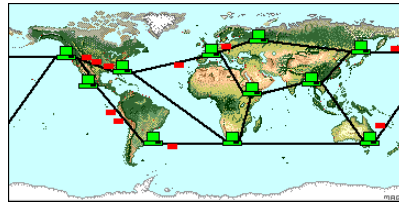
in 1969, Advanced Research Project Agency funded the ARPANET

- connected computers at UCLA, UCSB, SRI, and Utah
- allowed researchers to share data, communicate
56KB/sec communications lines (vs. 110 B/sec over phone lines)

Internet architecture

the Internet is distributed and decentralized

- each computer on the network is given a unique IP address
e.g., bluejay is 147.134.2.20
- each message to be transmitted is broken into packets (~1500 chars)
- each packet is labeled with its source & destination addresses, packet #
- individual packets are transmitted independently
- special purpose computers on the Internet called *routers* read the destination address on each packet and pass it on
- when the packets reach their destination, they are reassembled



Transmission Control Protocol (TCP):
specifies how to break up and reassemble the packets

Internet Protocol (IP):
specifies how messages are labeled and routed

Internet growth

throughout the 70's, the size of the ARPANET doubled every year

- decentralization made adding new computers easy
- ~1000 military & academic computers connected by 1984

in 80', U.S. government took a larger role in Internet development

- created NSFNET for academic research in 1986
- ARPANET was retained for military & government computers

by 90's, Internet connected virtually all colleges & universities

- businesses and individuals also connecting as computing costs fell
- ~1,000,000 computers by 1992

in 1992, control of the Internet was transferred to a non-profit org

- Internet Society: Internet Engineering Task Force
Internet Architecture Board
Internet Assigned Number Authority
World-Wide-Web Consortium
...

Internet growth (cont.)

Stats map of the Internet (Internet Valley, Inc.)

Date	Hosts	Domains	
July 98	36,739,000	4,300,000	A host is a computer that is connected to the Internet
July 97	19,540,000	1,301,000	
July 96	12,881,000	488,000	e.g., bluejay
July 95	6,642,000	120,000	
July 94	3,212,000	46,000	A domain is a subnetwork of computers on the Internet
July 93	1,776,000	26,000	
July 92	992,000	16,300	
July 89	130,000	3,900	e.g., creighton.edu
July 81	210		
1969	4		

Internet growth (cont.)

The Internet continues to grow at an amazing rate

www.netsizer.com provides up-to-date size estimates

Monthly Average Hosts in Millions			
Month	1999	2000	2001
Jan	44.2133	70.1557	100.464
Feb	48.0415	72.8768	103.959
Mar	50.4565	75.0223	107.187
Apr	53.4396	77.0664	109.828
May	55.7750	79.6372	112.072
Jun	57.3117	82.0340	114.804
Jul	59.2499	84.7899	117.288
Aug	61.3151	87.4436	-
Sep	63.2834	90.2814	-
Oct	65.5849	92.9145	-
Nov	67.2571	95.1408	-
Dec	68.7377	97.7584	-

As of July 2001,

~430 million Internet users

~117 million Internet hosts

Internet traffic

as projected by Larry Roberts, Internet traffic volume will exceed voice traffic volume (phone lines) by 2003



History of the Web

the idea of hypertext (cross-linked and inter-linked documents) traces back to Vannevar Bush in the 1940's

- online hypertext systems began to be developed in 1960's
 - e.g., Andy van Dam's FRESS, Doug Englebert's NLS
- in 1987, Apple introduced HyperCard

in 1989, Tim Berners-Lee at the European Particle Physics Laboratory (CERN) designed a hypertext system for linking documents over the Internet

- designed a (Non-WYSIWYG) language for specifying document content
 - which evolved into HyperText Markup Language (HTML)
- designed a protocol for downloading documents and interpreting the content
 - which evolved into HyperText Transfer Protocol (HTTP)
- implemented the first browser (text-based)

the Web was born!

History of the Web (cont.)

the Web was an obscure, European research tool until 1993

in 1993, Marc Andreessen (an undergrad at the University of Illinois' National Center for Supercomputing Applications) developed Mosaic, the first graphical Web browser

- the intuitive, clickable interface made hypertext accessible to the masses
- made the integration of multimedia (images, video, sound, ...) much easier
- Andreessen left NCSA to found Netscape in 1994
cheap/free browser popularized the Web (75% market share in 1996)
in 1995, Microsoft came out with Internet Explorer
Netscape bought by AOL in 1999 for \$10 billion in stock

today, the Web is the most visible aspect of the Internet

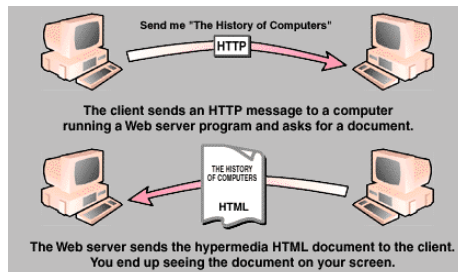
Web architecture

the Web is a giant client/server system

- Web server stores Web pages and acts as a broker
- Web browsers are the clients that request pages from the appropriate server and display them with proper formatting

e.g., when you type in a URL or click a link in a Web page

- a request is sent to the server that stores that page
- the server transmits the text of that page to the client machine
- the browser interprets the HTML code and displays text/images/...



HyperText Transfer Protocol (HTTP):
specifies how messages between the browser and the Web server are formatted

HyperText Markup Language (HTML):
specifies the content of the documents using tags (e.g., <center>)

Web growth

Stats map of the Internet & Web (Internet Valley, Inc.)

Date	Hosts	Domains	Web Sites	Web Sites/Hosts
July 98	36,739,000	4,300,000	4,279,000	12.000 %
July 97	19,540,000	1,301,000	1,200,000	6.200 %
July 96	12,881,000	488,000	300,000	2.300 %
July 95	6,642,000	120,000	25,000	0.400 %
July 94	3,212,000	46,000	3,000	0.100 %
July 93	1,776,000	26,000	150	0.010 %
July 92	992,000	16,300	50	0.005 %
July 89	130,000	3,900		
July 81	210			
1969	4			

recent estimates suggest 30-40 million Web sites, with 3-4 billion Web pages!

Static Web pages

most Web pages are *static* (contents are the same each time it is accessed)

- text/links/images are displayed in a page

e.g., online documents, most homepages

HyperText Markup Language (HTML) is used to specify text/image format

```
<html>
<head>
  <title>Greetings Page</title>
</head>
<body>
  <center>
    
    <p>
      <font color="red">Welcome to the
        World Wide Web</font>
    </center>
  </body>
</html>
```



Dynamic Web pages

as the Web moves towards online services and e-commerce, Web pages must also provide *dynamic* content

- pages must be fluid, changeable (e.g., rotating banners)
- pages must be able to react to the user's actions
- pages must be able to request and process information
- even desirable to store customer preferences and tailor services

e.g., Amazon.com

generating dynamic Web pages requires programming

- this course is about applying your programming skills to the development of dynamic Web pages and applications
- new issues arise due to the client-server nature of the Web
 - does computation occur on the client or the server?
 - portability? security? privacy?

Client-side programming

JavaScript

- a scripting language for Web pages, developed by Netscape in 1995
- uses a C++/Java-like syntax, so familiar to programmers
- simpler (not strongly typed, dynamic variables, primitive objects, ...)
- JavaScript code is embedded directly in HTML (interpreted by browser)
- good for adding dynamic features to Web page, controlling forms and GUI
- see <http://www.creighton.edu/~csc551/JavaScript/dice.html>

Java applets

- can define small, special-purpose programs in Java called applets
- provides full expressive power of Java (but more overhead)
- applets are included in Web pages using special HTML tags
- interpreted by the Java Virtual Machine embedded in the browser
- good for more complex tasks or data heavy tasks, such as graphics
- see <http://www.creighton.edu/~csc551/Java/Monte.html>

Server-side programming

CGI programming

- programs can be written to conform to the *Common Gateway Interface*
- such CGI programs reside on the Web server
- when a Web page executes a submit, data from the page is sent as input to the CGI program
- the CGI program executes on the server, sends its results back to the browser in the form of a Web page

- good if computation is large/complex or requires access to private data
- see <http://www.creighton.edu/~csc551/CGI/grades.html>

Active Server Pages (Microsoft)

Java Servlets (Sun)

Server Side Includes (Apache)

- vendor-specific alternatives to CGI
- provide many of the same capabilities but using HTML-like tags

In-class exercise

pick some of your favorite Web sites and try to identify

- static components?
- dynamic components?
 - client-side? JavaScript? Java applet?
 - server-side? CGI? ASP?

e.g., www.creighton.edu/~csc551

www.creighton.edu

www.thehungersite.com

At-home exercise

a 1998 survey estimated that 1 in 5 television commercials display a Web address (more during late night and sporting events)

- watch *at least* ½ hour of TV and pay attention to the commercials
- for next week, report
 1. the stations and times you watched (so I can filter out duplicates) the
 2. number of commercials with Web addresses
 3. the total number of commercials

we will combine the data accumulated by the class and come up with our own estimate

Next week...

basic HTML

- hypertext & hypermedia
- tags
- text formatting
- hyperlinks
- style sheets

Read Chapters 2-6

Be prepared for quiz on

- today's lecture (moderately thorough)
- the reading (superficial)