

Comando curl

La primera opción que he investigado es la de **guardar una página web** con el parámetro `-o`. Con esta opción podemos guardar una página web en un archivo html en nuestro disco duro. Esto solo creará una copia del HTML servido tal cual, sin el resto de ficheros estáticos como CSS, imágenes, o JavaScript. Esto se traduce en que, si en un futuro cambian de ubicación estos ficheros o no disponemos de conexión a internet, no se cargarán en la web almacenada.

La segunda opción investigada es la de obtener **cabeceras de una web** con el parámetro `--head`. En mi caso la he probado con la popular web de noticias Xataka:

```
sergio@sergio:~$ curl --head https://www.xataka.com
HTTP/2 200
content-type: text/html; charset=UTF-8
date: Thu, 24 Oct 2019 08:40:26 GMT
cache-control: public, s-maxage=60
surrogate-control: content="ESI/1.0"
vary: Accept-Encoding
x-clientip: 185.166.213.83
countrycode: ES
x-cache-hits: 26
accept-ranges: bytes
x-cache: Miss from cloudfront
via: 1.1 ea6b6651a564f3c1a19b54389dlf51e9.cloudfront.net (CloudFront)
x-amz-cf-pop: MXP64-C2
x-amz-cf-id: -xORKiwjWZcacYSHYVeDO3WxrG-GDn3tWBq5jj6Sf5tZuIZhNxMPYw==
```

En primer lugar, devuelve el código de respuesta HTTP, que ha resultado ser `200` (petición correcta). En la siguiente línea el tipo de contenido que devuelve el servidor (en este caso un documento HTML con la página web solicitada). También se recibe la fecha en GMT del momento en el que se envió la respuesta, la IP del cliente (nosotros), el país y más información que comúnmente incluyen las cabeceras.

He probado este mismo comando sobre una página que redirige, como es el dominio `msn.es` que redirige a <https://msn.com/es-es>, y este es el resultado:

```
sergio@sergio:~$ curl -L https://msn.es
curl: (51) SSL: no alternative certificate subject name matches target host name 'msn.es'
sergio@sergio:~$ curl --head msn.es
HTTP/1.1 301 Moved Permanently
Content-Length: 147
Content-Type: text/html; charset=UTF-8
Location: http://www.msn.com/es-es
Server: Microsoft-IIS/8.5
X-Powered-By: ASP.NET
Date: Thu, 24 Oct 2019 08:50:50 GMT
```

En `Location` podemos ver la dirección de destino de la redirección 301.

Precisamente para probar el último parámetro (-L) he empleado este mismo dominio, ya que esta opción se encarga de seguir el redireccionamiento para devolver el código de la página:

```
sergio@sergio:~$ curl -L msn.es
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE HTML PUBLIC "-//WAPFORUM/DTD XHTML Mobile 1.2//EN" "http://www.openmobilealliance.org/tech/DTD/xhtml-mobile12.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="es-ES" dir="ltr">

<head>
<!-- data-info:v:20191018_19030779;a:82de4549-d4a0-4314-93c6-6ff772069f41;cn:43;az:{did:b24a0ea2b3ba45a59fcd4d299c5ebc1, rid:
pc;th:start;PageName:startPage;m:es-es;cb;;l:es-es;mu:es-es;ud:{cid:,vk:homepage,n:,l:es-es,ck:};xd:BBpm8Mn;ovc:f;al::fxd:f;xdp:
258cf,audexedgelcf,samrtb-n,moneyhp3cf,compliancehplcf,artgly4cf,onetrustpoplive,msnapplcf,jsltlemetry;userOptOut:false;userO
<meta name="robots" content="index, follow"/>
<link rel="canonical" href="http://www.msn.com/es-es" />

<title>Hotmail, Outlook, noticias y hora#243;scopo en MSN Espa#241;a y Microsoft News</title>
<meta name="description" content="Consulta en MSN Espa#241;a y Microsoft News las #252;ltimas noticias de Espa#241;a y el mundo, famosos,
<meta name="viewport" content="width=device-width,initial-scale=1, maximum-scale=1" />
<link rel="stylesheet" href="//static-global-s-msn-com.akamaized.net/hp-neu/es-es/homepage/_sc/css/b443689a-4ab53bd3/direction
b-52fe22-a648eab2/86-b413d7-63b61fa3/dl-7801d5-68ddb2ab/7a-660b57-4534563a/75-a00681-654638bf/e7-7a4296-60b40184/25-a380ce-68bd
amrtb-n,moneyhp3cf,compliancehplcf,artgly4cf,onetrustpoplive,msnapplcf,jsltlemetry&csopd=20191004234831&csopdb=201910
<script type="text/javascript"><![CDATA[
(function(n,t){function h(n,i,r){typeof n!="string"&&(x=i,i=n,n=t);i&&i.splice||(x=i,i=[]);n==it?a=!0:n==rt&&(v=!0);l(n,i,r,!l
ned"&&(u[o]=[]);u[o].push(e)||{i:n,d:t,f:i,r:r,s:f});return}s=typeof i=="function";y=v&&a&&s&&!f;y?et(n,i,l,r):ft(s,n,i,l,r)}f)
```

Y para terminar de cerrar la investigación he probado a guardar la página web msn.es con el comando `curl -o msn.html msn.es`, obteniendo como era previsible la siguiente página:

Object Moved

This document may be found [here](#)

Así que aplicando lo aprendido con el parámetro -L ejecuté: `curl -o msn.html msn.es -L`, obteniendo ahora sí la página web que veríamos en el navegador. De hecho, no hay más que fijarse en la consola para comprobar como con la opción -L se ejecutan dos peticiones en vez de una:

```
sergio@sergio:~$ curl -o msn.html msn.es
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left   Speed
100  147  100  147    0     0   567      0 --:--:-- --:--:-- --:--:--   567
sergio@sergio:~$ curl -o msn.html msn.es -L
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left   Speed
100  147  100  147    0     0   565      0 --:--:-- --:--:-- --:--:--   565
100 44262  100 44262    0     0 96431      0 --:--:-- --:--:-- --:--:-- 96431
sergio@sergio:~$
```