

Que es Scrapy?

Scrapy es una plataforma colaborativa de código libre que corre en Python para extraer datos de páginas web usado para una serie de aplicaciones como minería de datos, procesamiento de información o registro histórico.

Scrapy no es una librería como pudiera parecer, sino un completo Framework de web scraping para Python.

A grandes rasgos, el funcionamiento de Scrapy es el siguiente:

- En primer lugar, el **Engine** recibe las **peticiones (Requests)** iniciales que le envía la **araña (Spider)**, las programa en el scheduler y solicita las siguientes peticiones a rastrear.
- El **Scheduler** va enviando las peticiones a procesar al Engine y este a su vez las envía al componente **Downloader**.
- La página es descargada y se crea una **respuesta (Response)** de esta página que el Engine se encarga de enviar a la araña para que sea procesada.
- Es entonces cuando la araña o rastreador devuelve al Engine los **ítems (Items)** con la información extraída de la página.
- Después, estos ítems son enviados a los **Pipelines** para procesar y almacenar la información.
- El proceso se repite hasta que el Scheduler se quede sin peticiones.

Dado que es un framework, Scrapy tiene una serie de herramientas poderosas para hacer el "scraping" o extraer información de webs de manera fácil y eficiente.

Una de las principales ventajas de este framework es que las peticiones se procesan y programan de manera asíncrona. Debido a ello, Scrapy no espera a que termine una petición para enviar otra y se pueden ejecutar a la vez de forma concurrente, acelerando el proceso en gran medida. Además, si una petición falla, el resto de las peticiones seguirán ejecutándose.

Una de las mayores desventajas en los framework es la curva de Aprendizaje que en si puede ser muy exhaustiva y cansadora a la hora de aprender todas su funciones.

Para conocer mas ventajas y desventajas: [✓ Ventajas y ✗ Desventajas de los Frameworks de Desarrollo Web para PHP 🙌 ¡NECESITAS SABER ESTO! - YouTube](#)



Referencias:

[Pros y contras de los frameworks de desarrollo web - Victor Robles | Victor Robles \(victorroblesweb.es\)](#)

Que es Scrapy: <https://elmundodelosdatos.com/extraccion-datos-sitios-web-productos-zara/>
<https://gidahatari.com/ih-es/extrae-informacion-de-paginas-web-con-scrapy/>

Framework: <https://www.edix.com/es/instituto/framework/>