Week 8:
HyperText Transfer Protocol - Clients - HTML

Johan Bollen
Old Dominion University
Department of Computer Science

jbollen@cs.odu.edu
http://www.cs.odu.edu/~jbollen
MIDTERM EXAM

See: http://www.cs.odu.edu/~jbollen/CS312
HTTP and the Web: some peculiarities

1. What’s index.html?
   (a) Absent filename
   (b) Server by default returns index.html
   (c) this allows:
   
   [http://www.cs.odu.edu/]

2. Why can I sometimes not complete URL or malform them?
   (a) Client fills in blanks
   (b) be careful though

3. What are those long URLs with lots of ampersands
   (a) Note question mark
   (b) Information is passed back to server
   (c) packed in URL
HTTP: some notes

1. Free for all
   (a) No password protection or authentication
      i. Unless specifically implemented by server
   ii. Reserved to HTTP/1.0 and up
   (b) Enables access for all that speak HTTP
   (c) Simple, efficient scheme to distribute data
   (d) Possibility of two-way traffic (not duplex!)

   i. URL encoding: http://www.cs.odu.edu/~jbollen/cgi-bin/test.cgi?dec=5
   ii. POST
   iii. PUT

2. Possibility of layering
   (a) Protocols on top of protocols (OAI)
   (b) Use HTTP to transfer commands associated with meta-protocol
   (c) Data transferred in HTML or any other format
Web clients

1. Functions as gateway between server and user

2. Interface:
   (a) Display documents retrieved from server
   (b) Select hyperlinks (URL)
   (c) Parses HTML documents
      i. Layout and structure
      ii. Displays images
      iii. Deals with other multimedia

3. Must adhere to standards
   (a) HTML
   (b) HTTP
   (c) MIME
   (d) CSS

4. Interface
   (a) GUI
   (b) text-only
Lynx


2. Available for variety of operating systems
   (a) Unix
   (b) Windows NT, W95
   (c) Mac OS X?

3. Pros:
   (a) Fast

(b) low overhead
(c) generally available
(d) Adheres to standards

4. Cons:
   (a) Poor web design
   (b) Generally not installed on Windows systems
   (c) Lack of graphical and multimedia content
Ten people were killed and 34 injured today when a Staten Island ferry slammed into a pier, the Fire Department of New York said. Police and local government officials said the ferry's master went home and attempted suicide after the accident. He was listed in critical condition.

FULL STORY
- Scenes from the accident: premium content Video | Gallery
- Witnesses: Passengers jumped for their lives
- Map: Site of accident | Interactive: Andrew L. Barberi data

MORE TOP STORIES Most Popular
- Judge OKs testimony on Bryant accuser's sex history | premium content Video
- Chinese spacecraft returns, ending successful mission | premium content Video
- Suspects rounded up in U.S. bombing in Gaza | premium content Video
- U.S. troops battle infiltrators at Iraqi-Syrian border | premium content Video

Arrow keys: Up and Down to move. Right to follow a link; left to go back.
Graphical Browsers: Netscape

1. Netscape History
   (a) Originally introduced in 1994
   (b) Grew out of Mosaic
   (c) Major browser until ± 1997
   (d) Overtaken by MS Internet Explorer
   (e) 1998: Move to Open Source
   (f) Slow death afterward
   (g) [http://www.blooberry.com/](http://www.blooberry.com/)

2. Now:
   (a) Mozilla 1.0.1: highly stable, efficient browser
   (b) Adheres to all standards
   (c) Special features
      i. Tabbed browsing
      ii. Open Source
      iii. Available for range of OS

[http://www.blooberry.com/indexdot/history/netscape.htm](http://www.blooberry.com/indexdot/history/netscape.htm)
Some common features

1. Bookmarks
   (a) User collects basket of interesting links
   (b) Can be organized according to theme, etc

2. History
   (a) Contains list of recently visited sites
   (b) May be important for multiple user systems!

3. URL completion
   (a) Client automatically completes links
   (b) Security issues!

4. Helper applications
   (a) Associate MIME types with specific applications
   (b) Informs client how to display certain files

5. View Page Source
   (a) Useful to manually parse web pages
   (b) Used more often than you’d expect

6. Similar pages service
   (a) Added service to list “similar” pages
   (b) Often external service which adds links
   (c) Alexa
Hypertext Markup Language (HTML)

1. HTML: markup language for web pages
   (a) Specification of what web pages must look like and what they must contain
   (b) all plain text = content + description
   (c) Content: regular text
   (d) description: HTML code or markup
      i. Commands: insert image, begin page body, etc.
      ii. Tags: specify formatting of content

2. All other files untouched by HTML:
   (a) Images
   (b) Sound
   (c) Video
   (d) Plain text

3. Browser:
   (a) retrieves HTML page
   (b) parse HTML code
   (c) Displays content according to specification

4. HTML editor
   (a) Any text editor
   (b) Make sure to save in text format
HTML: two conflicting principles

1. Markup describes what something is:
   (a) titles, headers, tables
   (b) browser decides how it looks

2. Markup describes what something must look like:
   (a) align right, make bold, italics
   (b) use different fonts
   (c) etc

3. HTML
   (a) Middle ground between semantics and cosmetics
   (b) browser has some freedom of interpretation
   (c) layout and form will depend on type of browser
   (d) evolution of HTML is toward increasing designer control
**HTML tags**

- **a tag is a predefined combination of characters to inform the browser of how to interpret and display page**

1. Tags: enclosed between `<` and `>`
2. For example: `<html>`, `<title>`
3. Tags are embedded in text of HTML page
4. **plain text**
   - (a) Case insensitive
   - (b) Can be written in any editor
   - (c) Does not require any specific coding other than ASCII
Tags

<TAG ATTR1='v1' ATTR2='v2'> text to be formatted </TAG>
HTML Tags

1. Tags
   (a) Enclose text
   (b) Begin and end tag must be the same (note slash)
   (c) Can be nested
      i. respect begin and end tag

2. Attributes
   (a) Modify behavior of tag
   (b) Can be more than one
   (c) quotes please!

   ii. take note of how it will affect marked up item
Structure of HTML pages

1. Document
   (a) ASCII file
   (b) Suffix: html
   (c) example: mypage.html
   (d) Remember index.html is special
   (e) Permissions: read other

2. Structure

(a) Head
   i. title
   ii. other “administrative” items

(b) Body
   i. Actual text and content
   ii. This is what will be displayed in main window
Example structure of HTML page

<html>
<head>
<title>Internet Concepts CS312</title>
</head>
<body>
This page concerns the fall 2003 version of Internet Concepts - CS312
</body>
</html>
Some formatting notes

1. Title: make it descriptive and distinctive

2. Given tags always start and end
   (a) spacing and indentation not required
   (b) HTML code may be one long

3. However!
   (a) Bad practice: your code will be unreadable
   (b) You need to be able to match start and end tags!
BLECH!

<html><head><title>Welcome to MSN.com</title><base href="http://g.msn.com/0US!s8.559_6217/" />
<meta http-equiv="pics-label" content="(pics -1.1 &quot;http://www.icra.org/ratingsv02.html&quot; comment &quot;Single file v2.0&quot; l gen true for &quot;http://www.msn.com&quot; r (nz 1 vz 1 lz 1 oz 1 cz 1)) &quot;http://www.rsac.org/ratingsv01.html&quot; l gen true for &quot;http://www.msn.com&quot; r (n 0 s 0 v 0 l 0))" />
Homework 3

Produce a simple homepage for yourself using tags discussed in this class only

**Requirements**

- Descriptive title
- List your full name
- Fake coordinates
- Fake phone number
- Fake hobby
- Your school or employer
- Try to use ASCII formatting

E-mail HTML code to me (no attachments!!), include your name and ID

Email title: “homework 3 - CS312 - Your name”

**Deadline**

Start of next class

**Scoring**

Grading on technical soundness and style, bad HTML is max half score