

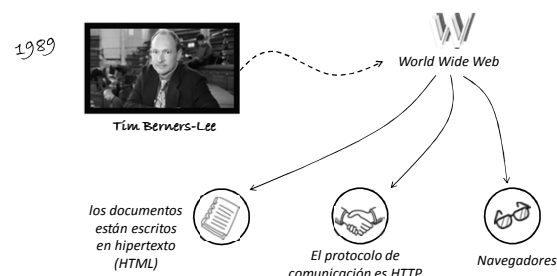
Ingeniería de Aplicaciones Web

Diego C. Martínez

Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Web



En 1994 Berners-Lee funda el **World Wide Web Consortium (W3C)** en el MIT, con apoyo de DARPA.
La idea central era asegurar la compatibilidad por medio de la definición de estándares, denominados **W3C Recommendations**.

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

HTTP

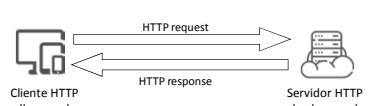
HTTP es el protocolo de red para la Web.

→ **"Hypertext Transfer Protocol"** → protocolo de aplicación para la entrega de documentos variados

protocolo de comunicación define:

- ✉ tipo y estructura de los mensajes
- 💬 las reglas del diálogo

ubicables por medio del URL.



HTTP es un protocolo sin estado, o *stateless protocol*:
no mantiene información sobre la conexión entre transacciones

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

URLs - sintaxis

El nombre estandarizado de un recurso en Internet, que lo identifica unívocamente se denomina URL - Uniform Resource Locator.

Es un subconjunto de una clase más general de identificadores (URI - Uniform Resource Identifier)

URI identifica
URL localiza

"Epic Games" URI

"Epic Games, Inc. 620 Crossroads Blvd Cary, NC 27518" URL
URL

Una URL describe el recurso por su *ubicación* (dónde está) y el *protocolo* que se entiende para obtenerlo (HTTP, HTTPS, FTP, SMTP, etc).

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

URLs - sintaxis

<scheme>://<user>:<password>@<host>:<port>/<path>?<query>#<frag>	
<scheme>	Protocolo para acceder al recurso
<user>:<password>	Usuario y password requerido para acceder al recurso
<host>:<port>	Nombre del host o dirección IP del servidor que tiene el recurso + port
<path>	Nombre local del recurso
<query>	Parámetros de entrada para el recurso
<frag>	Referencia a una porción del recurso (de uso para el cliente)

Ejemplos

http://cs.uns.edu.ar → port por defecto

http://www.uns.edu.ar:80

http://www.fis-ski.com/uk/disciplines/alpineskiing

https://www.bancogalicia.com.ar

ftp://ftp.prep.ai.mit.edu/pub/gnu

ftp://anonymous:my_passwd@ftp.prep.ai.mit.edu/pub/gnu

https://www.youtube.com/watch?v=ej9vLgraWXk

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

URLs - sintaxis

Una URL puede ser absoluta o relativa

las URL absolutas contienen toda la información necesaria para obtener el recurso
Las anteriores son todas absolutas

las URL relativas son interpretadas en relación con otra URL

`Go!`

La URL relativa se completa con

una URL base declarada en el recurso
`<base href="http://..." >`

la URL base del recurso que contiene la URL relativa

(VER RFC 2396 de la W3C)

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

URLs - sintaxis

Con el objetivo de que las URLs sean generales y portables por protocolos diferentes de las capas inferiores, algunos caracteres requieren codificación:

`http://www.unhost.com.ar/~usuario/un articulo para la web.html`
↓
`http://www.unhost.com.ar/%7usuario/un%20articulo%20para%20la%20web.html`

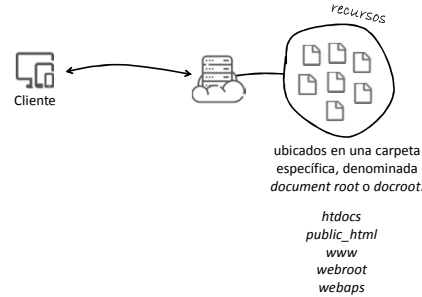
Otros caracteres están reservados y no pueden usarse sin codificación, entre ellos:
% / . _ # ; : ? < > { }

Visitar
`http://www.w3.org/Addressing`

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Servidores web

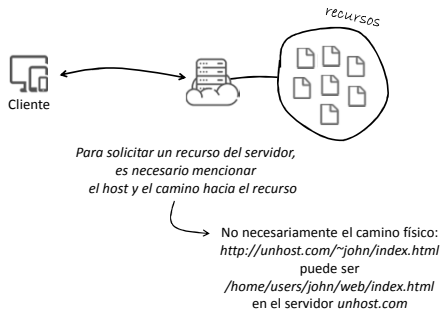
El servidor web actúa como una interfaz entre un conjunto de recursos y los clientes.



Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Servidores web

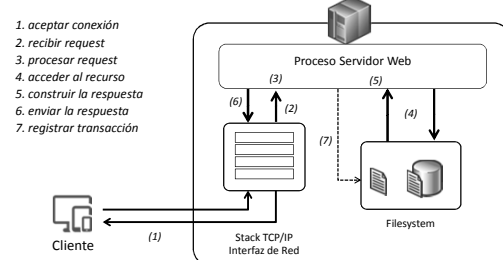
El servidor web actúa como una interfaz entre un conjunto de recursos y los clientes.



Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Servidores web

El esquema de trabajo de un servidor web es el siguiente:



Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Mensajes HTTP

Como protocolo de comunicación, HTTP define el tipo y estructura de los mensajes que se envían y las reglas del diálogo entre los dos participantes.

- mensajes hacia el servidor → *inbound messages/ requests*
- mensajes desde el servidor → *outbound messages / responses*

La estructura de los mensajes es la misma para todos

Línea inicial	Diferente para los requerimientos y respuestas.
Header	
...	Cero o más encabezamientos (header lines)
Header	
CRLF	Línea en blanco
Cuerpo del Mensaje	Contenido del mensaje: archivos, consultas a bases de datos, datos binarios, etc.

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Mensajes HTTP

Una línea inicial de request tiene tres partes

nombre del método en mayúsculas path del recurso solicitado versión de HTTP utilizada

`GET /path/to/file/index.html HTTP/1.0`
`POST subscribir.php HTTP/1.0`

Una línea inicial de response tiene tres partes

versión de HTTP código de status descripción del status

`HTTP/1.0 200 OK`
`HTTP/1.0 404 Not Found`

El primer dígito del código de status identifica la categoría:

- 1xx indica un mensaje de información únicamente
- 2xx indica éxito en general
- 3xx redirecciona el cliente a otro URL
- 4xx indica un error en la parte del cliente
- 5xx indica un error en la parte del servidor

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

El protocolo HTTP – códigos comunes

- ▮ **200 OK**
El pedido tuvo éxito, y el recurso es retornado en el cuerpo del mensaje.
- ▮ **404 Not Found**
El recurso solicitado no existe
- ▮ **301 Moved Permanently**
- ▮ **302 Moved Temporarily**
- ▮ **303 See Other (HTTP 1.1 only)**
El recurso se movió a otro URL y debe ser pedido automáticamente por el cliente.
- ▮ **500 Server Error**
Error inesperado. Típicamente un error en algún script alojado en el servidor que impide que se ejecute correctamente.

El detalle completo de los códigos puede encontrarse en las *especificaciones de la W3C*:
 HTTP 1.0: *RFC 1945*
 HTTP 1.1: *RFC 2616*

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

El protocolo HTTP – header lines

Las líneas de encabezamiento proveen información sobre el pedido (request) o la respuesta (response) o el objeto que se está enviando en el cuerpo del mensaje.

Header-Name: value

HTTP 1.0 define 46 headers, pero ninguno es requerido.
 HTTP 1.1 define 46 headers y requiere **Host** en los pedidos realizados.

Algunos headers del *cliente*:

- ▮ **From**: es la dirección de mail de quien realiza un pedido.
- ▮ **User-agent**: es el programa que realiza el pedido, de la forma "*Nombre/x.xx*"
Ejemplo: **User-agent**: Mozilla/3.0Gold

Algunos headers importantes del *servidor*:

- ▮ **Server**: identifica el software del servidor, en el mismo formato "*Nombre/x.xx*"
Ejemplo: **Server**: Apache/1.2b3-dev
- ▮ **Last-modified**: es la fecha de modificación del recurso otorgado
Ejemplo: **Last-Modified**: Fri, 31 Dec 1999 23:59:59 GMT

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Métodos HTTP

- GET → *Obtiene un documento del servidor. El mensaje no tiene cuerpo.*
- POST → *Envía datos al servidor para el procesamiento. Los datos se especifican en el cuerpo del mensaje.*
- HEAD → *Solicita solo los encabezados de un documento. El mensaje no tiene cuerpo.*
- PUT → *Almacena el cuerpo del request en el servidor, bajo el nombre indicado como URL.*
- TRACE → *Hace un rastreo del mensaje al servidor. El mensaje no tiene cuerpo*
- OPTIONS → *Determina qué métodos pueden operar en un servidor. El mensaje no tiene cuerpo*
- DELETE → *Remueve un documento del servidor. El mensaje no tiene cuerpo*

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Requests - responses

Si el mensaje incluye un cuerpo, entonces habrá dos *headers* que describen ese cuerpo:
Content-Type: *MIME-type* del cuerpo, como *text/html* or *image/gif*.
Content-Length: número de bytes en el cuerpo.

```

GET /path/file.html HTTP/1.0
From: home@snpp.com
User-Agent: SuperBrowser/1.0
[CRLF]

HTTP/1.0 200 OK
Date: Fri, 31 Dec 1999 23:59:59 GMT
Content-Type: text/html
Content-Length: 1354

<html>
<body>...</body>
</html>
    
```

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

El protocolo HTTP - métodos

Además de **GET**, los métodos más usados son **HEAD** y **POST**.

El método **HEAD** solicita al servidor únicamente los headers de la respuesta, sin cuerpo. *Permite ver si existe algún recurso, ver si ha sido modificado, etc.*

El método **POST** es esencial para la interacción web. Este método envía información al servidor para ser procesada por quien corresponda, como un script CGI o PHP.

Procesará los datos enviados

```

POST /empleados/despedir.cgi HTTP/1.0
From: burns@snpp.com
User-Agent: HTTPTool/1.0
Content-Type: application/x-www-form-urlencoded
Content-Length: 23

nombre=Homer&sector=7G
    
```

Cuerpo del mensaje

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

El protocolo HTTP – POST vs GET

```

POST /empleados/despedir.cgi HTTP/1.0
From: burns@snpp.com
User-Agent: HTTPTool/1.0
Content-Type: application/x-www-form-urlencoded
Content-Length: 23

nombre=Homer&sector=7G
    
```

```

GET /empleados/despedir.cgi?nombre=Homer&sector=7G HTTP/1.0
    
```

Sin cuerpo de mensaje

GET debe usarse *preferentemente* cuando los datos son pocos. En otro caso debe usarse el método **POST**.

En ambos casos, el URL será codificado (*URL-encoded*) para transmitir los caracteres especiales (espacio, &, %, etc).

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Documentos

Recordemos:
la Web surge como un servicio de Internet para estudiar y explorar documentos multimediales.

documentos de texto, con imágenes y/o sonido.
facilidades de exploración de información relacionada.



"The frontiers of a book are never clear-cut, because it is caught up in a system of references to other books, other texts, other sentences: it is a node within a network of references"

The archeology of knowledge
Michel Foucault, 1969

Esto requiere una forma especial de estructurar la información, para interrelacionarla y visualizarla adecuadamente.

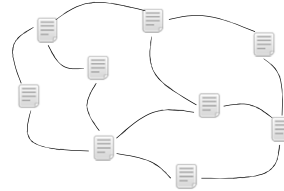
Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Hipertexto

Sabemos que la idea de *hipertexto* es anterior al surgimiento de la web, e incluso de las redes globales...

"Hypertext is a non-sequential writing, a text that branches and allows choices to the reader, best read at an interactive screen. As popularly conceived, this is a series of text chunks connected by links which offer the reader different pathways"

Theodor Nelson (1960)



Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

HTML

Propuesta de Berners-Lee:
Enriquecer un documento con marcas especiales para estructurar, vincular y componer la información, de manera tal de "convertirlo" en hipertexto.

+ <marcado> = HyperText Markup Language.

El aporte de Berners-Lee no se limita al HTML.
De hecho, ya existían lenguajes de marcado (SGML- Standard generalized markup language).

Su propuesta es una red global de documentos, unidos por

"global hypertext links"

Berners-Lee implementa el primer browser de documentos HTML y el primer servidor HTML de la historia, en una computadora NeXT.

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Estructura de HTML

Las marcas (tags) se encierran entre < y >

La versión original (de Berners-Lee) era simple, pero con posterioridad diferentes navegadores comenzaron a agregar sus propias modificaciones ("browsers-war").

Actualmente la definición de HTML de la W3C es estricta, pero los navegadores siguen procesando documentos sin la estructura apropiada.

Cada tag tiene un nombre y posiblemente varios atributos (attr=valor)

opening tag *closing tag*
datos<tag>contenido</tag>datos
datos<tag attr1=valor1 attr2=valor2>contenido</tag>datos
datos<tag_unario>datos

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Primera versión de HTML

Tags de la primera versión de HTML:

<TITLE> ... </TITLE> Título del documento
 ... Hipervínculo.
<ISINDEX> Indicador de una página índice
<PLAINTEXT> Texto a ser tomado en forma literal
<LISTING> ... </LISTING> Texto de font de tamaño fijo, tomado literalmente
<P>, <H1>, <H2>, <H3>, <H4>, <H5>, <H6> Párrafo y encabezados
<ADDRESS> text ... </ADDRESS> Direcciones, contacto, footers
<HP1>...</HP1> <HP2>... </HP2> Texto resaltado. Ya no se usa.

Listas descriptivas

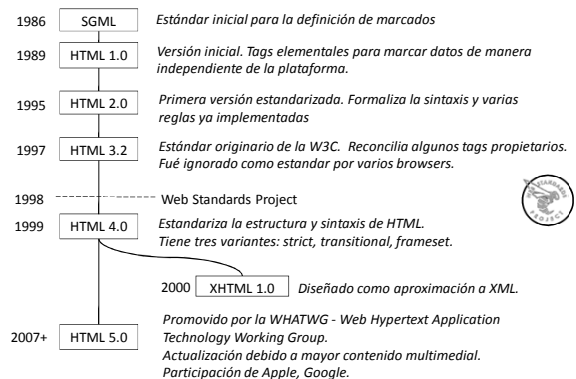
<DL>
<DT>Term<DD>definition paragraph
<DT>Term2<DD>Definition of term2
</DL>

Listas no numeradas

 list element
 another list element

Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Evolución de HTML



Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

