THE WORLD WIDE WEB
Outline

- A recipe for success
- The hyper-text transfer protocol
- Advanced features
- Performance improvement
A RECIPE FOR SUCCESS
A Web of Documents

Web pages contain references to others
A World Wide Web (WWW) “Anywhere” in the Internet
The Ingredients

- Servers
- Clients (browsers)
- Document format (HTML)
- Identifiers (URIs)
- Protocol (HTTP)
Web Pages

Different objects combined in a framework

Multimedia
Hyper-Text Markup Language

Some objects are “sensitive”
Universal Resource Identifier (URI)

- Identifies each object (resource)
- Where to find it
  - URL: Universal Resource Locator
- How to get it
  - Protocol
Universal Resource Identifier (URI)

http://www.polito.it/teach/cn/main.htm#bib

Schema and URI format

Host (server)

Path

Object (document)

Anchor

Port

http://12.18.43.2:8080
List of Publications of Mario Baldi

Books

These publications provided here are for personal use only as they have been published at copyrighted journals or conference proceedings. In particular works published by IEEE should be used in accordance to the following IEEE note:

IEEE © 2001 Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Personal use of this material is permitted.
Books

Web Browsers

→ “Visualize” pages
→ Download a new one upon a click
→ Support several protocols
  → HTTP, FTP, SIP
→ Object types
  → Images, video, sound
Success Factors

→ Intuitive
→ “Colorful”
→ Multimedia
HYPER-TEXT TRANSFER PROTOCOL (HTTP)
Features

- Text based
- Client-server
- Based on TCP
- Opened by client
- Usually port 80 (server)
Request (C)-Response (S) Protocol

TCP Open

HTTP Request

HTTP Response

Stateless
Message Format

Request/Status Line

Header Field [<name>:<value>]

... 

Header Field [<name>:<value>]

[CR LF]

Body
Request Line

<method> <URL> <ver>

- Methods
  - GET, POST, PUT, HEAD
- Version
  - HTTP/1.0
  - HTTP/1.1
Sample Request Message

GET /baldi/pubs/index.htm HTTP/1.1
Host: staff.polito.it
User-Agent: Mozilla/5.0
Accept: text/html,application/xhtml
Accept-Language: en-us
Accept-Encoding: gzip, deflate
Connection: keep-alive
Status Line

\(<\text{ver}>\) \(<\text{status code}>\)

- 200  OK
- 301  Moved Permanently
- 400  Bad Request
- 404  Not Found
- 500  Internal Server Error
Sample Response Message

HTTP/1.1 200 OK
Date: Sat, 15 Jun 2013 21:17:27 GMT
Server: Apache
Accept-Ranges: bytes
Content-Length: 93589
Keep-Alive: timeout=15, max=97
Connection: Keep-Alive
Content-Type: text/html

<html><head>
Status Code Categories

- 2xx Success
- 3xx Redirection
- 4xx Client problem
- 5xx Server problem
Authentication

To view this page, you must log in to this area on agateway.adp.com:443:
Portal Access [17:44:30:9112]
Your login information will be sent securely.

Name: 
Password: 

[ ] Remember this password in my keychain

Cancel  Log In

- First Time Users Register Here
- Need Help Getting Started?
- Update My Security Profile
- Change your Password
- Make This Site Your Home Page
- Forgot your User Id
- Forgot your Password

©2013 Automatic Data Processing, Inc.
Authentication

GET ...

401 Unauthorized
WWW-Authenticate: <chlg>

GET ...
Authorization: <cred>

200 OK

GET ...
Authorization: <cred>
Cookie

- Mechanism for the server to identify a returning client
- Enables stateful interaction
  - E.g. shopping chart
  - User preferences
PERFORMANCE IMPROVEMENT
Caching

- Clients store pages and objects
- What if they change?
  - HEAD method
  - If-modified-since: <date> header
  - 304 Not Modified
Proxy Server

Internet

www.netscire.it

GET www.netscire.it

index.htm

GET www.netscire.it

Corporate network

index.htm
Proxy

→ Reduce traffic
→ URL filtering
→ Centralized monitoring
→ Client protection
Content Delivery Network (CDN)