

2021-09-14

Anton-Bravo, Adolfo

Contents

1	Contenidos	1
1.1	DATOS	1
1.1.1	CSV	2
1.2	Cygwin	2
1.2.1	Mirrors	2
1.2.2	¿Qué es Cywgin?	3
1.3	MACOSX	3
1.4	SOFTWARE	4
1.5	Interfaces de uso de software	4
1.6	URL	4
1.6.1	Pregunta: ¿puede ser una carpeta un recurso, una <i>URL</i> ?	6
1.7	localhost	6
1.8	Pendientes	7
2	Pruebas	7

1 Contenidos

1.1 DATOS

- ¡Para el Periodismo de Datos necesitamos datos!
- Muchas veces es complicado encontrarlos pero también hay muchos sitios donde buscar e investigar datos ya preparados. Los iremos viendo.
- Tened en cuenta que en Github también vamos a encontrar repositorios que tengan datos o repositorios con recopilatorios de datos.

- Un truco/consejo: cuando queráis buscar algo en Github añadid el adjetivo "awesome" (fantástico, chulo, maravilloso) a la palabra que buscáis. Por ejemplo, buscad "awesome data" para buscar datos.
- Otro sitio interesante es kaggle.com, ya lo veremos más adelante.

1.1.1 CSV

- En la primera sesión hablamos de una de las cosas que era *Excel*: un programa para visualizar datos tabulados.
- Hay distintos tipos de datos. Hoy hemos mencionado los *CSV*, *Comma Separated Values* o *Valores Separados por Comas*.
- Son herederos de los *TSV*, *Tab Separated Values* o valores separados por tabuladores.
- Tened en cuenta que el tabulador crea un espacio más grande (4 o 5 caracteres) frente al espacio de la barra espaciadora que deja uno. Los espacios pueden estar presentes en algunos datos, como un nombre o una dirección, pero los tabuladores no. Por eso se utilizaba esta separación para que el ordenador la interpretara como otro campo de la tabla.
- El próximo día veremos esto con atención.

1.2 Cygwin

1.2.1 Mirrors

- Alguien ha dicho que dio error la página <https://cygwin.com/>. La verdad es que nunca me ha pasado pero si da error se puede volver a intentar más tarde.
- Algunas personas os habéis quedado en el punto en el que dudáis porque os da a elegir entre varios **mirrors** la descarga del software.

Esto es un guiño a la historia de Cygwin y de la cultura de Internet.

- Tened en cuenta que al principio las conexiones eran muy lentas. Algunas páginas con mucho contenido o contenido pesado, como puede ser el software, tenían **mirrors**, es decir, espejos, copias de ese contenido en otras para no saturar la original.

1.2.2 ¿Qué es Cygwin?

Aunque lo he explicado en clase, en su página web pone: *What is it?*

- Un conjunto amplio de herramientas GNU y de código abierto que ofrecen una funcionalidad similar a una distribución Linux en Windows.
- Una librería DLL (Dynamic Link Library) para Windows que ofrece buena parte de las funcionalidades de la API de POSIX.
- POSIX (*Portable Operating System Interface* o interfaz portable del sistema operativo) es una familia de estándares especificados por la IEEE Computer Society (*Institute of Electrical and Electronics Engineers*) para mantener la compatibilidad entre los sistemas operativos.

1.3 MACOSX

- Ha habido personas con problemas para instalar XCode.
- Parece que os pide actualizar software y/o sistema operativo para que tengáis compatibilidad con el XCode.
- Mirad antes si disponéis de otras versiones anteriores de XCode que valgan para vuestra versión del sistema operativo y así elegís entre actualizar todo, con lo que implica, o bien simplemente instalar una versión de XCode compatible con vuestra versión del sistema operativo.

1.4 SOFTWARE

Importante en este punto recordar dos cosas:

- Suele ser mejor tener las últimas versiones estables de un software. Pero no es una regla fija, hay situaciones diversas.
- Sin embargo, a veces, todavía se puede tener versiones antiguas siempre que estén soportadas, es decir, quienes hacen ese software se aseguren de estar probando que todavía funciona.
- Cuando el software deja de estar soportado empieza a recaer la responsabilidad exclusivamente en nosotros. Eso es lo que ocurrió, por ejemplo, cuando Windows dejó de soportar primero XP y luego W7. También ocurre en MacOSX y en cualquier otro software.

1.5 Interfaces de uso de software

Hay tres tipos de interfaces aunque hemos visto dos:

- *CLI* de *Command Line Interfaces* o interfaces de línea de comandos.
- *GUI* de *Graphical User Interfaces* o interfaces gráficas de usuario.
- Las de los móviles, capacitivas.

1.6 URL

- *URL* o *Unified Resource Locator* o *localizador de recursos uniforme* es la dirección web.
- Como dice Mozilla.org, la *URL* junto con el hipertexto (expresado en el *HTML*, aunque de esto exactamente no hemos hablado todavía) y *HTTP* son los conceptos claves de la Web.

- Tened en cuenta que la Web es una telaraña mundial, uno de los servicios de red de Internet. Sobre esta web lo que encontramos son "recursos", y a eso elude la *URL*.
- Cada URL valida apunta a un único recurso.
- Los recursos que normalmente conocemos son páginas web, pero también pueden ser documentos CSS, imágenes, JS, archivos de datos... cualquier archivo disponible en un servidor web.
- En una URL se pueden identificar 3 partes fácilmente. Veamos este ejemplo del *The Guardian*: <https://www.theguardian.com/world/2021/sep/14/taliban-governor-of-helmands-message-to-west-come-back-with-money-n>

1. `https://`, que indica el protocolo usado, `https` en el dominio que sea. La separación entre protocolo y dominio se realiza con `://`.

2. El dominio `www.theguardian.com`, es decir, lo que va entre `://` y la primera `/`. El dominio se lee de derecha a izquierda en orden de importancia. `.com` es el *TLD, Top Level Domain* o dominio de primer nivel. Una entidad lo gestiona y esa u otra que revende el servicio ha alquilado `theguardian.com` al medio inglés. La última (recordad, de derecha a izquierda) `www` es un subdominio de `theguardian.com`. A efectos de la Web o de funcionamiento, daría igual que no estuviera pero es también una forma de honrar al proyecto de Tim Berners Lee *WWW, World Wide Web* o gran telaraña mundial; y es también una forma de organizar contenidos o cosas que sirves a través de tu dominio. Por ejemplo, si tienen un servidor de correo, quizás se llame `imap.theguardian.com`. Estos subdominios sirven para identificar por los humanos los servicios.

3. La estructura de carpetas del servidor web, todo lo que hay a la derecha de la barra inclinada después de `.com`, es decir, `/world/2021/sep/14/taliban-governor-of-helmands-message-to-west-come-back-wi`. Cuando digo "estructura de carpetas del servidor web" es porque

no tienen por qué corresponderse con una estructura de carpetas reales. Probablemente solo lo último, el nombre del artículo, se corresponda con una carpeta.

1.6.1 Pregunta: ¿puede ser una carpeta un recurso, una URL?

- En realidad, al igual que todo software ha de tener al menos un archivo README, toda carpeta web tiene dentro al menos un archivo `index.html`.
- Entonces, no sé deciros cuándo, se empezó a dejar de mostrar el nombre del archivo `index.html` cuando se visitaba una carpeta que efectivamente tenía ese nombre de archivo.
- De esa manera el nombre del archivo se traslada al nombre de la carpeta.

1.7 localhost

- Os presento a vuestro ordenador, el localhost.
- localhost es el nombre de dominio de ese recurso, vuestro propio ordenador.
- Una cosa que vimos con los servidores web y las peticiones http desde el navegador como cliente fue la arquitectura cliente-servidor. Pero no dijimos que esta opera en todo lo que hace el ordenador.
- Y para ello, se conoce a si mismo como localhost.
- Si localhost es el dominio, su dirección IP es 127.0.0.1, que es la dirección IP reservada para localhost.

- Se puede echar un vistazo al archivo `/etc/hosts` (en GNU/Linux y Mac, en Windows buscad `hosts` en el buscador de Windows) y ver qué sale:

```
language=bash,label= ,caption= ,captionpos=b,numbers=none cat /etc/hosts
```

1.8 Pendientes

- Hipertexto
- Formatos de datos
- Tipos de datos

2 Pruebas

- Qué tipos de interfaces de datos hay
- Qué tipos de datos hay?
- Elige una URL de una noticia de un medio de comunicación y explícala tal como hicimos en clase.
- ¿Qué significa el funcionamiento "cliente-servidor"?
- ¿Qué relación tiene el formato CSV con Excel?