CGT 141/CPT 141 Lecture 1 Wk 1
Introduction to the Course, the Web, & Overview of HTML, XHTML and History

Introduction to Course

- About Professor
- About TAs
- Course Syllabus
    - Meeting times
    - Text(s)
    - Equipment
    - Course Administration and Attendance
    - Resources: Reference Articles and Q&A
    - Assignments
        - Exercises
            - Due period assigned.
        - Projects
        - Exams
    - Course is based on 60% Application, 40% "Theory"

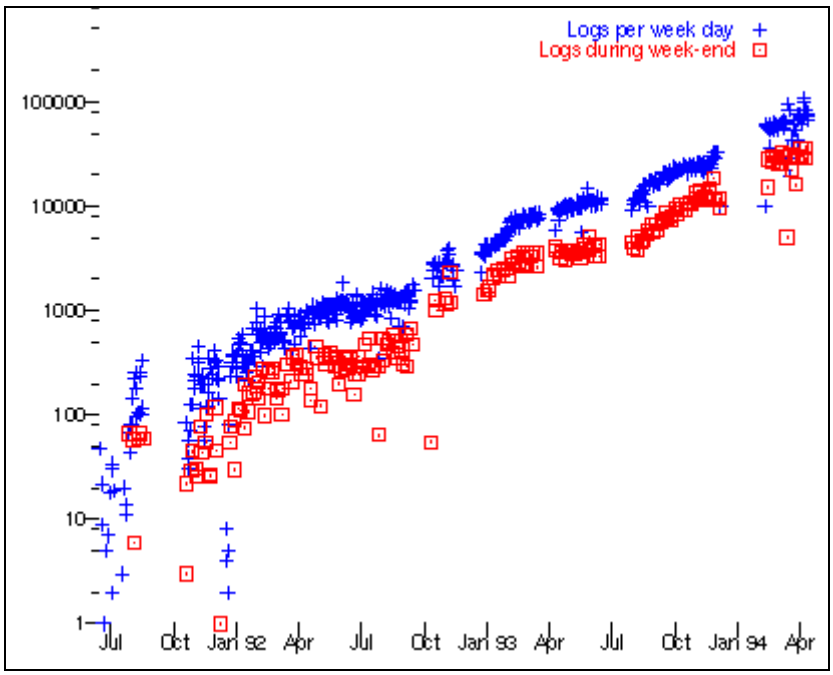
Overview of HTML, xHTML and History

Introduction to the Web (Source: http://www.w3c.org/history.html)

- 1941-45
    - Vannevar Bush
        - Associated with studies in programmed learning, shortly after WWII…much of what is done in education today such as learning objectives, learning goals and methods of instruction emerged in an effort to train the military (relates to objectives/goals on your syllabus). Needed ability to train people quickly and efficiently.
        - Wrote an article about a photoelectrical-mechanical device called Memex, short for memory extension, which could make and follow links between documents on microfiche.
            - Remember Microfiche? Similar to that concept
            - First sighting of hyperlinks

- 1957
    - Advanced Research Projects Agency (ARPA) launched in response to Sputnik. ARPA began focusing on computer networking and communications technology and who developed the first US satellite in 18 months of its formation.
- 1960s
    - Doug Engelbart
        - He prototypes the "oNLine System" (NLS) that does hypertext browsing editing, email, and so on. He invents the mouse for this purpose.
    - Ted Nelson
        - Coined the terms hypertext and hypermedia in his writing.
        - Envisioned a computer-based system that was **platform independent**, **universal** and **multi-protocol**.
        - However, computing technology at the time did not reach these goals totally.
- 1972
    - ARPANET
        - First public demonstration of ARPANET by the US Department of Defense… continued from 1957 above.
- 1989
    - Timothy-Berners Lee
        - Worked at CERN (European Organization for Nuclear Research) where he wrote a program to link between arbitrary nodes.
        - Wrote a proposal for a hypertext communications system (titled, *Information Management: A Proposal*).
        - Developed his hypertext system on the now defunct NextStep development environment, which he called the “WorldWideWeb.”
- 1993
    - Marc Andreessen
        - As a graduate project, he developed Mosaic for X Windows – the first browser – which was release to the public in February of 1993.
        - Worked at the National Center for Supercomputing Applications (NCSA) at the University of Illinois and did a seminar at CERN concerning Mosaic.
        - Left NCSA in 1994 to create Mosaic Communications Corp., which eventually became Netscape Communications Corp.

- 1994-1995
    - Commercialization!
    - The first W3C meeting was held in December of 1994
    - In June of 1994, this was the load on the first web server (info.cern.ch) was 1000 times what it was 3 years earlier:



Advantages of the Web

- Global – can be reached from anywhere
- Universal – covers everything imaginable
- Time-Independent – Always there
- Dynamic – change each time you're there
- Real-time information resource – library? What is a library?...

Disadvantages of the Web

- Connection/Access
    - 97% of world still not connected as of 2001
    - More than 50% of American households are connected as of Feb 2002 (CNN)
    - The USA makes up 29% of the world's internet users as of Dec 2002
    - Many underdeveloped nations still without access
    - Global internet population grows an average of 4% each year
    - For more information – http://www.nielsen-netratings.com/pr/pr_030220.pdf
- Cost for access and space/services
- Speed/Bandwidth
- Technology Compatibility
- Content/Author Validity
- Universality

What is the Web?

- Subset of the Internet
- Composed of various protocols
  - A protocol is a set of rules for data transfer and communication
- Internet includes:
  Web – HTTP or Hypertext Transfer Protocol
  - FTP – File Transfer Protocol
  - Gopher – pretty much defunct
  - SMTP, DHCP, POP, IMAP, Other...

What Makes the Web Work?

- Masses of interconnected computers
- Packets of information are sent and received using TCP/IP
  - Composed of two protocols:
    - Transmission Control Protocol – data assembling/disassembling
    - Internet Protocol – Routing
    - PC – route – PC
      TCP – IP – TCP

- Using Internet Protocol Addresses (IP Addresses), each computer on the Internet is identified by 4 sets of numbers separated by periods.
  - IPv4 uses 32-bit numeric address
  - Each component can be a number from 0-255
  - Example: 128.45.12.1 or 255.255.255.255 or 67.12.1.0
  - IP addresses running out???
    - IPv6 is gradually replacing the old system with a classless scheme called CIDR – Classless Inter-Domain Routing
      - CIDR – A single IP address can be used to designate many unique IP addresses
      - These IP addresses end with a slash and a number
        - 128.45.12.1/15
      - CIDR is also called supernetting

- Domain Name Servers associate a Domain Name to an IP address
  - IP Addresses and Domain Names are linked together in "master computers" called Domain Name Servers (DNS Servers).
  - Enables English representation of computers addresses, rather than numerical
  - Example: Domain Name – tech.purdue.edu

- Universal Resource Locators (URLs) are the total address needed to access a web

page. They differ slightly from the raw Domain Name.
    - Differs from the Domain name in that it includes potential prefixes and may include a path to data.
    - Example: URL – http://www.tech.purdue.edu/
    - Thus, in the URL http://www.tech.purdue.edu/cgt/courses/index.html
        - tech.purdue.edu is the domain
        - cgt/courses/index.html is the path and file name
- URLs can include:
    - HTTP://
    - FTP://
    - GOPHER://
    - Or IP Addresses

What Does the Browser Do?

- Browser Looks Up the Server
    - When a URL is entered, the browser hits a DNS server to determine the IP address of the appropriate web location.
- Browser Requests Data
    - The browser then requests information from the HTTP server, given the location identified by the DNS.
- Server Sends Content
    - In "chunks" or packets across the Internet. Various packets may traverse different paths to the user's computer. TCP is used on server end to break up the data and on the user's end to reassemble it. IP takes care of routing the data from server to user.
- Content Is Received
    - On the user's end, when data is received it is assembled and stored in a special directory on the user's computer, often called a cache.
- Rendering of Content
    - Once the content is available, the user's browser then renders/composites the graphics and text on the fly. (HTML, JPG, GIF, & PNG elements).
    - If scripting is encountered, the browser uses internal engine (if available) to parse the information (JavaScript, VBScript, Java)
    - If unknown media elements or code are encountered, the browser determines what to do with the data (MIME).

What is the **H**yper**T**ext **M**arkup **L**anguage (HTML)?

- ASCII (text) based. – (American Standard Code for Information Interchange)
- Subset of Standard Generalized Markup Language (SGML)
- An HTML file is a plain text file
    - Contains a finite set of markup tags
    - Markup tags tell the browser how to display the text
- An HTML file must have either a .html or a .htm file extension
    - Pick one or the other and stick with it – do not use switch back and forth
- An HTML file can be created with any text editor
    - Notepad, XRay, any text editor

    - Save it as PLAIN text – do not save it as rich text
- De facto standards body is the World Wide Web Consortium
    - W3C for short
    - Composed of many companies and individuals
    - Several "working groups" devoted to specific technologies
- Is a markup language, not a formatting language.
- Designed so that many different applications could be written to read HTML, not just browsers we are familiar with.
    - Often called user agents, renders, or simply, browsers.
- HTML markup identifies various parts of the document (text, graphics, and other elements) and what they are. The browser then uses a Document Type Definition (DTD) to determine how elements should be rendered.
- The HTML DTD provides the general rules for the browser to layout the page; it is a loose definition, meaning "Items may shift during transit!"
- The HTML DTD does not excel at coordinate specific layout, dynamic content, interactivity or multimedia.
- The browsers alone can only interpret GIF, JPEG, and PNG images and text.

What is XHTML?

- E**X**tensible **H**yper**T**ext **M**arkup **L**anguage
- Basically HTML with stringent rules for formatting.
    - We will follow many of these rules this semester (even simply HTML does not require it).
    - XHTML extends HTML
    - Is intended to replace HTML
    - Is almost identical to HTML 4.01
- It is XML-based
- It is designed to work with XML-based user agents
    - You can view, edit, and validate them using XML editors such as XRay
- Conforms to both HTML and XHTML at the same time
    - Can be viewed in the browser as an HTML page
    - Can be viewed as an XML page

For this entire course, your pages must not only be HTML compliant, but also conform to XHTML and the wellformedness constraints of XML that are covered in the next lecture.

Toward the end of the semester, students will see what is further required to make their documents valid as well as wellformed xHTML, along with creating XML documents.

Resources:

- http://www.xhtml.org
- http://www.w3.org/TR/xhtml1/
- http://www.w3schools.com/xhtml/