World Wide Web Pages and Browsing

Chapter 35

Browser Interface

- The World Wide Web (WWW) is a large-scale, online repository of information accessed with a browser.
- WWW is a hypermedia system, an extension of a hypertext system. Documents contain active links to other documents.
- A WWW document is called a page and is represented using HyperText Markup Language (HTML).
- HTML does not contain detailed formatting. The browser must format and display the page.

HTML Format and Representation

- Each HTML document has a head and a body.
- It is a text file that contains tags (most come in pairs).
  `<head>` junk junk `</head>`
- White space is ignored.
- Tags provide structure and format hints.

General Form of HTML Document

```
<html>
  <head>
    <title>text of the document title</title>
  </head>
  <body>
    text of the body of the document
  </body>
</html>
```

Here is the same HTML text, just much harder to read. It will be displayed exactly the same by the browser.

HTML Formatting Tags

- `<br>` causes a line break.
- `<h1>` indicates a heading of level one (the largest heading).
- `<h2> ... <h6>` other heading levels.
- `<img src="fred_photo.gif"`> an image to displayed on the page and where to get it.
- `<img src="fred_photo.gif" align=middle>` the same image now displayed in the middle of the page.
**HTML Formatting Tags**

<table>
<thead>
<tr>
<th>An unordered list (bullets) of four names.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Here is a list of four names.</td>
</tr>
<tr>
<td><code>&lt;ul&gt;</code></td>
</tr>
<tr>
<td><code>&lt;li&gt;</code> Scott <code>&lt;/li&gt;</code></td>
</tr>
<tr>
<td><code>&lt;li&gt;</code> Sharon <code>&lt;/li&gt;</code></td>
</tr>
<tr>
<td><code>&lt;li&gt;</code> Jan <code>&lt;/li&gt;</code></td>
</tr>
<tr>
<td><code>&lt;li&gt;</code> Stacey <code>&lt;/li&gt;</code></td>
</tr>
<tr>
<td><code>&lt;/ul&gt;</code></td>
</tr>
</tbody>
</table>

This is how the list would look:

- Scott
- Sharon
- Jan
- Stacey

---

**Identifying a WWW Page**

- Uniform Resource Locator (URL) general form:
  - protocol://computer_name:port/document_name
- Example URL:
  - [http://www.netbook.cs.purdue.edu/toc/toc01.htm](http://www.netbook.cs.purdue.edu/toc/toc01.htm)
- A hypertext link to another document is activated when the text item is selected.

This book is published by [Prentice Hall](http://www.prenhall.com) one of the larger publishers of textbooks

---

**Client-Server Interaction**

- The connection between the browser (client) and the web server lasts only for one page.
  - Browser opens the connection
  - Browser requests the page.
  - Server sends the page.
  - The connection is closed.
- To display an image, the above procedure is repeated (for each image).
- The browser interacts with the server using HyperText Transfer Protocol (HTTP)

---

**Web Document Transfer and HTTP**

- Browser requests a page, server supplies it.
- HTTP supports four basic operations that a browser can specify when making a request:
  - GET requests a specified item from the server. The server returns a heading that contains status information followed by a blank line followed by the requested item.
  - HEAD requests status information about an item. The server returns status information without returning a copy of the item itself.

---

**Example HTTP Header Returned by Server**

```
HTTP/1.0 200 OK
Date: Mon, 30 Oct 2000 01:22:22 GMT
Server: Apache/1.2.5
Last-Modified: Sat, 28 Oct 2000 01:03:37 GMT
ETag: "130fe-81-3883bbe9"
Content-Length: 129
Accept-Ranges: bytes
Connection: close
Content-Type: text/plain
```
Optional Clients in a Browser

A browser can contain other clients besides an HTTP client. Typical ones are email and FTP.

Many of the examples in this text are available online. Source code from 
<ftp://ftp.cs.purdue.edu/pub/comer/netbook/client.c> an example client program is available.

Caching in Web Browsers

- Browsers use a cache to improve document access.
- Large images can fill the cache.
- Cache is useless for checking on the bids for an item on ebay.com.
- HTTP allows the server to specify a timeout for a page or not allow caching at all.
- Browser can flag a request not to be supplied from a cache.
- Browser can validate a cached page that has timed out with the HEAD operation.