
SISTEMAS WEB

CURSO 2022/2023

HTTP - HyperText Transfer Protocol

Compresión



Web Sistemak by [Oskar Casquero](#) & [María Luz Álvarez](#) is licensed under a [Creative Commons Reconocimiento 4.0 Internacional License](#).

FUNCIONAMIENTO DE HTTP:

COMPRESIÓN

- ¿Cómo se reduce el tamaño de las respuestas?

- En general, el contenido de una respuesta HTTP es texto: HTML, XML, JSON, CSS.
- El texto tiene buena respuesta a la compresión.
- El protocolo HTTP permite comprimir el cuerpo del mensaje, siendo el algoritmo más utilizado para ello el *gzip*.
- El algoritmo de compresión se indica en dos cabeceras:
 - En la **solicitud**: “**Accept-Encoding**”
 - Si el cliente no tiene capacidad para descomprimir un contenido comprimido, este encabezamiento tomará el valor de "Identity".
 - En la **respuesta**: “**Content-Encoding**”
 - Si el cuerpo del mensaje se envía sin comprimir, esta cabecera no se debe introducir en la respuesta.

FUNCIONAMIENTO DE HTTP:

COMPRESIÓN

- Supongamos que un cliente, capaz de trabajar con contenidos comprimidos, solicita un recurso con URI : <http://www.google.es/>


Solicitud de ejemplo

```
GET / HTTP/1.1
Host: www.google.es
Accept: text/html
Accept-Encoding: gzip,identity;q=0.5
Accept-Language: en-US,es-ES;q=0.8
User-Agent: Mozilla Windows Escritorio
```

Respuesta de ejemplo

```
HTTP/1.1 200 OK
Date: Mon, 30 Nov 2015 21:03:14 GMT
Content-Encoding: gzip
Content-Length: 5543
Content-Type: text/html; charset=ISO-8859-1
```

CONTENIDO COMPRIMIDO



```
00Ú|ñ^â¥ÊîF|@eL^N6|æYA|HÀ?c'S|Ö■Ü_Q1|EF±E|_#'^-|
|e%0E|L■B

8Ó_T0u$°L^8UâS y BX^|B^~i5]?Âû":ã%Ç_TôT@M-:ÓA(=ô!|Â^X#çLd||8±`
^@70Ç ä      ^:ád0Tê'LH?NÇx2VâÉÉi4B°Ç-âMÇæ^Ai2É-üy0a:8îâÇAI
(â^)^æDr^~P!±ú=E·âc^TqûÐÉ_TÇî||c 4$F|â||ô:h'1Séü'=ÇÖbâtiüüýüB
```

FUNCIONAMIENTO DE HTTP:

COMPRESIÓN - EJEMPLO

```
import requests
import sys
import zlib

metodo = 'GET'
uri = "https://www.google.es/"
cabeceras = {'Host': 'www.google.es'}

compressed = False
if len(sys.argv) == 1:
    cabeceras['Accept-Encoding'] = 'identity'
elif sys.argv[1] == 'compress':
    compressed = True
    cabeceras['Accept-Encoding'] = 'gzip'
else:
    print("Error! Erabilera: python compression_es.py compress")
    exit(0)

respuesta = requests.request(metodo, uri, headers=cabeceras, allow_redirects=False, stream=True)

codigo = respuesta.status_code
descripcion = respuesta.reason
print(str(codigo) + " " + descripcion)
for cabecera in respuesta.headers:
    print(cabecera + ": " + respuesta.headers[cabecera])

print("RESPONSE CONTENT LENGTH: " + str(len(respuesta.raw.data)) + " byte")
if compressed:
    contenido_compressed = respuesta.raw.data
    contenido_uncompressed = zlib.decompress(contenido_compressed, 16+zlib.MAX_WBITS)
    print("UNCOMPRESSED RESPONSE CONTENT LENGTH: " + str(len(contenido_uncompressed)) + " byte")
```

```
graph TD
    A[stream=True] --> B[raw.data]
    B --> C[len(respuesta.raw.data)]
```

- La librería requests descomprime automáticamente los formatos de compresión gzip y deflate.
- Si queremos un acceso directo al contenido de la respuesta, introducimos el parámetro `stream = True` en la solicitud y leemos el campo `raw.data` en la respuesta

FUNCIONAMIENTO DE HTTP:

COMPRESIÓN - EJEMPLO

```
(venv) C:\Users\cvzcaio\Dropbox (Personal)\docencia\grado\Sistemas Web\SW 2020\Praktikal>python compression.py compress
200 OK
Date: Wed, 12 Feb 2020 07:39:19 GMT
Expires: -1
Cache-Control: private, max-age=0
Content-Type: text/html; charset=ISO-8859-1
P3P: CP="This is not a P3P policy! See g.co/p3phelp for more info."
Content-Encoding: gzip
Server: gws
X-XSS-Protection: 0
X-Frame-Options: SAMEORIGIN
Set-Cookie: 1P_JAR=2020-02-12-07; expires=Fri, 13-Mar-2020 07:39:19 GMT; path=/; domain=.google.es; Secure, NID=197=TuCqh
sRxyTUWCZOVawM6ScZObD0; expires=Thu, 13-Aug-2020 07:39:19 GMT; path=/; domain=.google.es; HttpOnly
Alt-Svc: quic=":443"; ma=2592000; v="46,43",h3-Q050=":443"; ma=2592000,h3-Q049=":443"; ma=2592000,h3-Q048=":443"; ma=2592000
Transfer-Encoding: chunked
RESPONSE CONTENT LENGTH: 5584 byte
UNCOMPRESSED RESPONSE CONTENT LENGTH: 13116 byte
```

- Tamaño de la respuesta comprimida: 5584 octetos
 - Tamaño de la respuesta sin comprimir: 13116 octetos
- Un 57% de ahorro en el tamaño del cuerpo del mensaje