

HTML

Lecture 5

cs193i – Internet Technologies
Summer 2004
Stanford University

Administrative Stuff

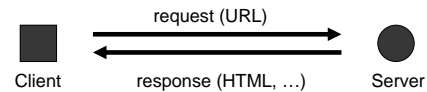
- HW #1 due July 12
- Silas' review Perl review session
 - 7/13, 2:15-3:05
 - Skilling 193
- Lab #2 due July 14
- Midterm Monday, July 19, in class

Today's Discussion

- HTML
- Details on HTTP Request/Response Headers
- Advanced Topics

Review: Client and Server

- User uses HTTP client (Web Browser)
- It has a URL (e.g. `http://www.yahoo.com/`)
- Makes a request to the server
- Server sends back data (the response)
- User clicks on the client side...



HTML Characteristics

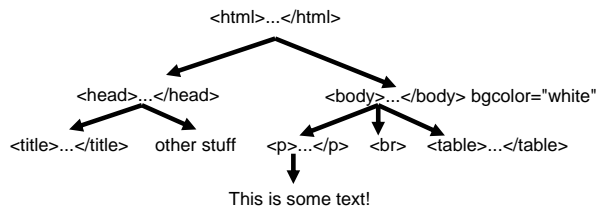
- Just a Text File!
 - + Portable
 - + Human Readable/Writable
- Defines the Structure (not Appearance) of the Document
 - Client (Browser) defines the appearance
 - Font preferences, window width, ...
 - Pours into Browser (PDAs, Bigger/Smaller)

Document Structure

```
<html>
<head><title>My First Web Page</title>
</head>
<body bgcolor="white">
<p>A Paragraph of Text.</p>
</body>
</html>
```

Nested Tags

- Like a tree, each element is contained inside a parent element
- Each element may have any number of attributes



Basic Tags

```
<html>
<head><title>My First Web Page</title>
</head>
<body bgcolor="white">
<p>A Paragraph of Text.</p>
</body>
</html>
```

Basic Tags

- `<!doctype html public "-//w3c//dtd html 4.0 transitional//en">`
 - Preamble which identifies content as HTML
- `<h1>...</h1>`
 - H1-6 where larger number means smaller heading
- `<p>`
 - Includes vertical whitespace unlike `
`

Basic Tags

- `<hr>` horizontal rule
- `
` new line
- `...` bold
- `<i>...</i>` italicize text in between

Lists

- Unordered Lists
 - ``
 - ` Apples`
 - ` Oranges`
 - ``
 - Ordered Lists
 - ``
 - ` One`
 - ` Two`
 - ``
 - Can be nested
- Apples
 1. Fuji
 2. Granny Smith
 - Oranges

Image Files

- ``
- JPEG
 - Best for photos
 - Public standard
- GIF
 - Best for simple images
 - Older standard
- PNG – Portable Network Graphics
 - Public standard replacement for GIF
- SVG – Scalable Vector Graphics
 - Series of drawing commands
 - Uses XML.

Tables

- `<table>...</table>`
- `<tr>...</tr>` for each row
- `<td>...</td>` for each element in a row
- `<th>...</th>` for header row

Table Example

```
<table border="1">
<tr>
<td>row 1, cell 1</td>
<td>row 1, cell 2</td>
</tr>
<tr>
<td>row 2, cell 1</td>
<td>row 2, cell 2</td>
</tr>
</table>
```

row 1, cell 1	row 1, cell 2
row 2, cell 1	row 2, cell 2

Comments

- `<!-- This is a comment -->`
- `<!--`
This paragraph,
is also a
comment...
`-->`

Special HTML

- `<` → `<`
- `>` → `>`
- `&` → `&`
- ` ` → space

Anchor Tag (Links)

Absolute HREFs specify fully qualified URLs.

- `Yahoo!`
- `In this directory!`
- `In sub-directory a!`

Relative HREFs are relative to the directory containing the current HTML file.

What is the WWW?

- A hypertext system that runs on top of the Internet, based on Three Main Standards
 - URL
 - HTTP
 - HTML

Client/Server Timeline



HTTP Request Structure

Method URI HTTP-version
General Header
Request Header
Entity Header

URL, URN, or URI?

- URN is location-independent resource identifier
 - urn:ietf:rfc:3187
 - urn:isbn:0451450523
- URL is the location
- URI is the superset of URL & URN

URL Structure

<scheme>://<user>:<password>@<host>:<port>
/<path>;<params>?<query>#<frag>

Unsafe Characters

- Some Characters need to be encoded
 - ~ [ASCII: 126 (0x7E)]
 - SPACE [ASCII: 32 (0x20)]
 - % [ASCII: 37 (0x25)]
- Examples
 - `http://www.bob.com/%7Ekelly/`
 - `http://www.bob.com/my%20home%20page.html`
 - `http://www.bob.com/100%25Crankiness.html`

Empty-String Path

- `http://www.yahoo.com`
- Assume the path is `"/`
- Client should send
`GET / HTTP/1.0\r\n\r\n`

Relative Headers

- Client Side
Given a URL in a file, if it is relative, will add base address to the relative URL
- Last requested path is `http://foo.com/b/index`.
 - in `index.html` see `link</`
 - base address is `http://foo.com/b/`
- Client requests `http://foo.com/b/a.html`

Request Header

- GET / HTTP / 1.1
<HEADERS>

Request Header Example

```
GET / HTTP/1.1
Host: localhost:8181
Connection: keep-alive
Referer: http://localhost/~ronyeh/
User-Agent: Mozilla/5.0 (Macintosh; U; PPC Mac OS
X; en-us) AppleWebKit/124 (KHTML, like Gecko) Safari/125
Accept: */*
Accept-Encoding: gzip, deflate;q=1.0, identity; q=0.5, *;q=0
Accept-Language: en-us, ja;q=0.62, de-de;q=0.93
```

General Headers

- Used by clients & servers
Seen in both requests and responses
- Date: Tue, 3 Oct 1974 02:16:00 GMT
- Connection: keep-alive

Request Headers

- Client-IP: 123.23.34.52
- Host: hostmachine.com
- Referer: `http://wherefrom.com/`
- User-Agent: Mozilla/5.0
- UA-OS
- If-Modified-Since

Request Headers

- Accept: */*
- Accept: text/html
- Accept-Language: en-us, ja
- Accept-Encoding: gzip

HTTP Response Structure

HTTP-version	Status-code	Reason-phrase
General Header		
Response Header		
Entity Header		
Entity Body		

Example Response

HTTP/1.0 200 OK
Content-type: text/plain Content-length: 19
Hi! I'm a message!

HTTP Server Response Codes

- 200 OK
- 3XX -- Minor Client Error
 - 301 -- File Moved Permanently
 - 302 -- Moved Temporarily
 - 304 -- Not Modified
- 4XX -- Major Client Error
 - 400 -- Syntax Error
 - 401 -- Unauthorized
 - 403 -- Forbidden, Permission Denied
 - 404 -- Not Found!

HTTP Server Response Codes

- 5XX -- Server Errors
 - 500 -- Internal Server Error
 - 503 -- Service Unavailable

General Headers

- Used by clients & servers
 - Seen in both requests and responses
- Date: Tue, 3 Oct 1974 02:16:00 GMT
- Connection: close

Server Response Headers

- Server: GWS/2.1
- Content-Length: 2136
- Content-Type: text/html
- Location
- Expires
- Last-Modified

MIME

- Multipurpose Internet Mail Extensions
- type, subtype, & optional parameters
- type/subtype; param1=value1

MIME types

- application/*
- audio/*
- image/*
 - image/jpeg
 - image/tiff
- text/*
 - text/xml
 - text/rtf
 - text/html
 - text/plain
- video/*
 - video/quicktime
 - video/mpeg
 - video/x-msvideo

Pages with Multiple Types

- Each entity (ex. image) is standalone HTTP request
 - Page with many pictures creates many connections
- Each response therefore has appropriate MIME settings

Mapping URL Path

- Server can map URLs to any place on the file system. Doesn't have to be under the Document Root. It's the server's choice!!!
- User names
 - ~kashaw
 - May map to /users/kashaw/WWW
- /a/b/ => maps to a default file
 - index.html, default.html, index.htm, index.shtml
- /a/b/ => if default file doesn't exist, may list the directory's files

Trailing Slash

- What if Client asks for /a/b
- Say file b doesn't exist
- Utilize the 301 Redirect to /a/b/
- Client re-does request

What happens if server does NOT issue 301,
but gives the client the right file anyways?

Advanced Topics

- Redirection
- Caching
- Performance
- HTTP 1.1

When NOT to Redirect

- Client requests /a/b/
 - Server maps to /a/b/index.html and sends back html file
 - `A Link`
- Client takes base address /a/b/ concatenates with c.html
- Client requests /a/b/c.html which is correct!

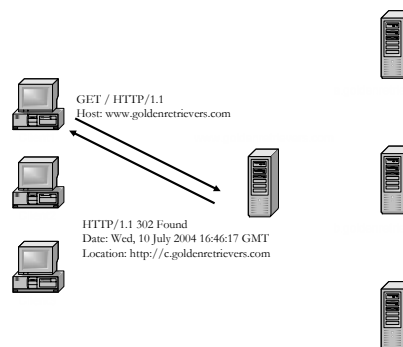
When to Redirect

- URL missing trailing slash
 - No file named /class/cs193i
 - But, directory named /class/cs193i/
- If Redirect did NOT happen
 - Client thinks base address is /class/
 - a relative href="schedule.html" in cs193i will be mapped by client to /class/schedule.html
 - Server will return 404 Not Found

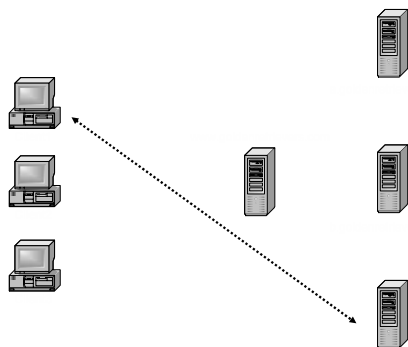
Why Redirect?

- Reliability
(Find Live Hosts)
- Minimize Delay
(Find Shortest Path)
- Conserve Network Bandwidth
(Spread out Requests Geographically)
- Load Balancing
(Distribute Requests Temporally)

Load Balancing Example



Load Balancing Example



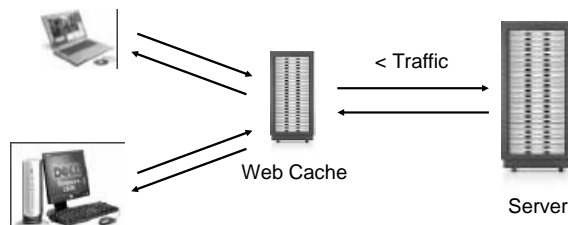
Redirection Tradeoffs

- HTTP Redirection
 - Every request initially goes through the www.goldenretrievers.com machine
 - Must Customize www Web Server
- Alternative: DNS Redirection
 - DNS server decides which IP address to return (from a list of OK IP Addr)
- Alternative: Hardware Redirection
 - NAT Box! Packet rewriting!

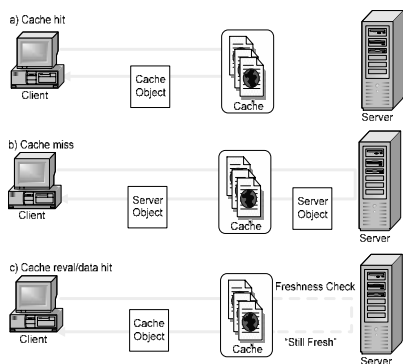
Caching Motivation

- Redundant Data Transfer
- Network Bandwidth Bottlenecks
- Server Demand
- Distance Delays (Latency)

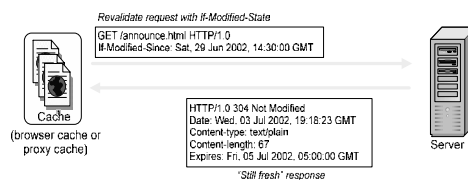
Adding Caching



Hit, Miss, Revalidate



Revalidate Options

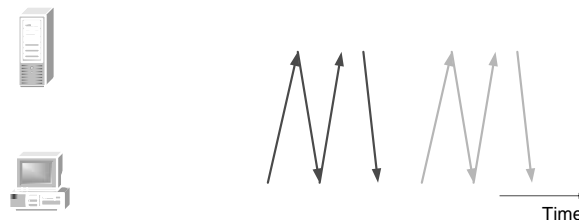


- Check via If-Modified-Since...Not Modified
 - Suffers from 2X latency between cache & server
- Just assume, and have a timeout, refresh cache automatically

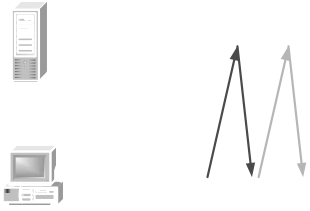
Request / Response Timeline



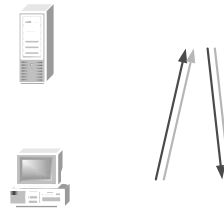
Web Pages w/ Multiple Requests



Persistent Connections



Pipelining



Connection: Keep-Alive

Persistent Connections

- HTTP 1.0 -- Connections close by Default
 - No need for Content-length, end signaled by EOF (in-band signal)
- HTTP 1.1 -- Persistent by Default
 - Must use Content-length

Chunked-Transfer Encoding

- Problem: Content-length costly for server
- Solution
 - Server omits Content-Length
 - Transfer-encoding: chunked
 - Send Data in Chunks, Prefixed by length in Hex
- End is marked with Chunk Length 0 (in band signal like in POP)