

✓ [Actividad extracurricular 12] web scraping

Nombre: Adrian Francisco Correa Anrango

fecha: 04/01/2025

tema: web Scraping

- Revisar qué es web scraping
- Realizar una prueba en python para dos librerías diferentes
- Realizar scraping de un sitio web de su elección
- Subir pdf del notebook e informe.

link de la pagina web: <https://www.guiahardware.es/top-10-supercomputadores/>

```
import requests
from bs4 import BeautifulSoup

# URL del sitio web
url = "https://www.guiahardware.es/top-10-supercomputadores/"

# Descargar la página
response = requests.get(url)
response.raise_for_status() # Asegura que la solicitud fue exitosa

# Analizar el contenido HTML
soup = BeautifulSoup(response.text, 'html.parser')

# Extraer títulos (modificar según la estructura del sitio)
titles = soup.find_all('h2') # Ajustar si es necesario
descriptions = soup.find_all('p') # Ajustar si es necesario

# Mostrar resultados
for i, title in enumerate(titles[:10]): # Limitar a los 10 primeros
    print(f"Supercomputador {i + 1}: {title.text.strip()}")
    if i < len(descriptions):
        print(f"Descripción: {descriptions[i].text.strip()}")
    print("-" * 50)
```



Supercomputador 1: ¿Qué es HPC?

Descripción: Actualizado a: 19 de enero de 2024

Supercomputador 2: ¿Qué son los supercomputadores?

Descripción: Los supercomputadores son unos grandes desconocidos, a pesar de que muchos de nosotros hacemos uso de ellos a diario sin saberlo, como cuando usamos ciertos servicios de la nube. Este

Supercomputador 3: ¿Para qué se usan los supercomputadores?

Descripción: Contenido

Supercomputador 4: Top 10 supercomputadores más potentes de la actualidad

Descripción: La informática de alto rendimiento (HPC) es la capacidad de procesar datos y realizar cálculos complejos a gran velocidad. La High Performance Computing (HPC) logra estos objetivos me

```
-----
Supercomputador 5: Futuro: Computación cuántica
Descripción: Gracias a reunir diversas tecnologías, como la arquitectura de computación, los algoritmos, el software de sistemas, los programas y la electrónica, bajo un mismo paraguas, la HPC es
-----
```

```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.service import Service
from webdriver_manager.chrome import ChromeDriverManager

# Configurar Selenium
service = Service(ChromeDriverManager().install())
driver = webdriver.Chrome(service=service)

# Navegar al sitio web
driver.get("https://www.guiahardware.es/top-10-supercomputadores/")

# Extraer títulos y descripciones (ajustar selectores según sea necesario)
titles = driver.find_elements(By.TAG_NAME, "h2")
descriptions = driver.find_elements(By.TAG_NAME, "p")

# Mostrar resultados
for i, title in enumerate(titles[:10]):
    print(f"Supercomputador {i + 1}: {title.text.strip()}")
    if i < len(descriptions):
        print(f"Descripción: {descriptions[i].text.strip()}")
    print("-" * 50)

# Cerrar el navegador
driver.quit()
```



```
Supercomputador 1: ¿Qué es HPC?
Descripción: Actualizado a: 19 de enero de 2024
-----
Supercomputador 2: ¿Qué son los supercomputadores?
Descripción: Los supercomputadores son unos grandes desconocidos, a pesar de que muchos de nosotros hacemos uso de ellos a diario sin saberlo, como cuando usamos ciertos servicios de la nube. Est
-----
Supercomputador 3: ¿Para qué se usan los supercomputadores?
Descripción: Contenido
-----
Supercomputador 4: Top 10 supercomputadores más potentes de la actualidad
Descripción: La informática de alto rendimiento (HPC) es la capacidad de procesar datos y realizar cálculos complejos a gran velocidad. La High Performance Computing (HPC) logra estos objetivos me
-----
Supercomputador 5: Futuro: Computación cuántica
Descripción: Gracias a reunir diversas tecnologías, como la arquitectura de computación, los algoritmos, el software de sistemas, los programas y la electrónica, bajo un mismo paraguas, la HPC es
-----
```



