

Hypertext Transfer Protocol (HTTP)

Raj Jain

Washington University in Saint Louis

Saint Louis, MO 63130

Jain@cse.wustl.edu

These slides are available on-line at:

<http://www.cse.wustl.edu/~jain/cse473-05/>



- q Hypertext Transfer Protocol (HTTP)
- q Hypertext Markup Language (HTML)
- q Key HTTP Terms
- q URI vs URL vs URN
- q Intermediate HTTP Systems
- q HTTP Message Structure
- q HTTP 1.1 Features

Hypertext Transfer Protocol (HTTP)

- q Hypertext: Documents contain pointers to other documents
- q HTTP: Protocol used between web browsers and web servers
- q Originally designed for hypertext. Not limited to text. Used for all types of media.
- q Transaction oriented client/server protocol
- q Uses TCP connections
- q Stateless: New TCP connection for each transaction (Changed in HTTP v1.1)
- q Presentation separate from content
 - q Presentation controlled by the browser
 - q Content provided by the server
- q Hypertext markup language (HTML)
- q Hypertext transfer protocol (HTTP)

Hypertext Markup Language (HTML)

- q Header, Body, Anchors, Hyper References
- q **Sample Code:**

```
<HTML>
<HEAD>
<TITLE>Hello</TITLE>
</HEAD>
<BODY>
<A
  HREF="http://www.cse.wustl.edu/~jain/greetings.html">Ho
  w are you?</A>
</BODY>
</HTML>
```

Output

Hello

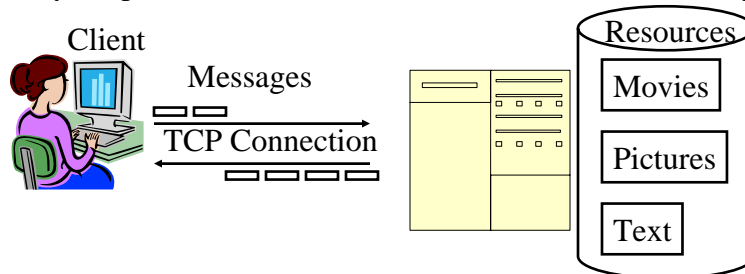
[How are you?](#)

A Sample List of HTML Tags

<code><A></code>	<code></code>	Anchor (link or name)
<code><BODY></code>	<code></BODY></code>	Contents
<code>
</code>		Break
<code><FORM></code>	<code></FORM></code>	Input form
<code><H1></code>	<code></H1></code>	Heading level 1
<code><HEAD></code>	<code></HEAD></code>	Header of a document
<code><HR></code>		Horizontal Rule
<code><HTML></code>	<code></HTML></code>	The doc type is HTML
<code></code>		List Item
<code></code>	<code></code>	Ordered List
		Paragraph break
<code><PRE></code>	<code></PRE></code>	Preformatted text
<code><TITLE></code>	<code></TITLE></code>	Document title
<code></code>		Unnumbered list

Key HTTP Terms

- q User agent: Browsers, spiders
- q Client and Server systems
- q Connection: TCP
- q Message
- q Resource: Object or service. Identified by URI/URN/URL
- q Entity: Representation of a resource with a header and body



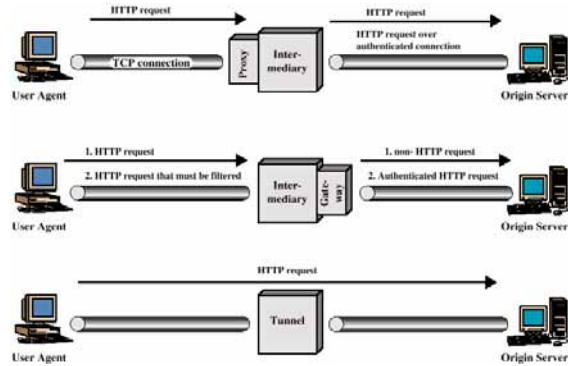
URI vs URL vs URN



- q Uniform Resource Name (URN): Host/path/object#name
www.cse.wustl.edu/~jain/refs/wir_refs.htm#consortia
- q Uniform Resource Identifier (URI):
Name + Scheme (How to get it)
Scheme://host/path/object#name
http://www.cse.wustl.edu/~jain/refs/wir_refs.htm#consortia
- q Uniform Resource Locator (URL): URI based on location rather than any other attributes
- q URI vs URL
 - q IDs need not be location based \Rightarrow URLs are subset of URIs
 - q Most URIs are based on location \Rightarrow URLs \cong URIs
In practice, URL and URI are used interchangeably.
 - q URI is preferred in technical documentation.

Intermediate HTTP Systems

- q Proxy: For performance enhancements. **Caches** pages.
- q Gateway: For security. Used at firewall boundary.
- q Tunnel: simple relay
- q Origin server: Holds original content



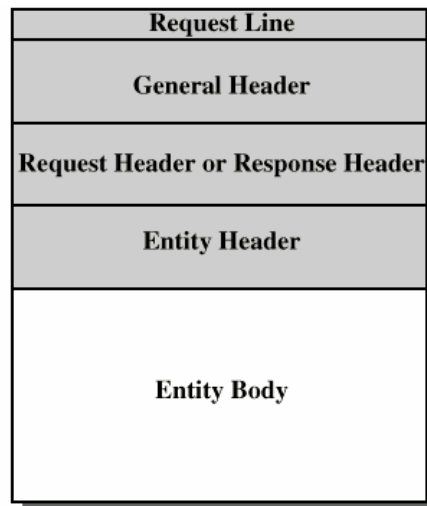
Washington University in St. Louis

CSE473s

©2005 Raj Jain

20-9

HTTP Message Structure



Washington University in St. Louis

CSE473s

©2005 Raj Jain

20-10

Sample HTTP Exchange

```
telnet www.cse.wustl.edu http
Connected to hydra.cs.wustl.edu.
Escape character is '^]'.
GET /~jain/index.html HTTP/1.0      } Request Line
Accept: text/plain, text/html      } Request Header

HTTP/1.0 200 OK                      } Response Line
Server: Netscape-Enterprise/2.0a    }
Date: Tue, 25 Feb 2005 05:04:11 GMT }
Accept-ranges: bytes                }
Last-modified: Tue, 25 Feb 2005 05:03:07 GMT }
Content-length: 84                  }
Content-type: text/html              }
```

Sample HTTP Exchange (cont)

```
<HTML>                                }
<HEAD>                                } Entity Header
<TITLE>Hello</TITLE>
</HEAD>
<BODY>                                }
Hello! How are you?                   } Entity Body
</BODY>
</HTML>

Connection closed by foreign host.
```

HTTP Requests

- q **GET** Return the contents
- q **HEAD** Return the header
- q **POST** Treat the document as a script and send some data to it
- q **PUT** Replace the contents with some data
- q **DELETE** Delete the indicated document
etc.

HTTP Request Headers

Header	Description
From	Email address of user
User-Agent	Client software
Accept File	File types that client will accept
Accept-encoding	Compression methods
Accept-Language	Languages
Referrer	URL of the last document the client displayed
If-Modified-Since	Return document only if modified since specified
Content-length	Length (in bytes) of data to follow

HTTP Status Codes

Code	Text
2xx	Success
3xx	Redirection
301	Moved
302	Found
4xx	Client Errors
400	Bad Request
401	Unauthorized
404	Not found
5xx	Server Errors
500	Internal Error
502	Service overloaded

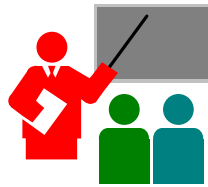
HTTP Response Headers

Header	Description
Server	Server software
Date	Current Date
Last-Modified	Modification date of document
Expires	Date at which document expires
Location	The location of the document in redirection responses
Pragma	A hint, e.g., no cache
MIME-version	
Link	URL of document's parent
Content-Length	Length in bytes
Allowed	Requests that user can issue, e.g., GET

HTTP 1.1 Features

- q Persistent TCP Connections: Remain open for multiple requests
- q Partial Document Transfers: Clients can specify start and stop positions
- q Conditional Fetch: Several additional conditions
- q Better content negotiation
- q More flexible authentication

Summary



- q Content separate from presentation
- q Protocol messages in plain text
- q HTML to define the media
- q HTTP to transfer the media
- q HTTP 1.1 allows for persistent connections

Reading Assignment

- q Read Section 22.2 of Stallings' 7th edition