

WEB SISTEMAK

2020/2021 IKASTURTEA

HTTP - HyperText Transfer Protocol

Konpresioa eta Cache-a

2021-02-12 M (1. zatia)



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

HTTP-REN FUNTZIONAMENDUA: KONPRESIOA

- **Zelan laburtzen da erantzunen tamaina?**

- Orokorrean, HTTP erantzun baten edukia testua da: HTML, XML, JSON, CSS.
- Testuak konpresioarako erantzun ona aurkezten du.
- HTTP protokoloak mezuaren gorputza konprimatzeko aukera ematen du, horretarako algoritmorik erabiliena *gzip* delarik.
- Konpresio algoritmoa bi goibururen bitartez adierazten da:
 - Eskaeran: “Accept-Encoding”
 - Bezeroak konprimatutako eduki bat deskonprimatzeko gaitasunik ez badu, goiburu honek “identity” balioa hartuko du.
 - Erantzunean: “Content-Encoding”
 - Mezuaren gorputza konprimatu gabe bidaltzen bada, ez da goiburu hau adierazten.

HTTP-REN FUNTZIONAMENDUA: KONPRESIOA

- Demagun eduki konprimatuekin lan egiteko gai den bezero batek ondorengo URI-a duen baliabidea eskatzen duela: <https://www.google.es/>


Eskaera baten adibidea

```
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 Desktop
```

Erantzun baten adibidea

```
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
```

KONPRIMATUTAKO EDUKIA



```
8Ó|ñ"â¥ÊîF@eL"Ñ6|æYÁ|HÀ?c'S|Ö■_Q_1■EF±E|_#'^|
|æ%ôE||■B
8Ó_T0üs°L18Uâs y ÆX"|-B*-i5]²Äü":â%Ç_Tô¶@*M-:óA(=ô!|Ä²X#çL_d||ð±'
²@70Ç ä      ²²ád0Tê'■H²NÇx2VâÊ£i4B°Ç-âMÇæ²Ai2É-üy0a:8îâÇAÎ
(¬â²)²æDr*-B!±ú=E·äc²¶üÐÉ_TÇi||c 4$F|â||ô:h'1Séu²=ÇÖbätüüYüB||
```

ADIBIDEA: KONPRESIOA

```
import requests
import sys
import zlib

metodoa = 'GET'
uria = "https://www.google.es/"
goiburuak = {'Host': 'www.google.es'}

compressed = False
if len(sys.argv) == 1:
    goiburuak['Accept-Encoding'] = 'identity'
elif sys.argv[1] == 'compress':
    compressed = True
    goiburuak['Accept-Encoding'] = 'gzip'

erantzuna = requests.request(metodoa, uria, headers=goiburuak, allow_redirects=False, stream=True)

kodea = erantzuna.status_code
deskribapena = erantzuna.reason
print(str(kodea) + " " + deskribapena)
for goiburua in erantzuna.headers:
    print(goiburua + ": " + erantzuna.headers[goiburua])

print("RESPONSE CONTENT LENGTH: " + str(len(erantzuna.raw.data)) + " byte")
if compressed:
    edukia compressed = erantzuna.raw.data
    edukia uncompressed = zlib.decompress(edukia compressed, 16+zlib.MAX_WBITS)
    print("UNCOMPRESSED RESPONSE CONTENT LENGTH: " + str(len(edukia_uncompressed)) + " byte")
```

The diagram consists of two red arrows. The first arrow starts from the `stream=True` parameter in the `requests.request` call and points to a red box containing `erantzuna.raw.data`. The second arrow starts from this box and points to the `edukia compressed` variable in the decompression logic.

- requests liburutegiak *gzip* eta *deflate* konpresio formatuak automatikoki deskonprimatzen ditu.
- Erantzunaren edukira atzipen zuzena nahi badugu, eskaeran *stream=True* parametroa sartu eta erantzunean *raw.data* eremua irakurri behar dugu

ADIBIDEA: KONPRESIOA

```
(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
```

- Konprimatutako edukiaren luzeera: 5584 zortzikote
- Konprimatutako gabeko edukiaren luzeera: 13116 zortzikote
%57-ko aurrezpena mezuaren gorputzaren tamainan

HTTP-REN FUNTZIONAMENDUA: CACHE

- **Zelan ekiditzen da beharrezkoa ez den trafiko trukea?**

- HTTP protokoloak baditu bezeroak azkenegoz eskatu zituenetik aldatu ez diren baliabideen transferentzia saihesteko moduak:
 - Status: 304
 - Goiburuak:
 - Eskameran: Cache-Control, If-Modified-Since, If-None-Match
 - Erantzunean: Cache-Control, Date, Last-Modified, ETag
- Horrela, banda-zabaleraren erabilera optimizatu eta web zerbitzariak eskaera eta erantzun erredundanteak prozesatu eta sor ditzan saihesten da.
- Cache-aren inguruan, hiru dira kontuan hartu beharreko alderdiak:
 - Freskotasuna
 - Balidazioa bezeroan
 - Balidazioa zerbitzarian

HTTP-REN FUNTZIONAMENDUA: CACHE

- Demagun nabigatzaile batek web orri bat eskatzen duela. Erantzuna (HTML dokumentua) jasotzen duenean, nabigatzaileak web orriaren barnean irudi bati erreferentzia egiten dion **** elementua dagoela ikusten du.
- Beraz, nabigatzaileak irudia eskuratzeko bigarren HTTP eskaera bat egiten du. Web zerbitzariak irudia HTTP erantzun batean itzultzen du, erantzun horretako goiburu jakin batzuetan meta-datu batzuk sartzen dituelarik.

Eskaera baten adibidea

```
GET /image.jpg HTTP/1.1
Host: sw2016.com:8080
Accept: image/*
User-Agent: Mozilla Windows Escritorio
```

Erantzun baten adibidea

```
HTTP/1.1 200 OK
Date: Thu, 20 Nov 2015 20:25:52 GMT
Last-Modified: Tue, 17 Sep 2015 13:00:02 GMT
ETag: "1a968-3ec-4e693e61bb8b6"
Cache-Control: max-age=2592000
Content-Length: 12405
Content-Type: image/jpg

IRUDIA (eduki binarioa)
```

HTTP-REN FUNTZIONAMENDUA: CACHE:

FRESKOTASUNA

- Demagun erabiltzaileak nabigatzailea itxi eta hurrengo egunean web orri bera eskatzeko nabigatzailea berriro zabaltzen duela.

Aurreko adibideko erantzuna

```
HTTP/1.1 200 OK
Date: Thu, 20 Nov 2015 20:25:52 GMT
Last-Modified: Tue, 17 Sep 2015 13:00:02 GMT
ETag: "1a968-3ec-4e693e61bb8b6"
Cache-Control: max-age=2592000
Content-Length: 12405
Content-Type: image/jpg

IRUDIA (eduki binarioa)
```

“Cache-Control” goiburua ([RFC 2616, 14.9 atala](#)) bezero eta proxy-en cache-en alderdi ezberdinak kudeatzeko erabiltzen da.

Kasu honetan, zerbitzariak nabigatzaileari ondorengoa dinotso: “hilabete batez, ez iezadazu irudi honengaitik galdetu”.

Horrela, nabigatzaileak irudia bere cache-tik zuzenean hartzen du.

HTTP-REN FUNTZIONAMENDUA: CACHE:

BALIDAZIOA BEZEROAN

- Demagun bi hilabete geroago erabiltzaileak web orri bera berriro eskatzen duela, baina, nahiz eta web orriaren edukia aldatu den, irudia ez da aldatu.
- Nabigatzaileak, irudiari lotutako cache gordetze denbora iraungi dela ikustean, eskaera berri bat egiten du:

Aurreko adibideko erantzuna

```
HTTP/1.1 200 OK
Date: Thu, 20 Nov 2015 20:25:52 GMT
Last-Modified: Tue, 17 Sep 2015 13:00:02 GMT
ETag: "1a968-3ec-4e693e61bb8b6"
Cache-Control: max-age=2592000
Content-Length: 12405
Content-Type: image/jpg

IRUDIA (eduki binarioa)
```

Eskaera berriaren adibidea

```
GET /image.jpg HTTP/1.1
Host: sw2016.com:8080
Accept: image/*
If-Modified-Since: Tue, 17 Sep 2015 13:00:02 GMT
If-None-Match: "1a968-3ec-4e693e61bb8b6"
User-Agent: Mozilla Windows Desktop
```

“If-Modified-Since” goiburuan aurreko erantzunean “Last-Modified” goiburuan itzulitako balioa sartzen da.

“If-None-Match” goiburuan aurreko erantzunean “ETag” goiburuan itzulitako balioa sartzen da.

HTTP-REN FUNTZIONAMENDUA: CACHE:

BALIDAZIOA ZERBITZARIAN

- *If-Modified-Since* datatik hona irudia aldatu ez bada eta *If-None-Match*-en adierazitako kodea zerbitzariak irudiari esleituta dion *Etag* kodearekin bat badator, orduan HTTP erantzunean *304 Not Modified* erantzun kodea itzultzen da eta ez da edukirik txertatzen.

Erantzun baten adibidea

```
HTTP/1.1 304 Not Modified
Date: Thu, 20 Jan 2016 20:25:52 GMT
Last-Modified: Tue, 17 Sep 2015 13:00:02 GMT
ETag: "1a968-3ec-4e693e61bb8b6"
Cache-Control: max-age=2592000
```

HTTP-REN FUNTZIONAMENDUA: CACHE: ARIKETA

Zerbitzari batek baliabide bat eskuragarri du ondorengo URI-an: <http://localhost:8080/SW2016/servlet/HolaMundoCache>

OHARRA: ondoren azaltzen diren kasuetan ez da baliabidea aldatzen.

1. KASUA: Aldiune jakin batean, nabigatzaile batek baliabidea eskatzen du; bezeroak eta zerbitzariak ondorengo HTTP trafikoa elkar aldatzen dute:

```
GET /SW2016/servlet/HolaMundoCache HTTP/1.1
Host: localhost:8080
User-Agent: Mozilla/5.0 (Windows NT 6.1; WOW64; rv:44.0) Gecko/20100101 Firefox/44.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: es-ES,es;q=0.8,en-US;q=0.5,en;q=0.3
Accept-Encoding: gzip, deflate
Connection: keep-alive
```

```
HTTP/1.1 200 OK
Server: Apache-Coyote/1.1
Date: Tue, 17 May 2016 19:19:05 GMT
Last-Modified: Tue, 17 May 2016 19:19:05 GMT
Cache-Control: max-age=3600
Content-Type: text/html; charset=utf-8
Content-Length: 475

<!DOCTYPE html><html>...</html>
```

2. KASUA: 61 minutu geroago nabigatzaileak bigarren eskaera bat egiten du; bezeroak eta zerbitzariak ondorengo HTTP trafikoa elkar trukutzen dute:

```
GET /SW2016/servlet/HolaMundoCache HTTP/1.1
Host: localhost:8080
User-Agent: Mozilla/5.0 (Windows NT 6.1; WOW64; rv:44.0) Gecko/20100101 Firefox/44.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: es-ES,es;q=0.8,en-US;q=0.5,en;q=0.3
Accept-Encoding: gzip, deflate
Connection: keep-alive
If-Modified-Since: Tue, 17 May 2016 19:19:05 GMT
```

```
HTTP/1.1 304 Not Modified
Server: Apache-Coyote/1.1
Date: Tue, 17 May 2016 20:20:05 GMT
Last-Modified: Tue, 17 May 2016 19:19:05 GMT
Cache-Control: max-age=3600
```

Nabigatzaileak eskaera bera egiten badu, zergaitik jasotzen du erantzun ezberdina?

Erantzuna: **Zerbitzarian baliabidea aldatu ez delako (HTTP/1.1 304 Not Modified). Beraz, nabigatzailearen cache-an gordetako baliabideak baliozkoa izaten jarraitzen du eta ez da berriro bidali behar (erantzunak ez du edukirik edo gorputzik).**