

Introducción al programa de capacitación para crawlers.

Contenido del primer módulo:

- Conociendo la store info.
- introducción al manejo de cookies.
- Introducción a custom headers
- introducción a webs multdivisas o cambios de monedas.

1.1 Store info.

¿Qué es la store info?

La store info es un array que contiene la información referente a las stores que manejamos en el rubik, es un array indexado que contiene en sus posiciones el tipo de currency o moneda que maneja la tienda , el código de país o country code el lenguaje de la tienda entre otros.

¿Cómo acceder a la store info?

Para acceder a la store info basta con hacer dump a la variable \$store_info dentro del script editor en el rubik.

```
3 dump($store_info);  
4
```

Esto nos devolverá un arreglo con la siguiente información.

```
(pre>array(67) {  
    ["StoreID"]=>  
    string(5) "86963"  
    ["Domain"]=>  
    string(18) "basket4ballers.com"  
    ["CustomerID"]=>  
    string(5) "13857"  
    ["CountryCode"]=>  
    string(2) "fr"  
    ["Currency"]=>  
    string(3) "eur"  
    ["DefaultLang"]=>  
    string(2) "fr"  
    ["Sufix"]=>  
    string(2) "71"  
    ["StoreIDLogical"]=>  
    string(1) "0"  
    ["Sleep"]=>  
    string(7) "3000000"  
    ["InheritProxyType"]=>  
    string(1) "0"  
    ["ProxyType"]=>  
    string(3) "all"  
    ["ImagesProxyType"]=>  
    string(0) ""  
}
```

Para utilizar esta información accedemos como en cualquier arreglo accediendo a su posición.

Ejemplo:

Estamos trabajando una web que tiene varias herencias y debemos construir una URL para luego trabajarla.

<https://www.basket4ballers.com/?currency=eur>

Si tenemos ese mismo dominio .de y .com por ende necesitamos construir la URL para que funcione para todas las herencias entonces haciendo uso de la store info en su posición Domain y su posición Currency nos permitirá construir la url de la siguiente manera.

```
2  
3 dump($store_info['Domain']);  
4
```

esto nos devolvera:

```
<pre>string(18) "basket4ballers.com"
```

```
dump($store_info['Currency']);
```

esto nos devolvera:

```
<pre>string(3) "eur"
```

Quedaría de la siguiente forma:

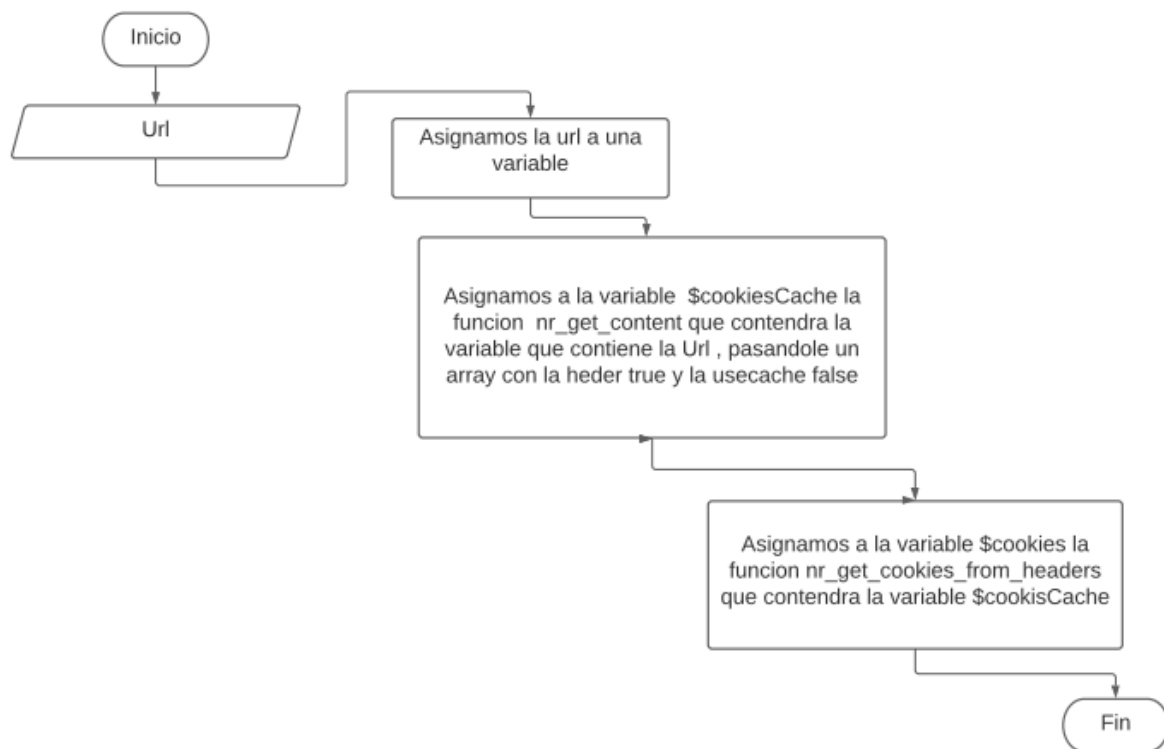
```
$_url = 'https://www.'.$store_info['Domain'].'/?currency='.$store_info['currency'];
```

De esta manera podemos usar de manera dinámica la información de la store para manipularla según nuestro requerimiento.

2.2 Introducción al manejo de cookies.

Las cookies son identificadores de sesión que se utilizan para indicar que se está trabajando con algo en específico dentro de un sitio web, bien sea un producto o algún atributo selector de cambio de moneda o geolocalización.

Para la captura de las cookies de cualquier URL que estemos manipulando debemos seguir los pasos señalados en el diagrama de flujo.



\$url =

'https://www.basket4ballers.com/fr/chaussure-de-basket/21289-adidas-pro-bounce-2019-low-noiess-ftwbla-ecarla-eg2818.html';

\$cacheCookies = nr_get_content(\$url, array ('header'=>true , 'usecache'=>false));

\$cookies = nr_get_cookies_from_headers(\$cacheCookies);

esto nos devolvera:

```

<pre>array(6) {
  [0]=>
  string(37) "SP5I=f9272097f6c473f19302598175317d21"
  [1]=>
  string(93) "SP5E=Er21cQ7TU1/BfUxrZ6MvDS84d0WaLohgeCF3vvwwfuYtsag1n1eJqVHUxqpbs2nxLRY6OK11k1LV4VhNUM3Awg=="
  [2]=>
  string(39) "spcsrf=a6b36bca4ee45d39f448ce5930905ea3"
  [3]=>
  string(11) "adOtr=obsv1"
  [4]=>
  string(44) "UTGv2=D-h4f827401b3dd38d200cb6bce2e99047a879"
  [5]=>
  string(31) "sp_1it=uuhyQyvyS1YYunkMTRjCg=="
}
```

Una vez obtenidas las cookies las enviamos mediante un array dentro del nr_get_content que estamos consultando de la siguiente manera :

```

1 <?php
2 function details($url){
3     global $import;
4
5     $import->url($url);
6
7     $cacheCookies = nr_get_content($url, array ('header'=>true , 'usecache'=>false));
8     $cookies = nr_get_cookies_from_headers($cacheCookies);
9
10    $cache = nr_get_content($url, array('cookies'=>$cookies));
11
12

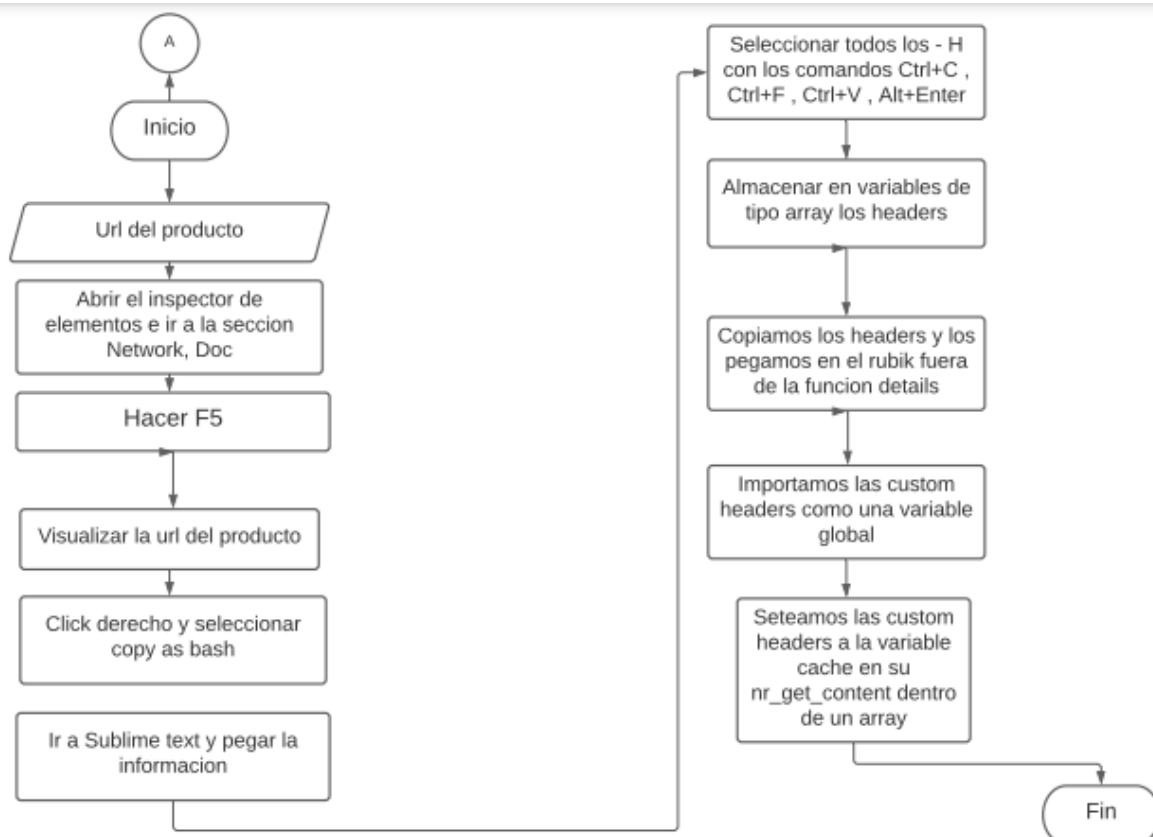
```

`$cache = nr_get_content($url, array('cookies'=>$cookies));`

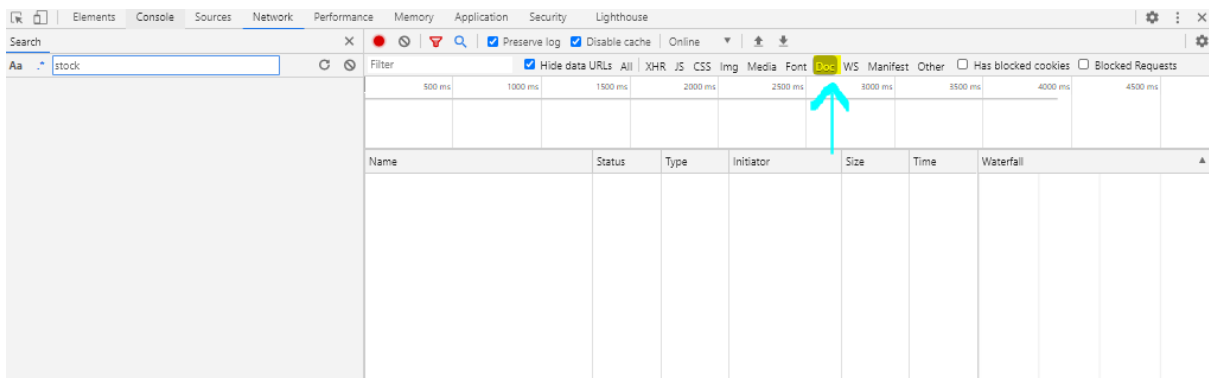
3.3 Introducción a custom headers

Los custom headers son las cabeceras usadas por las url las cuales contienen información como el tipo de archivo que envía o maneja, cookies, parámetros post entre otros.

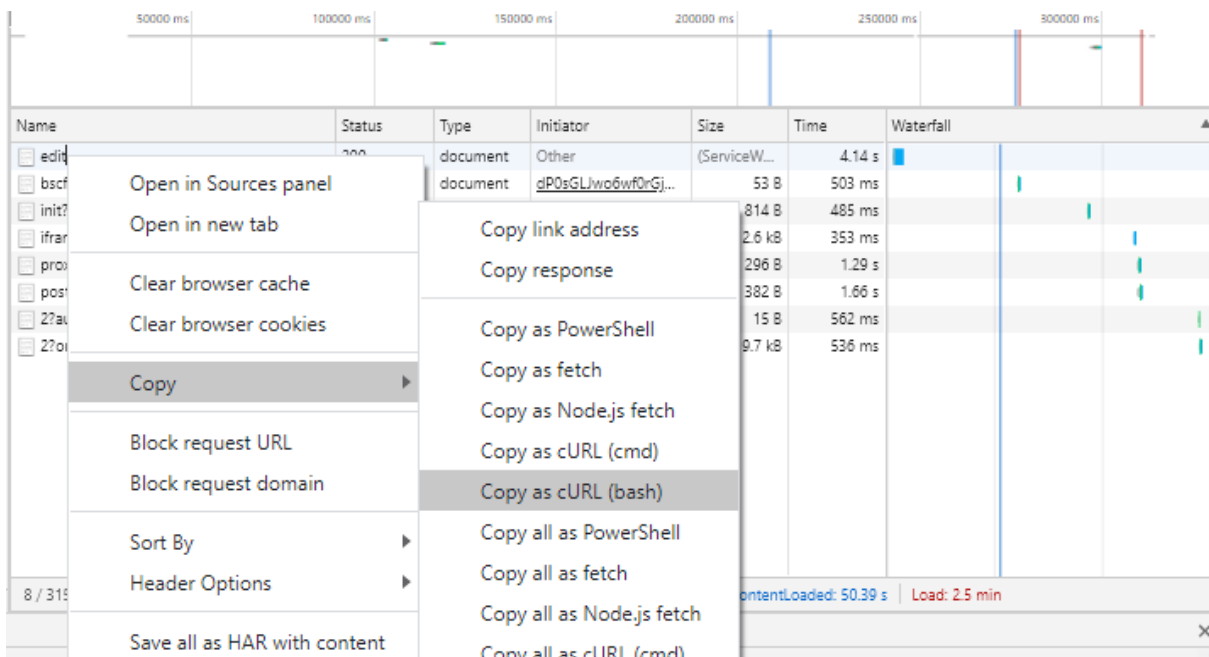
Para acceder a esta información debemos seguir los pasos señalados en el diagrama de flujo.



Paso 2



paso 5



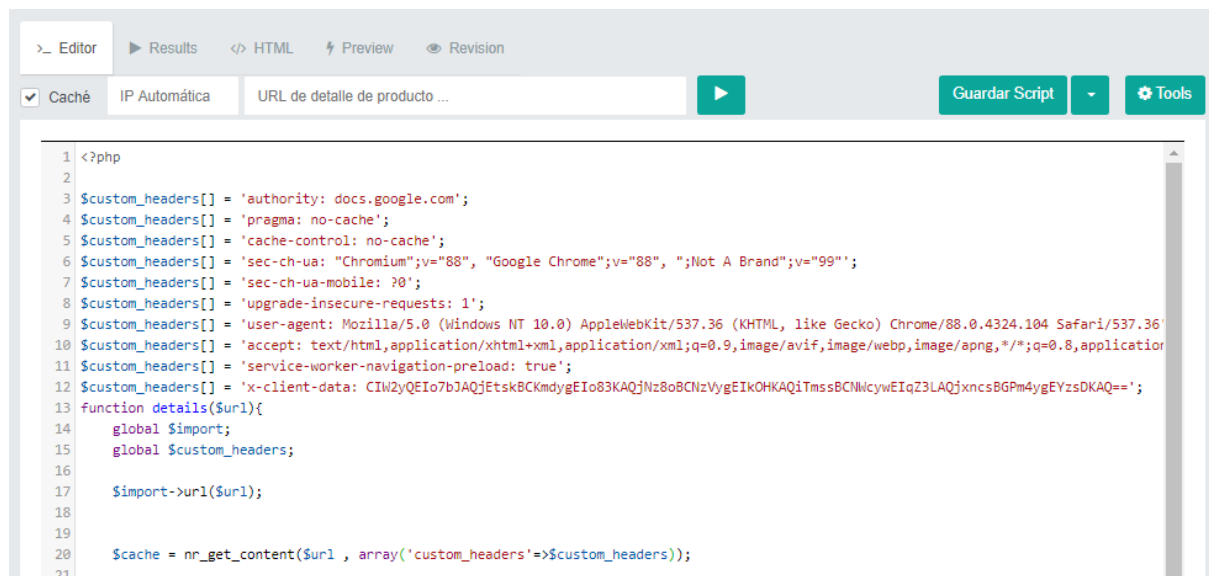
paso 6

```
-H 'authority: docs.google.com' \  
-H 'pragma: no-cache' \  
-H 'cache-control: no-cache' \  
-H 'sec-ch-ua: "Chromium";v="88", "Google  
Chrome";v="88", ";Not A Brand";v="99"' \  
-H 'sec-ch-ua-mobile: ?0' \  
-H 'upgrade-insecure-requests: 1' \  
-H 'user-agent: Mozilla/5.0 (Windows NT 10.0)  
AppleWebKit/537.36 (KHTML, like Gecko)  
Chrome/88.0.4324.104 Safari/537.36' \  
-H 'accept: text/html,application/xhtml+xml,application/  
xml;q=0.9,image/avif,image/webp,image/  
apng,/*/*;q=0.8,application/signed-exchange;v=b3;q=0.9' \  
-H 'service-worker-navigation-preload: true' \  
-H 'x-client-data: CIW2yQEIo7bJAQjEtskBCKmdygEIo83KAQjNz  
8oBCNzVygEIkOHKAQiTmssBCNWcywEIqZ3LAQjxncsBGPm4ygEYzsDKA
```

paso 8

```
$custom_headers[] = 'authority: docs.google.com';  
$custom_headers[] = 'pragma: no-cache';  
$custom_headers[] = 'cache-control: no-cache';  
$custom_headers[] = 'sec-ch-ua: "Chromium";v="88",  
"Google Chrome";v="88", ";Not A Brand";v="99";  
$custom_headers[] = 'sec-ch-ua-mobile: ?0';  
$custom_headers[] = 'upgrade-insecure-requests: 1';  
$custom_headers[] = 'user-agent: Mozilla/5.0 (Windows  
NT 10.0) AppleWebKit/537.36 (KHTML, like Gecko)  
Chrome/88.0.4324.104 Safari/537.36';  
$custom_headers[] = 'accept: text/html,application/  
xhtml+xml,application/xml;q=0.9,image/avif,image/  
webp,image/apng,/*/*;q=0.8,application/  
signed-exchange;v=b3;q=0.9';  
$custom_headers[] = 'service-worker-navigation-preload:  
true';  
$custom_headers[] = 'x-client-data: CIW2yQEIo7bJAQjEtskB  
CKmdygEIo83KAQjNz8oBCNzVygEIkOHKAQiTmssBCNWcywEIqZ3LAQjx  
ncsBGPm4ygEYzsDKAQ==';
```

pasos 9 , 10 y 11



The screenshot shows a web browser interface with a developer tools panel open. The 'Console' tab is active, displaying a series of PHP code snippets. The code is as follows:

```
1 <?php
2
3 $custom_headers[] = 'authority: docs.google.com';
4 $custom_headers[] = 'pragma: no-cache';
5 $custom_headers[] = 'cache-control: no-cache';
6 $custom_headers[] = 'sec-ch-ua: "Chromium";v="88", "Google Chrome";v="88", ";Not A Brand";v="99";';
7 $custom_headers[] = 'sec-ch-ua-mobile: ?0';
8 $custom_headers[] = 'upgrade-insecure-requests: 1';
9 $custom_headers[] = 'user-agent: Mozilla/5.0 (Windows NT 10.0) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/88.0.4324.104 Safari/537.36';
10 $custom_headers[] = 'accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9';
11 $custom_headers[] = 'service-worker-navigation-preload: true';
12 $custom_headers[] = 'x-client-data: CIW2yQEIo7bJAQjEtskBCkmdyEIo83KAQjNz8oBCNzVygEIkOHKAQITmsBCNwcywEIqZ3LAQjxncsBGPm4ygEYzsDKAQ=';
13 function details($url){
14     global $import;
15     global $custom_headers;
16
17     $import->url($url);
18
19
20     $cache = nr_get_content($url , array('custom_headers'=>$custom_headers));
21 }
```

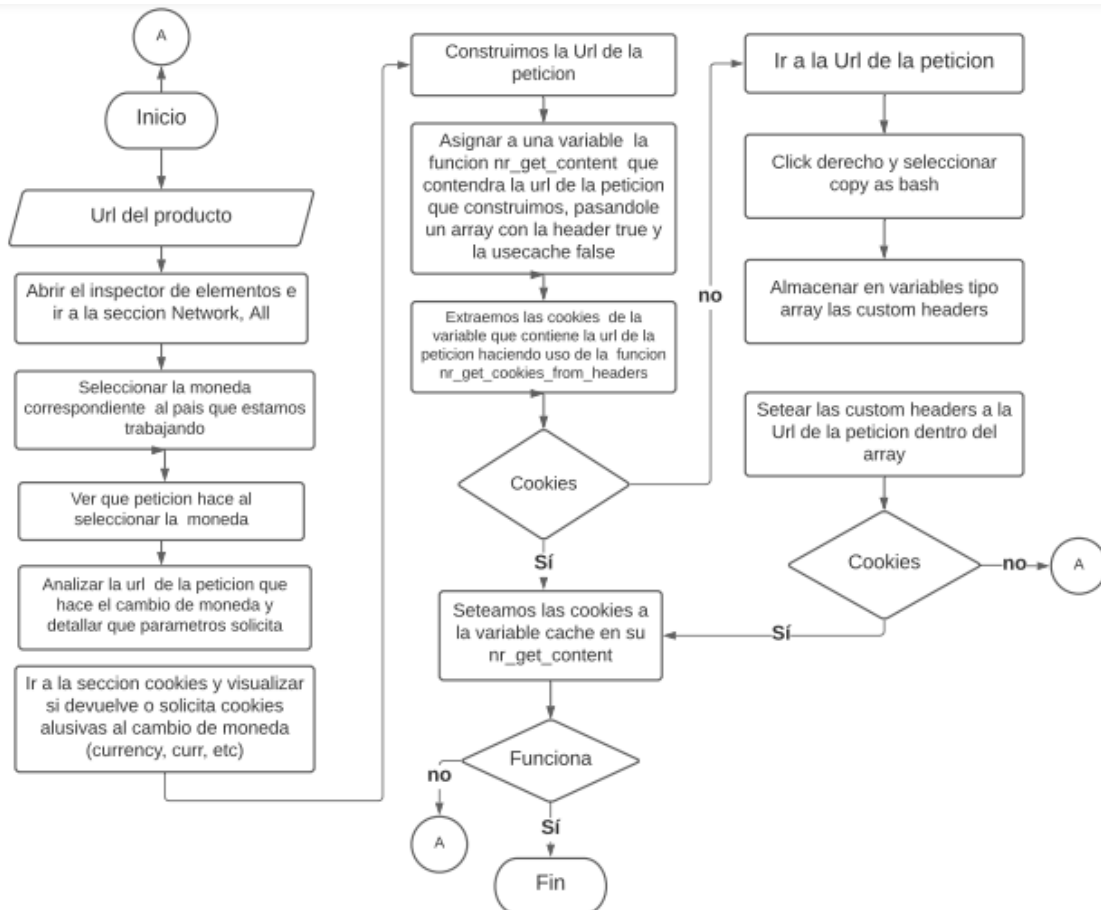
De esta manera seteamos las custom headers.

```
$cache = nr_get_content($url , array('custom_headers'=>$custom_headers));
```

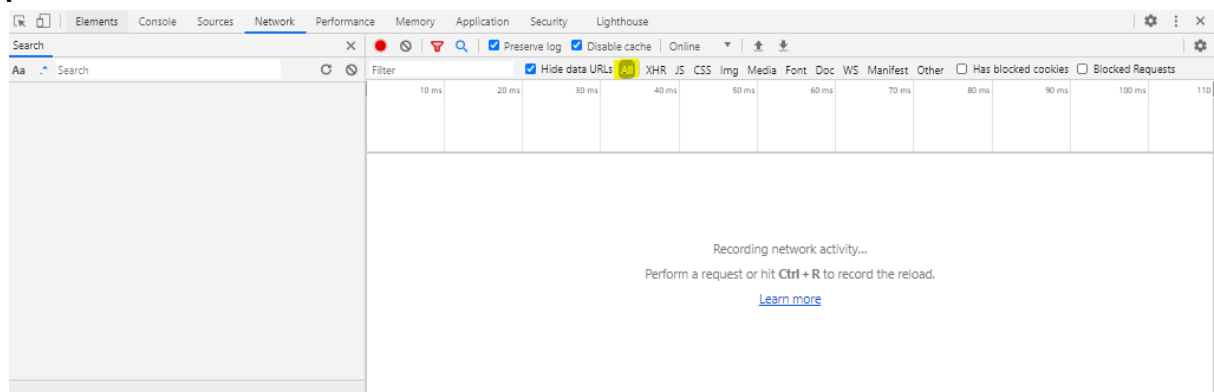
3.3 Webs multidivisas.

En algunas webs nos conseguiremos que la store vende sus productos para varios países los cuales manejan divisas distintas como es el caso de <https://varuste.net/> esta web ofrece sus productos no solo a la unión europea sino al reino unido y estados unidos por lo que hace cambio de moneda según el país de destino.

Para estos casos los pasos a seguir para setear la moneda correcta es la señalada en el diagrama de flujo.



paso 2



paso 3

1626.51 SEK	Выбрать
258.68 SGD	Выбрать
5827.83 THB	Выбрать
1433.30 TRY	Выбрать
2958.53 ZAR	Выбрать
142.19 GBP £	Выбрать
20253 JPY ¥	Выбрать
160.48 EUR €	Выбрать
14825.82 RUB ₹	Выбрать

paso 4

Name	Headers	Preview	Response	Initiator	Timing	Cookies
vaihda_valuuttakurssi.php	<p>General</p> <p>Request URL: https://varuste.net/ru/ajax/vaihda_valuuttakurssi.php</p> <p>Request Method: POST</p> <p>Status Code: 200</p> <p>Remote Address: 81.22.242.48:443</p> <p>Referrer Policy: strict-origin-when-cross-origin</p> <p>Response Headers</p> <p>access-control-allow-origin: *</p> <p>content-encoding: gzip</p> <p>content-type: text/html; charset=utf-8</p> <p>date: Sun, 31 Jan 2021 00:24:06 GMT</p> <p>p3p: CP="x"</p> <p>server: nginx</p> <p>set-cookie: maaid=VE; expires=Mon, 31-Jan-2022 00:24:06 GMT; Max-Age=31536000; path=/; SameSite=Lax</p> <p>set-cookie: valuutta=EUR; expires=Mon, 31-Jan-2022 00:24:06 GMT; Max-Age=31536000; path=/; SameSite=Lax</p> <p>vary: Accept-Encoding</p> <p>x-shuriken-jsid: "250"</p> <p>x-shuriken-title: "From the depths to the summits Varuste.net \u0420\u0443\u0441\u0441\u0438\u0439"</p>					

paso 5

Form Data	view parsed
change_currency=EUR	

paso 6

Headers	Preview	Response	Initiator	Timing	Cookies																																																																																																														
<p>Request Cookies <input type="checkbox"/> show filtered out request cookies</p> <table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> <th>D</th> <th>P</th> <th>E</th> <th>S</th> <th>H</th> <th>Se...</th> <th>Sa...</th> <th>Pri...</th> </tr> </thead> <tbody> <tr> <td>shurikenscript</td> <td>250</td> <td>v</td> <td>/</td> <td>/</td> <td>3</td> <td>1</td> <td></td> <td>Lax</td> <td>M...</td> </tr> <tr> <td>viimeksi_katsotut_tuott...</td> <td>98528</td> <td>v</td> <td>/</td> <td>/</td> <td>2</td> <td>3</td> <td></td> <td>Lax</td> <td>M...</td> </tr> <tr> <td>maaaid</td> <td>VE</td> <td>v</td> <td>/</td> <td>/</td> <td>2</td> <td>7</td> <td></td> <td>Lax</td> <td>M...</td> </tr> <tr> <td>_ga</td> <td>GA1.2.1941062129.1612052261</td> <td>...</td> <td>/</td> <td>/</td> <td>2</td> <td>3</td> <td></td> <td></td> <td>M...</td> </tr> <tr> <td>_gid</td> <td>GA1.2.497609402.1612052261</td> <td>...</td> <td>/</td> <td>/</td> <td>2</td> <td>3</td> <td></td> <td></td> <td>M...</td> </tr> <tr> <td>_fop</td> <td>fb.1.1612052264806.418004975</td> <td>...</td> <td>/</td> <td>/</td> <td>2</td> <td>3</td> <td></td> <td>Lax</td> <td>M...</td> </tr> <tr> <td>valuutta</td> <td>USD</td> <td>v</td> <td>/</td> <td>/</td> <td>2</td> <td>1</td> <td></td> <td>Lax</td> <td>M...</td> </tr> </tbody> </table> <p>Response Cookies</p> <table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> <th>D</th> <th>P</th> <th>E</th> <th>S</th> <th>H</th> <th>Se...</th> <th>Sa...</th> <th>Pri...</th> </tr> </thead> <tbody> <tr> <td>maaaid</td> <td>VE</td> <td>v</td> <td>/</td> <td>/</td> <td>3</td> <td>8</td> <td></td> <td>Lax</td> <td>M...</td> </tr> <tr> <td>valuutta</td> <td>EUR</td> <td>v</td> <td>/</td> <td>/</td> <td>3</td> <td>9</td> <td></td> <td>Lax</td> <td>M...</td> </tr> </tbody> </table>						Name	Value	D	P	E	S	H	Se...	Sa...	Pri...	shurikenscript	250	v	/	/	3	1		Lax	M...	viimeksi_katsotut_tuott...	98528	v	/	/	2	3		Lax	M...	maaaid	VE	v	/	/	2	7		Lax	M...	_ga	GA1.2.1941062129.1612052261	...	/	/	2	3			M...	_gid	GA1.2.497609402.1612052261	...	/	/	2	3			M...	_fop	fb.1.1612052264806.418004975	...	/	/	2	3		Lax	M...	valuutta	USD	v	/	/	2	1		Lax	M...	Name	Value	D	P	E	S	H	Se...	Sa...	Pri...	maaaid	VE	v	/	/	3	8		Lax	M...	valuutta	EUR	v	/	/	3	9		Lax	M...
Name	Value	D	P	E	S	H	Se...	Sa...	Pri...																																																																																																										
shurikenscript	250	v	/	/	3	1		Lax	M...																																																																																																										
viimeksi_katsotut_tuott...	98528	v	/	/	2	3		Lax	M...																																																																																																										
maaaid	VE	v	/	/	2	7		Lax	M...																																																																																																										
_ga	GA1.2.1941062129.1612052261	...	/	/	2	3			M...																																																																																																										
_gid	GA1.2.497609402.1612052261	...	/	/	2	3			M...																																																																																																										
_fop	fb.1.1612052264806.418004975	...	/	/	2	3		Lax	M...																																																																																																										
valuutta	USD	v	/	/	2	1		Lax	M...																																																																																																										
Name	Value	D	P	E	S	H	Se...	Sa...	Pri...																																																																																																										
maaaid	VE	v	/	/	3	8		Lax	M...																																																																																																										
valuutta	EUR	v	/	/	3	9		Lax	M...																																																																																																										

paso 7

```
3 $currency = strtoupper($store_info['Currency']);
4 $domain = $store_info['Domain'];
5 $countryCode = $store_info['CountryCode'];
6 function details($url){
7     global $import;
8     global $currency;
9     global $domain;
10    global $countryCode;
11
12    $url = nr_url_encode($url);
13
14    $import->url($url);
15
16    //CONSTRUIMOS LOS PARAMETROS Y ARMAMOS LA URL
17    $_params = 'change_currency='.$currency;
18    $_url = 'https://'.$domain.'/'.$countryCode.'/ajax/vaihda_valuuttakurssi.php';
19
```

pasos 8 , 9

```
40 // SOLICITAMOS LAS COOKIES
41 $cacheCookies = nr_get_content($_url, array('header'=>true, 'usecache'=>false, 'post'=>$_params));
42 $cookies = nr_get_cookies_from_headers($cacheCookies);
43
```

Esto nos devuelve:

```
<pre>array(3) {
  [0]=>
  string(8) "maaid=FI"
  [1]=>
  string(12) "valuutta=EUR"
  [2]=>
  string(18) "shurikenscript=250"
}
```

paso 10

```
$cache = nr_get_content($_url, array('usecache'=>false, 'cookies'=>$cookies));
```

De esta manera obtenemos exitosamente el cambio de divisas como se puede apreciar en las siguientes imágenes.

Vista del sitio web:

Haglöfs Roc Fusion Pant Men



Цена

160.48 EUR € ▾ + расходы на доставку

Цена с НДС 0%

Количество

1 **WT**

Цвет

Vista del rubik:



Haglöfs Roc Fusion Pant Men

160.48 EUR

Code: 98528 DB Current Code: 98528

Marca: Haglöfs

Categorías: [спортивный](#) » [походы](#) » [одежда](#) » [штаны](#) » [soft shell](#) »

URL:

<https://varuste.net/ru/Hagl%25252525252525C3%252525252525B6fs%252525252525tu=98528>