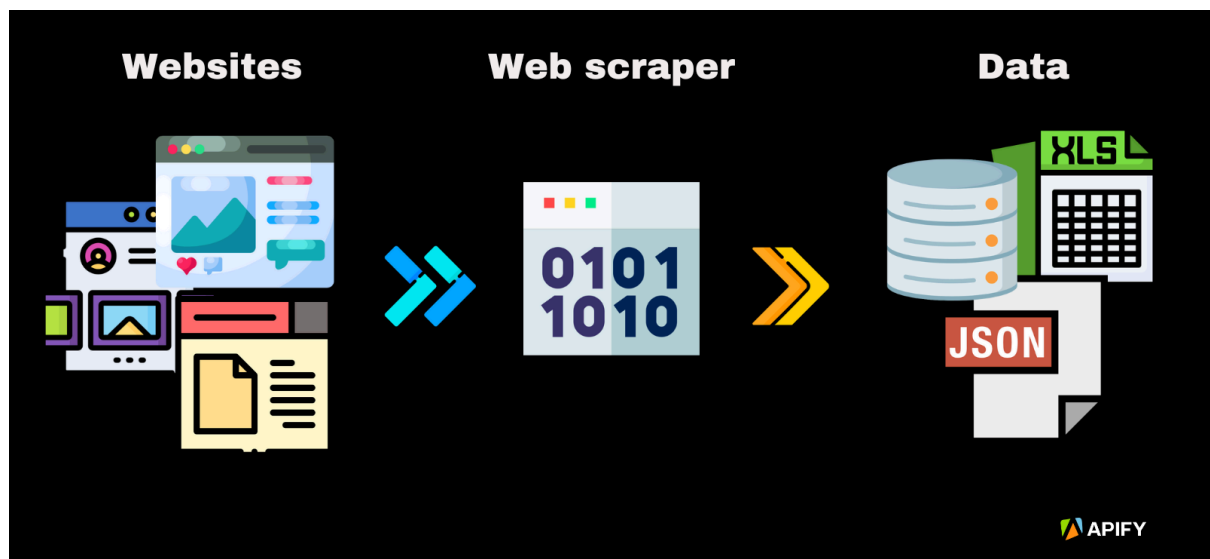


Documentación del web Scrapping



Índice

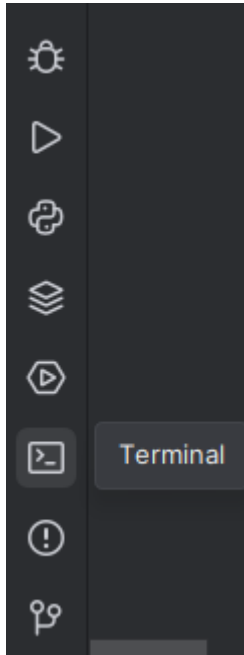
1. Importar bibliotecas necesarias
2. Primera función
3. Segunda función
4. Tercera función

1. Importar bibliotecas necesarias

Lo primero que tendremos que hacer será instalar en nuestro nuevo proyecto las dos bibliotecas que necesitaremos para conseguir hacer el web scraping. Serían estas dos:

```
import requests
from bs4 import BeautifulSoup
```

Para instalarlas necesitaremos acceder al terminal del pycharm la cual está situada abajo a la izquierda y se vería tal que así:



Una vez abierto tendremos que escribir estas dos cosas:

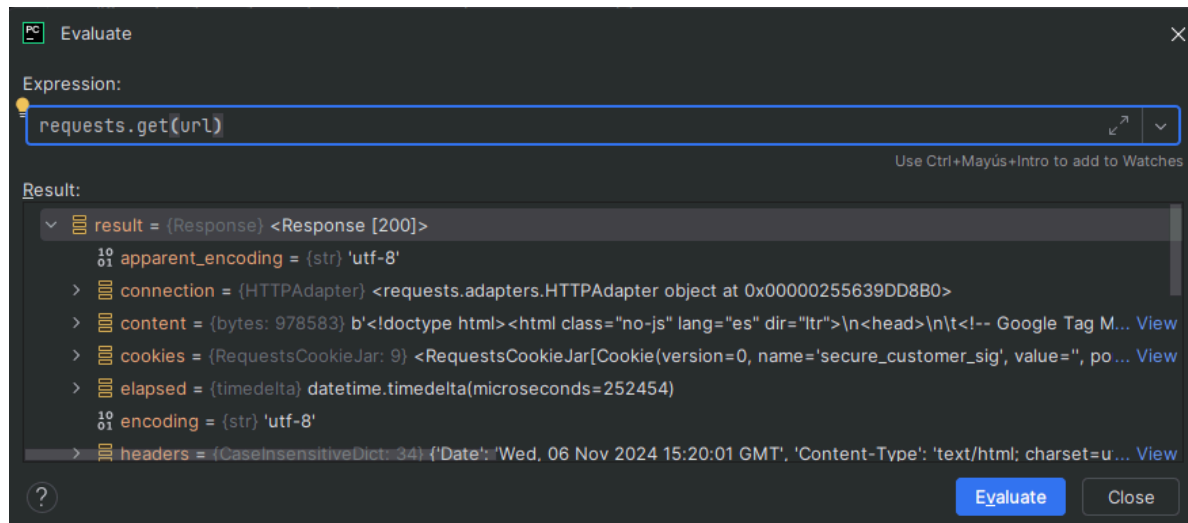
- pip install requests
- pip install beautifulsoup4

2. Primera función

Esta primera función lo que hace es comunicarse con la página web para que nos dé la respuesta que queremos que sea un 200.

Lo primero es guardar el enlace de la página web en una variable.

Después será utilizar en otra variable la primera biblioteca que hemos instalado previamente, primero utilizando `.get()` y dentro del paréntesis el nombre de la variable en la que hemos guardado el enlace de la página web para que nos de `<response [200]>`, se vería así.



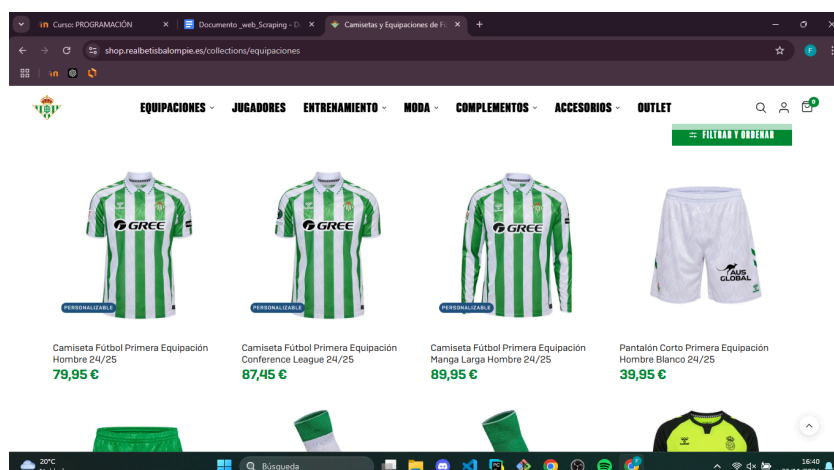
Y luego añadiremos el segundo atributo que será `.content` que lo que hará será sacarnos todo el código fuente de la página web.

Por último pondremos un `return` y la variable `response` para utilizarlo después.

```
def primera_respuesta_url(): 1 usage
    #Guardar la enlace de la página web en una variable
    url = "https://shop.realbetisbalompie.es/collections/equipaciones"

    #Pedir solicitud para comprobar a la página web (tiene que dar un 200)
    response = requests.get(url).content
    return response
```

Foto de la página web:



3. Segunda función

Esta función lo que hace es transformar el código como en una especie de sopa para que posteriormente podamos ir cogiendo elementos que queramos.

Lo primero que tenemos que hacer es escribir una nueva variable y dentro de esa nueva variable poner el nombre de la primera función que hemos creado.

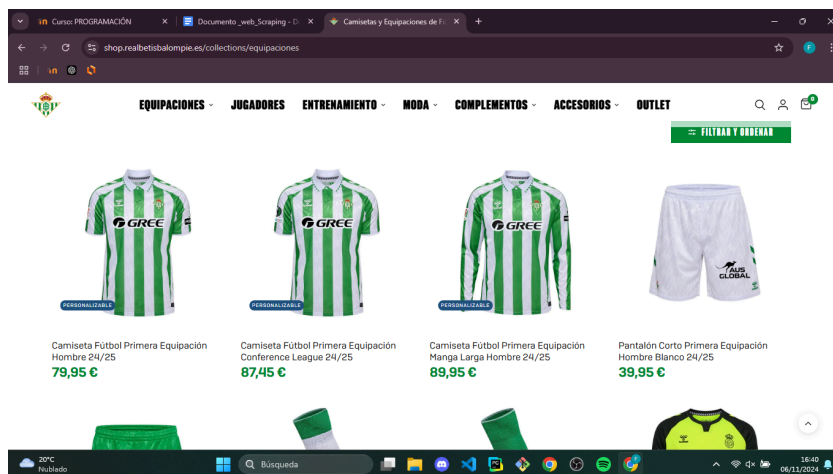
Luego lo que tenemos que hacer será transformar el contenido de la página web en una sopa para ello tenemos que escribir una nueva variable y utilizar la segunda biblioteca que hemos instalado previamente en el paso 1.

Por último escribimos un return con el nombre de la variable en el que hayamos hecho dicha transformación del código.

```
def transformar_sopa(): 1 usage
    response = primera_respuesta_url()

    #Transformar el contenido de la página web
    soup = BeautifulSoup(response, features: 'html.parser')
    return soup
```

Foto de la página:



4. Tercera función

Aquí tendremos varios elementos, lo primero que tendremos que hacer será crear una variable con el nombre de la segunda función que hemos creado:

```
soup = transformar_sopa()
```

Lo segundo que tendremos será acceder a los contenedores necesarios de la página web:

```
news_headlines = soup.find_all( name='li', class_='column')
```

Lo siguiente que tendremos que hacer será crear una plantilla de diccionario de lo que queramos guardar en él:

```
ropa = {  
    "Nombre": "",  
    "Foto": "",  
    "Tipo_prenda": "",  
    "Temporada": "",  
    "Precio": "",  
}
```

Lo siguiente que tendremos que hacer será hacer una copia de la plantilla del diccionario para ir metiendo todos los elementos que saquemos de la página web:

```
ropa_2 = ropa.copy()
```

Después lo que tenemos que hacer será crear una lista vacía porque allí meteremos todos los diccionarios una vez que los creamos:

```
lista_ropa = []
```

Lo siguiente que tenemos que hacer será hacer un bucle for que lo que hará es recorrer los elementos que le indiquemos de la página web para irlos añadiendo a los diccionarios y posteriormente a la lista. Después pondremos el nombre de la lista con su función .append() (dentro el nombre de la copia del diccionario) para ir añadiendo todos los diccionarios a la lista. Al final de dicho bucle pondremos un return para que nos devuelva todos los diccionarios añadidos a la lista:

```
for headline in news_headlines:

    #Nombre
    nombre_producto = headline.find('a')['title']
    ropa_2['Nombre'] = nombre_producto

    #Foto
    foto = headline.find('img')['data-src']
    ropa_2['Foto'] = foto

    #Tipo prenda
    tipo_prenda = nombre_producto.split()[0]
    ropa_2['Tipo_prenda'] = tipo_prenda

    #Temporada
    temporada = nombre_producto.split()[-1]
    ropa_2['Temporada'] = temporada

    #Precio
    precio = float(headline.find('span', class_='amount').text.replace("€", "").replace(", ", "."))
    ropa_2['Precio'] = precio

    #Guardar todos los diccionarios en una lista
    lista_ropa.append(ropa_2)

return lista_ropa
```

Foto de la página web:

