Introduction to the World Wide Web

Sara Bouchenak

Sara.Bouchenak@imag.fr
http://membres-liglab.imag.fr/bouchenak/teaching/

Agenda

Lecture, Tuesday, 09:45 – 12:45	Lab, Tuesday, 09:45 – 12:45
Introduction to distributed systems	
	Distributed applications with RMI (Part I)
Distributed Web applications	
	Distributed applications with RMI (Part II)
Interruption week	
Event-based systems & MapReduce systems	
	Distributed Web applications with Servlets (Part I)
Cloud computing	
	Distributed Web applications with Servlets (Part II)
Advanced techniques for efficient distributed systems	
	Caching with Memcached
Event-based systems & MapReduce systems	
Interruption week	
Advanced techniques for dependable distributed systems	
	Evaluation

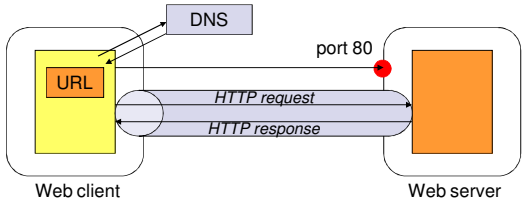
Brief history

- Basic idea
  - A set of distributed documents interlinked with *hypertex links*
- Initial objective (Tim Berners-Lee, CERN, 1989-1990)
  - Creating a tool for collaborative work on shared data, used by a distributed physicians community
  - Late 1993, 250 servers, 1% of the traffic of the Internet
- Real starting up (1994)
  - First Web browsers and search engines: Mosaic, Netscape, AltaVista, Yahoo!
  - Creation of the World Wide Web Consortium (W3C): *www.w3c.org*
  - Late 1994, ~10 000 servers
- Today (2010)
  - Intel: 100 000 servers
  - Facebook: 30 000 servers
  - An estimate of a total of 75 million servers

Basic elements of the Web

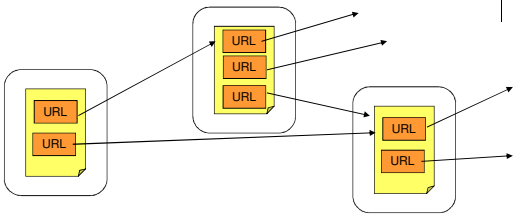
- A global naming space for resource identification
  - URL: Uniform Resource Location
- A protocol for client/server interaction and transfer of documents
  - HTTP: HyperText Transfer Protocol
- A markup language for the description of hypertext documents
  - HTML: HyperText Markup Language
- Extensions
  - Scripts languages (e.g. Java applets on the client-side, CGI or Java Servlets on the server-side)
  - Multiple types of documents (HTML, image, audio, video, etc.)

General principles of the World Wide Web



- The Web follows the **client/server architecture**
- The requested service is located by a **URL**
- The interaction (request-response) protocol is **HTTP**
- A Web server can simultaneously serve multiple clients

General principles of the World Wide Web (2)



- Web documents (e.g. HTML, XML, etc.) use **hypertext links**, i.e. *links to other documents*
- Hyperlinks are materialized by URLs
- This structure is the origin of the « Web »

Resource identification in the Web

- A resource (e.g. Web document) is identified in the Web by a URL (Uniform Resource Locator)
- A URL determines the location of a resource and the access protocol to that resource
- Example
  - http://www.upm.es/index.html
  - Protocol: HTTP
  - Web server location: www.upm.es
  - Requested resource in the Web server: index.html
- Other protocols
  - FTP (remote file access)
  - File (local file access)
  - mailto (SMTP mail)
  - etc.

URL interpretation

- The Web client interprets the beginning of the URL to determine
  - Which protocol to use (e.g. HTTP, FTP, etc.)
  - What is the location of the Web server (using DNS)
  - What is the port of the Web server (implicit default port of the protocol, or explicit port specified in the URL after the name of the server)
- The Web server interprets the end of the URL to determine
  - What is the requested resource
  - Default rules (e.g. index.html)

HTTP – HyperText Transfer Protocol

- HTTP: standard protocol of the World Wide Web
  - A client-server protocol, built on top of TCP/IP
  - Main use: between a Web browser and a Web server
- Main HTTP methods
  - GET: a method to return the whole Web page identified by the URL
  - HEAD: a method to return the head of the Web page identified by the URL (the head contains summary information such as title, date of creation, etc.)
  - PUT: a method to send a page content to the Web server and store it at the specified URL, replaces previous content if any
  - POST: similar to PUT, extends previous content
  - DELETE: a method to delete the content identified by the URL

HTTP – HyperText Transfer Protocol (2)

- HTTP methods need specific authorization rules, depending on Web client access rights and protection rules applied to resources on the Web server
- The response to an HTTP method has
  - a status code
  - E.g. 200 for Ok, 401 for Unauthorized, 404 for Not found, etc.
  - and possibly a result associated to the requested method (e.g. with the GET method)

HTML – HyperText Markup Language

- A markup language is a system for annotating a text (i.e. acontent) in a way which is syntactically distinguishable from that text (content)
- E.g. revision instructions by editors, traditionally written with a red pencil on authors' manuscripts
- Markup is typically omitted from the version of the text which is displayed for end-user consumption
- HTML includes structural markers (HTML tags)
- HTML has presentation semantics: its specification describes how the structured data is to be presented

HTML example

- Structure of an HTML document
  <HTML>
  <HEAD> head </HEAD> *title, date, meta-data*
  <BODY> body </BODY> *content of the document*
  </HTML>
- Presentation tags
  - Format
    <B> **bold text** </B> <I> *italic text* </I>
  - Titles
    <H1> title of level 1 </H1>
  - Paragraph
    <P> paragraph </P>

 line break

HTML example (2)

- Image inclusion

  <IMG
  src="http://www.somecompany.com/People/Ian/vacation/family.png" alt="A photo of my family at the lake.">

- Hyperlinks

  Visit W3Schools!
  - A Web browser can use a particular formatting to present such a hyperlink (e.g. in blue)
  - A click on this link with a Web browser results in the call of the GET method with the URL specified in the hyperlink

References

- The World Wide Web Consortium (W3C). http://www.w3c.org/
- Elizabeth Castro. HTML 4 for the World Wide Web. Peachpit Press, Oct. 1999.
- Leon Shklar and Rich Rosen. Web Application Architecture: Principles, Protocols and Practices. Wiley, May 2009.
- This lecture is partly based on lectures given by Sacha Krakowiak, http://sardes.inrialpes.fr/people/krakowia/