

Descargas masivas con wget

(webs, mp3, videos, etc...)

Hoy le toca el turno a una pequeña utilidad poco conocida por los neófitos, pero muy útil para realizar descargas desde la shell de Linux, estamos hablando del fantástico wget. Muchos de nosotros hemos utilizado el wget para hacer alguna descarga y poco más...pero eso llega a su fin, ya que hoy vamos a ver como convertir el wget en "El Gestor de Descargas de Linux".

Lo primero es instalarnos el wget, yo creo que viene por defecto con cualquier instalación base, pero por si acaso ya sabeis "apt-get install wget" o "rpm -i wget", etc...

Una vez instalado podemos realizar la descarga de un archivo sin más que teclear:

- `wget ruta_del_archivo`

Esto nos puede parecer muy bonito...pero si solo podemos hacer esto vaya un truño. Tranquilos, que hay más; vamos a ver como descargarnos todos los archivos mp3 de una lista de páginas que le indiquemos, el comando a emplear será el siguiente (después explicaré qué significa cada opción paso a paso):

- `wget -r -l1 -H -t1 -nd -N -np -A.mp3 -erobots=off -i ~/listurl.txt`

Ahora veremos que es cada opción de las que hemos añadido:

- `-r` indica a wget que haga la descarga de forma recursiva, es decir va a una URL, y sigue cada enlace que encuentra.
- `-l1` esto le dice a wget que siga los enlaces, pero solo hasta 1 nivel, es decir que no siga los enlaces encontrados en las páginas enlazadas.
- `-H` esto se usa para que baje archivos en otros dominios, no sólo en el actual.
- `-np` indica que no siga enlaces a directorios padre.
- `-A.mp3` le dice a wget que sólo baje archivos terminados en mp3, evidentemente podemos indicarle el patrón que queramos, pero en este caso sólo nos interesan los mp3.
- `-nd` esta opción sirve para que wget no nos cree la estructura de directorios, sino que lo descarge todo en el mismo directorio.
- `-t1` se usa para que si falla una descarga haga solo 1 reintento más, así no se nos atascará en un archivo si este falla.
- `-erobots=off` es para que ignore los archivos robots.txt
- `-N` esto es para no descargar el mismo archivo 2 veces, a menos que sea una copia más reciente.
- `-i listurl.txt` este último comando sirve para indicarle cual es el archivo en el que tenemos la lista de URL's que queremos saquear :P

Se podrían añadir algunas opciones más, como por ejemplo:

- `-b` esto pondría la descarga en background (segundo plano), así podemos hacer más cosas mientras se bajan los archivos.
- `-o archivo_de_logs.log` con esto le decimos que guarde en un log lo que va haciendo...es útil cuando mandamos el proceso a un segundo plano, para ver como ha ido todo.

Hay muchas más opciones que hacen de wget una verdadera maravilla, para consultarlas todas ya sabéis:

- `man wget`

A darle caña al wget, un saludo a todos!

Fuentes: `man wget` y [Jeffre Ven Blog](#)

Escrito por Tolito viernes, 06 de octubre de 2006

Manual de wget

Cuántas veces el usuario novel de Linux al cambiar su windows de toda la vida por el sistema del pingüino se ve perdido en medio de programas que no conoce o bien necesita una herramienta y desconoce su homólogo en el mundo linuxero.

Y claro está, en su búsqueda de programas se centra en las “ventanitas de toda la vida” y en los ya clásicos golpes de ratón, ignorando que en Linux hay un arma poderosísima: la consola de comandos.

Este manual se centra en uno de esos comandos maravillosos que hace tanto o más que un programa gráfico, estamos hablando de *wget*, comando que nos permitirá bajarnos webs enteras, archivos, bajar algo de un ftp, etc, pero sobre todo veremos como construirnos una especie de p2p para poder continuar las descargas siempre que nos apetezca.

Lo primero un pequeño resumen de las principales opciones de *wget*:

Descargar una pagina web o un fichero:	wget http://www.servidor.com/fichero.zip
Guardar información que aparece por pantalla en un fichero log:	wget -o log.txt http://www.servidor.com/fichero.zip
Para obtener mas información sobre la descarga:	wget -d http://www.servidor.com/fichero.zip
Para que haga la descarga sin mostrar ningún log:	wget -q http://www.servidor.com/fichero.zip
Descargar las URLs que aparezcan en un fichero:	wget -i urls.txt
Continuar una descarga que se cortó:	wget -c http://www.servidor.com/fichero.zip
Descargar algo que requiere usuario y clave:	wget --http-user=usuario --http-passwd=clave http://www.servidor.com/fichero.zip
Aparentar ser un navegador distinto:	wget --user-agent=Opera http://www.servidor.com/fichero.zip
Descargar una pagina con todos los elementos gráficos, javascript, etc:	wget -p http://www.servidor.com
Descargar toda una web:	wget -r http://www.servidor.com
Descargar hasta tres niveles de profundidad de una web:	wget -r --level=2 http://www.servidor.com
Descargar un archivo de un ftp:	wget ftp://ftp.servidor.com/fichero.zip

Bien, ¿y cómo podemos descargar algo muy grande siempre que nos apetezca?. Para mayor comodidad primero crearemos un directorio para albergar todas nuestras descargas:

mkdir descargaswget

Y entramos en él:

cd descargaswget/

Ahora creamos un archivo con nuestro editor favorito:

nano archivo

Y dentro del mismo escribimos lo siguiente:

wget -c www.servidor.com/archivoquenosqueremosbajar

Lo siguiente es hacerlo ejecutable:

chmod +x archivo

Y siempre que queramos seguir bajándolo, entraremos en el directorio que hemos creado y lo ejecutaremos:

./archivo

Él solito se lo ira bajando y detectando la cantidad de fichero que ya se ha bajado, finalizando la descarga cuando lo haya bajado todo. E incluso podemos escribir dentro del archivo una lista de cosas a bajar, cuando ejecutemos el archivo él mismo irá mirando por el principio de la lista viendo si ya se han bajado los primeros elementos.

Este es un claro ejemplo de cómo no necesitamos programas gráficos para desenvolvernos por internet.