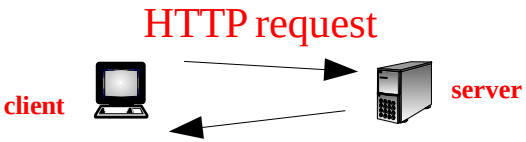
Outline

- DNS
- Charsets
- Email
- **Web**
- HTML

Web

- World Wide Web, www: was started by Tim John Berners-Lee in 1989 and developed in the 90s to provide an easy access to information in the Internet.
- Components:
  - Transport layer: TCP, well-known port: 80.
  - Application layer protocol: HyperText Transfer Protocol (HTTP). RFC1945 (HTTP-1.0 Y1996), RFC2616 (HTTP-1.1 Y1999).
  - HyperText Markup Language (HTML): Language used to format web documents.



Web browser:
- firefox
- explorer
- ...

HTTP response (HTML page)

Web server:
- apache
- ...

Source: wikipedia

Web elements

- Protocol
  - **HTTP** (HyperText Transfer Protocol)
- Information (format)
  - **HTML** (HyperText Markup Language)
- LINK to information
  - **URI** (Uniform Resource Identifier):

    URN (Name), **URL** (Locator)

Web – links

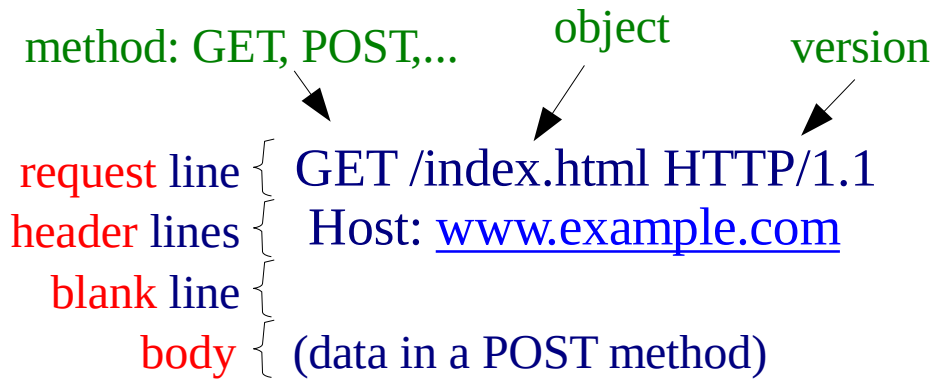
- Uniform Resource Identifier (URI) RFC3986
  - Generic syntax to identify a resource.
- Uniform Resource Locator (URL) RFC1738
  - Subset of URIs identifying the locating a resource in the Internet.
- The URL general syntax is

scheme://username:password@domain:port/path?query_string#fragment_id

  - scheme: Purpose, and the syntax of the remaining part. http, gopher, file, ftp...
  - domain name or IP address gives the destination location. The port is optional.
  - query_string: contains data to be passed to the server.
  - fragment_id: specifies a position in the html page.
  - Examples:
    - http://tools.ietf.org/html/rfc1738
    - http://147.83.2.135
    - http://studies.ac.upc.edu/FIB/grau/XC/#Practs
    - file:///home/llorenc/gestio/2010/cd/autors.html
    - http://www.amazon.com/product/03879/refs9?pf_ra=ATVPD&pf_rd=07HR2

Web – HTTP Messages, RFC2616

- Client (HTTP request):



- Header: Allows the client to give additional information about the request and the client itself.
  - Host:
    - host of the resource being requested
    - mandatory in HTTP/1.1

Web – HTTP Messages, RFC2616

- **Methods:**
  - GET Typical command. Requests an object.
  - POST Request an object qualified by the data in the body. This data is the contents of the HTML form fields, provided by the client.
  - HEAD the server returns only the header
  - OPTIONS request communication options
  - PUT store entity
  - PATCH modify an existing resource
  - DELETE delete entity
  - TRACE final recipient echoes the received message back
  - CONNECT used with a proxy
- NOTES
  - Most used: GET, POST
  - Safe and mandatory: GET, HEAD

Web – HTTP Messages, RFC2616

- POST uses MIME types: application/octet-stream, to send raw binary data, and application/x-www-form-urlencoded, to send name-value pairs. Example:

request line { POST /login.jsp HTTP/1.1

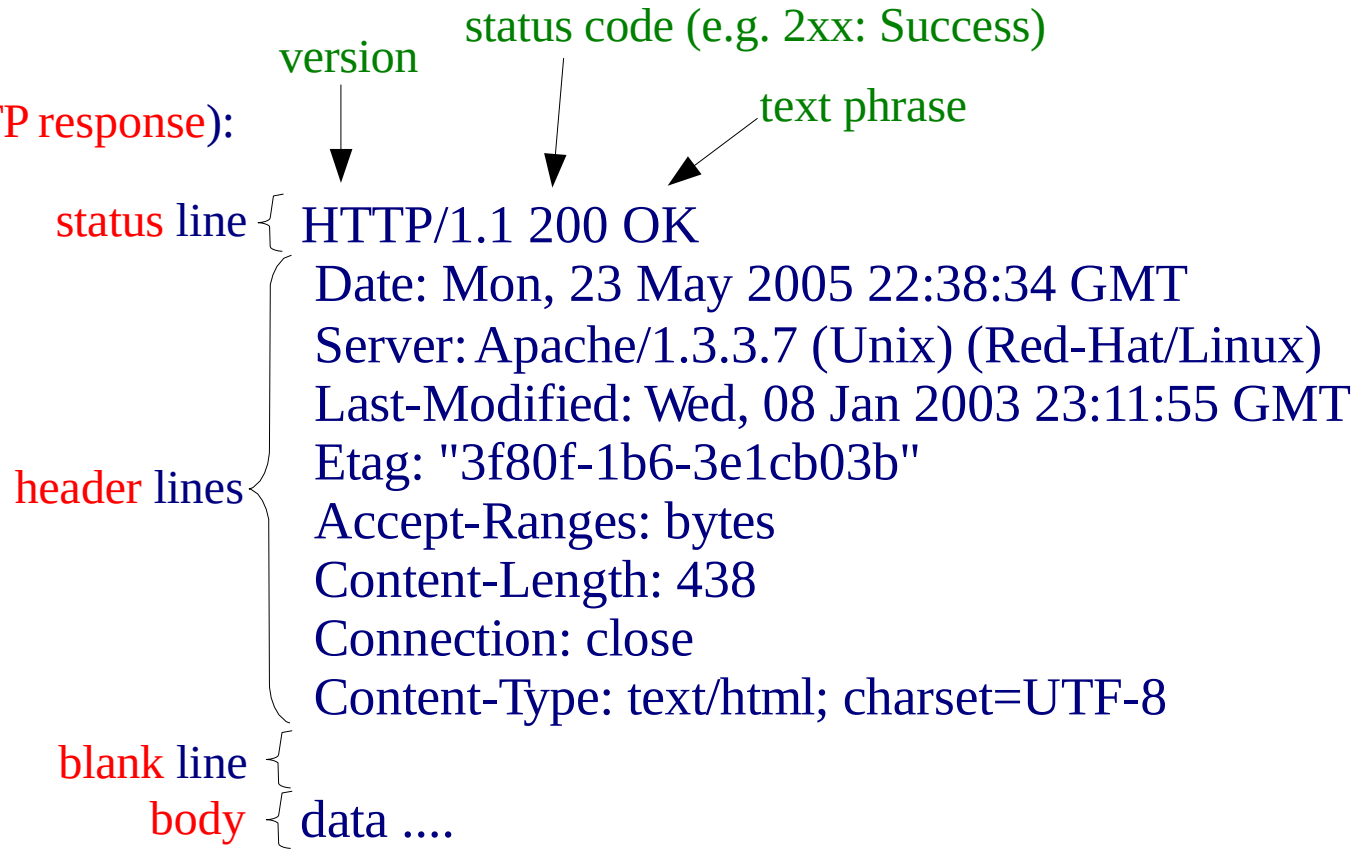
header lines {
Host: www.mysite.com
User-Agent: Mozilla/4.0
Content-Length: 27
Content-Type: application/x-www-form-urlencoded

blank line {

body { userid=llorenc&password=mypassword

Web – HTTP Messages, RFC2616

- Server (HTTP response):



Web – HTTP Messages, RFC2616

- Header
  - Last-Modified: date, used in conditional retrieval.
  - Etag: id, used in conditional retrieval.
  - Connection: keep-alive/close, controls whether or not the network connection stays open after the current transaction.
  - Accept: <MIME_type>/<MIME_subtype>, acceptable mime types.
  - ...

Web – Persistent/non Persistent connections

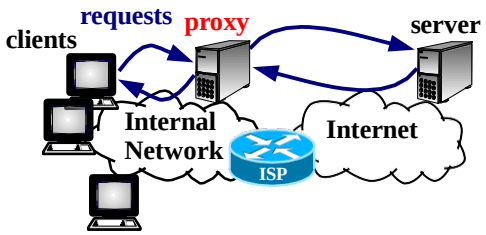
- Non persistent (default in HTTP/1.0): The server closes the TCP connection after every object. E.g, for an html page with 10 jpeg images, 11 TCP connections are sequentially opened.
- Persistent (default in HTTP/1.1) : The server maintains the TCP connection open until an inactivity time. All 11 objects would be sent over the same TCP connection.
- Persistent connections with pipelining (supported only in HTTP/1.1): The client issues new requests as soon as it encounters new references, even if the objects have been not completely downloaded.

Web – Caching and Proxies

- Caching: The client stores downloaded pages in a local cache. Conditional GET requests are used to download pages if necessary. It can use the Date and/or Etag:

```
GET /index.html HTTP/1.1
Host: www.example.com
If-Modified-Since: October 21, 2002 4:57 PM
If-None-Match: "686897696a7c876b7e"
```

- Proxy server: Acts as an intermediary for requests from clients.
  - Advantages:
    - Security (the proxy may reject the access to unauthorized servers)
    - Logs
    - Caching
    - Save public IP addresses (only the proxy may have access to the Internet)
    - ...



Web – web based applications

- Components:
  - Presentation: A web browser (client side).
  - Engine generating “on the fly” HTML pages (server side).
    - Languages:
      - Java.
      - Hypertext Preprocessor (PHP): Embedded program language and HTML code (http://www.php.net).
      - Other: ASP, CGI, ColdFusion, Perl, Python...
  - Storage: a database (e.g. mysql).

- Benefits:
  - Fast to deploy and upgrade (only server side).
  - Only a compatible browser is required at the client side.
  - Provide cross-platform compatibility (i.e., Windows, Mac, Linux, etc.)