Definitions & Basics

- **World Wide Web (WWW) - the hypermedia system used on the Internet to display information containing text, images, audio and video clips, and references to other pages.**

- **browser - a program that accesses and displays information from the WWW.**

- **point and click - clear.**

  A browser interface is point-and-click. Each selected document is displayed on the screen.

- **hypertext - refers to documents containing text and embedded references to other documents.**

Definitions & Basics

- <u>hypermedia</u> - refers to documents containing text, graphics, video and audio, animation, and embedded references to other hypermedia documents.

- <u>link</u> - a pointer or reference to another document.
  - A link is represented by an "active area" on the screen.

    A graphic button.

    Highlighted and/or underlined text.
  - Selecting a link gets the referenced document for display.

Definitions & Basics

- **nondistributed system** - all documents stored locally.

  Links remain valid and consistent.

- **distributed system** - documents stored on a variety of computers.

  Links can become invalid and inconsistent.

- **page** - a WWW document.

  A page can contain different types of information.

- **homepage** - the initial page for an individual or an organization.

Definitions & Basics

- <u>markup language</u> - a text formatting language which specifies formatting with commands called <u>control sequences</u>.

  - A word-processing file using a markup language contains no actual formatting. All formatting is specified with control sequences.

  - Familiar WYSIWYG word-processors like Word do not use a markup language. Such files contain a lot of formatting.

  - TeX is a markup language.

  - People argue the pros and cons of markup vs WYSIWYG. Each has its place.

Definitions & Basics

- **<u>HyperText Markup Language (HTML)</u> - the standard markup language used to create documents on the WWW.**

  - **HTML includes formatting commands along with the text to be displayed.**

  - **See book for a basic description.**

- **<u>Uniform Resource Locator (URL)</u> - a special syntactic form to identify a page on the WWW.**

  - **The general form of a URL is**

    ```
    protocol://computer_name:port/
            document_name
    ```

  - `protocol` **can be** `http, ftp, file, mailto.`

Definitions & Basics

- `computer_name` **is the computer's DNS name.**

- `port` **specifies an optional TCP port. The default port is 80.**

- `document_name` **is the path on the host computer to the page.**

- **HyperText Transport Protocol (HTTP)** **- the protocol used to transport a WWW document from one computer to another.**

  - **The currently accepted version is HTTP 1.0 (RFC 1945).**

  - **HTTP 1.1 (RFC 2616) is a backward compatible upgrade that appeared in 1998.**

Definitions & Basics

- **HTTP specifies:**

  how a client and a server establish a connection;

  how the client requests data from the server;

  how the server responds to the client;

  how the connection is closed.

- **HTTP connections use TCP/IP for data transfer.**

Client-Server Interaction

- **The connection between a Web browser and a server is short.**

- **Once a connection is made and a file transferred, the connection is closed.**

- **HTTP 1.0 is a <u>stateless</u> protocol using <u>nonpersistent</u> connections.**

  - **HTTP 1.0 opens and closes a connection for each request and the server does not store any state information about the client.**

  - **The main improvements in HTTP 1.1 are <u>state</u> and <u>persistent</u> connections.**

Browser Architecture

- **Web browsers are more complex than Web servers.**

- **All a server has to do is serve up Web documents repeatedly on request.**

- **A browser consists of much more.**

  - **An HTTP client to request documents.**

  - **Other clients for other protocols such as FTP.**

  - **A driver to display output on the screen.**

Browser Architecture

- Interpreters to interpret the documents.

  An HTML interpreter.

  Other optional interpreters such as Shockwave.

- A controller to manage everything.

  Accept input from a mouse and keyboard.

  Call on other components to perform requests specified by the user.

Caching in Browsers

- Downloading HTML documents can be slow.

    - Internet congestion.

    - Slow dialup connection.

    - Busy server.

- Returning to previous HTML documents requires reload from the server.

- Instead, a local cache can be used to hold copies of recently visited pages.

Proxy

- <u>proxy</u> - a network entity that satisfies HTTP requests on behalf of a client.

- An organization can set up an HTTP proxy that caches documents for multiple users.

- AOL runs one of the largest proxy-server farms in the world to speed data transfer to users.

Identifying a User

- An HTTP server is stateless and does not keep track of users.

- HTTP provides two mechanisms to help a server identify a user.

  - Authentication.

  - Cookies.

Authentication

- Used to restrict user access.

- Requires a user to log in with a user name and a password.

Cookies

- **Used to serve content as a function of user identity.**

How Cookies Work

- **An HTTP client contacts a Website using cookies.**

- **The HTTP server's response includes a** `Set-cookie:` **header with an ID number.**

  Example: `Set-cookie: 1234567`

- **When the HTTP client receives the response, it appends a line to a special cookie file stored on the client machine. The line includes the host name of the server and the associated ID number.**

How Cookies Work

- In subsequent requests to the same server, the client request includes a `Cookie:` request header with the corresponding ID number.
Example: `Cookie: 1234567`
- The server does not know the user's name, but the server does know this is the same user who contacted the server earlier.

How Cookies Are Used

- If a server requires authentication but doesn't want to hassle a user with a user name and password prompt each time the user visits a Website, it can set a cookie.
- If a server wants to remember a user's preferences so that it can provide targeted advertising during subsequent visits, it can set a cookie.
- If a user is shopping at a site, the server can use cookies to keep track of the items a user is purchasing—that is, to create a virtual shopping cart.

Cookies

- Cookies don't work well for a nomadic user who accesses the same site from different machines.
- The use of cookies is controversial. For example, see
  `www.cookiecentral.com`
- If you don't want your browser to accept cookies, you can disable them. In Netscape, go to Preferences->Advanced.
- Sample cookie file and additional information.