Introduction to the HTML

Origins

US DoD: ARPAnet - late 1960s and early 1970s
- Network reliability
- For ARPA-funded research organizations

City university of New York: BITnet, CSnet - late 1970s & early 1980s
- email and file transfer for other institutions

National Science Foudation: NSFnet - 1986
- Originally for non-DOD funded places(non military uses)
- Initially connected five supercomputer centers
- By 1990, it had replaced ARPAnet for non- military uses
- Soon became the network for all (by the early 1990s)

NSFnet eventually became known as the Internet

Internet

- A world-wide network of computer networks
- At the lowest level, since 1982, all connections use **TCP/IP**
- TCP/IP hides the differences among devices connected to the Internet

Internet Protocol (IP) Addresses

- Every node has a unique numeric address
- Form: 32-bit binary number(4four 8 bit numbers separated by'.')
- New standard, IPv6, has 128 bits (1998)

Domain names

- Form: host-name.domain-names
- Last domain specifies the type of organization
- Fully qualified domain name - the host name and all of the domain names
- DNS servers - convert fully qualified domain names to Ips
- Mid 1980s on top of TCP/IP, telnet, ftp,Usenet,mailto protocols came.
- **Prob: Each protocol has its own user interface and own purpose**

The World-Wide Web

www protocol is the possible solution to the problem mentioned.
Origins:

   - Tim Berners-Lee at CERN proposed the Web in 1989

Purpose: To allow scientists to have access to many databases of scientific
       work through their own computers, exchange documents.

   - Document form: hypertext
   - Pages? Documents? Resources?
   - We'll call them documents
   - Hypermedia – more than just text – images,
      sound, etc.

Web or Internet?

   - Web uses the protocols, http, that runs on the Internet– among several
    others (telnet, mailto, etc.)

Web Browsers

  - **Mosaic** - NCSA (Univ. of Illinois), early 1993

  - First to use a GUI, led to explosion of Web use

  - Initially for X-Windows, under UNIX,

  - ported to other platforms by late 1993

- Browsers are clients - always initiate,
- Servers react (although sometimes servers require responses)

- Most requests are for existing documents, using **Hypertext Transfer Protocol (HTTP)**

- But some requests are for program execution, with the output being returned as a document

Web Servers

- Provide responses to browser requests,
    - existing documents or
    - dynamically built documents

- Browser-server connection is now maintained through more than one
  request-response cycle

- All communications between browsers and servers use HTTP

- Web servers run as background processes in the operating system

- Monitor a communications port on the host, accepting HTTP messages
  when they appear

- All current Web servers came from either
  1. The original from **CERN**
  2. The second one, from **NCSA**

Web server file structure: Directories

Document root (servable documents): Web documents to which the server has direct access

- These files available to the clients through top-level URLs.
- Clients do not access the document root directly in URLs, rather the server maps the requested URL to document root.
- Ex: http://www.abc.com/hello.html---->admin/web/topdocs/hello.html

 Server root (server system software): Descendent directories and it supporting software

-**Virtual document trees**
-**Virtual hosts**
-**Proxy servers**

Web servers: **Apache –fast, reliable, open source, UNIX based**
               IIS   –windows based

URLs

- Used to identify documents (resources) on the Internet.

- General form: **scheme: object-address**
- The scheme is a communications protocol: telnet/ftp/http/file/mailto
- For the http protocol, the object-address is: fully qualified domain name/doc path

- URLs cannot include spaces or any of a collection of other special characters (semicolons, colons, ...)

- A path that includes all directories along the way is: **Complete Path**
- Else the doc path may be abbreviated as a **partial path**

- If the doc path ends with a slash, it means it is a directory
  Ex: http://www.google.com/
- If the doc path ends without slash , it means it is a file
  Ex: http://www.google.com/welcome.html

The Hypertext Transfer Protocol

 The protocol used by ALL Web communications: latest version **http1.1**

HTTP contains two phases: **Request Phase** and **Response Phase**

HTTP communication between web server and browser.
Communication contains: **header ->Info about communication**
body   ->data of communication

Request Phase: The general form of HTTP request is:

     1. HTTP method   domain part of URL   HTTP ver.
     2. Header fields
     3. blank line
     4. Message body

    - An example of the first line of a request:
      GET /cs.uccp.edu/degrees.html HTTP/1.1

Most commonly used methods:

GET - Fetch a document

POST - Execute the document, using the data in body

HEAD - Fetch just the header of the document

PUT - Store a new document on the server

DELETE - Remove a document from the server

Response Phase general form:

    1. Status line

    2. Response header fields

    3. blank line

    4. Response body

- Status line format:

    HTTP version   status code   explanation

- Example: **HTTP/1.1 200 OK**

Status code is a three-digit number; first digit specifies the general status

      1 => Informational

      2 => Success

      3 => Redirection

      4 => Client error

      5 => Server error