

Scrap de enlaces para página web

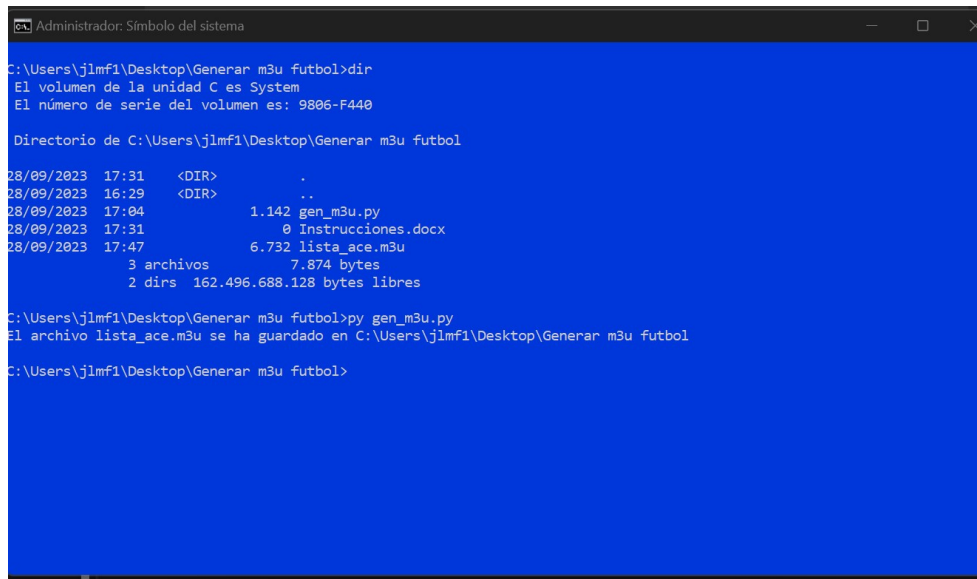
Utilidad programada en Python para scrapear enlaces de una web.

Necesitamos tener instalado **Python** y **Ace Player** en el pc y ejecutarla desde una pantalla de comandos como administrador:

Nos posicionamos en la carpeta que contiene el ejecutable y escribimos

py gen_m3u.py

Nos generará la lista y la guardará en la misma carpeta.



```
Administrador: Símbolo del sistema
C:\Users\jlmf1\Desktop\Generar m3u futbol>dir
El volumen de la unidad C es System
El número de serie del volumen es: 9806-F440

Directorio de C:\Users\jlmf1\Desktop\Generar m3u futbol
28/09/2023  17:31    <DIR>          .
28/09/2023  16:29    <DIR>          ..
28/09/2023  17:04                1.142 gen_m3u.py
28/09/2023  17:31                0 Instrucciones.docx
28/09/2023  17:47                6.732 lista_ace.m3u
                3 archivos             7.874 bytes
                2 dirs 162.496.688.128 bytes libres

C:\Users\jlmf1\Desktop\Generar m3u futbol>py gen_m3u.py
El archivo lista_ace.m3u se ha guardado en C:\Users\jlmf1\Desktop\Generar m3u futbol
C:\Users\jlmf1\Desktop\Generar m3u futbol>
```

El programa nos busca textos y enlaces de acestream en la página <https://futbolsinpagar.pages.dev/> . Esto se puede modificar por la que tu quieras. Puedes imprimir por pantalla los resultados si quieres (en el código adjunto está comentado para que se lo salte). Una vez que nos encuentra los enlaces, nos lo vuelca a una lista m3u con el nombre que le hemos puesto en “filename”, en el mismo directorio en el que tenemos el ejecutable Python. La abriremos con el acestream player.

El código Python es el que sigue (modificable a voluntad):

```
1. from bs4 import BeautifulSoup
2. import requests
3. import os
4.
5. def scrape_links(url):
6.     # Realizar la solicitud HTTP
7.     response = requests.get(url)
8.
9.     # Crear el objeto BeautifulSoup
10.    soup = BeautifulSoup(response.text, 'html.parser'
11.    )
12.    # Encontrar todos los elementos 'a' en la página
    (que corresponden a los enlaces)
13.    links = soup.find_all('a')
14.
15.    # Extraer el atributo 'href' y el texto de cada
    enlace
16.    urls = [(link.text, link.get('href')) for link in
    links if link.get('href').startswith('acestream://')]
17.
18.    return urls
19.
20.    def create_m3u_file(links, filename="lista_ace.m3u"):
21.        with open(filename, "w") as f:
22.            f.write("#EXTM3U\n")
23.            for name, url in links:
24.                f.write(f"#EXTINF:-1,{name}\n")
25.                f.write(f"{url}\n")
26.
27.            print(f"El archivo {filename} se ha guardado en
    {os.getcwd()}")
28.
29.    # Prueba la función con cualquier URL
30.    url = "https://futbolsinpagar.pages.dev/"
31.    links = scrape_links(url)
32.
33.    # Imprime en pantalla todos los enlaces y sus nombres
34.    #for name, link in links:
35.        # print(f" {name}, {link}")
36.
37.    # Crea el archivo M3U
38.    create_m3u_file(links)
```