

## HERRAMIENTAS DE SCRAPING

El web scraping es una técnica que se utiliza para la extracción y almacenamiento de información de cualquier página web a través de un programa de software, que suele ser un crawler.

Hacer web scraping es algo relativamente sencillo, todo depende de la cantidad de información que necesites extraer de un sitio web.

Puedes hacerlo de forma manual, aunque también hay herramientas y software que ayudan en la tarea.

## WEB SCRAPING MANUAL

El scraping manual es tan sencillo como **seleccionar, copiar y pegar los datos o el contenido de una página web**. Se utiliza cuando se quiere scrapear una página pequeña, o una sección concreta de cualquier web.

Cuando la página es muy grande, o es necesario recopilar información compleja, esta técnica es muy laboriosa y casi no se utiliza en esos casos, ya que puede llevar mucho tiempo y trabajo extraer información de muchos sitios web para estudiarlos de manera manual. No es rentable para el scraper.

## . WEB SCRAPING AUTOMÁTICO

Es la forma más habitual de hacer scraping web. Se utiliza para **obtener grandes cantidades de datos** de una o muchas páginas web. Para llevarlo a cabo es necesario el uso de un algoritmo o software que extraiga la información.

Hay diferentes maneras de hacerlo:

### Mediante el uso de bots

Están programados para hacer diferentes tareas de manera automática; en este caso, extraer información de una web.

### A través de un parser o analizador sintáctico

Un parser **convierte una pieza de texto** en otro tipo de estructura para almacenar la información.

### Análisis de textos

**Este método es para scrapers experimentados.** Utiliza la función “grep” de Unix para encontrar algunos términos de la web en Perl o Python. Este método exige mucho más trabajo que la simple utilización de un software.

HERRAMIENTAS MANUALES tienen sus **desventajas** que serían el poco alcance, o el mayor tiempo que lleva hacer el trabajo, no son escalables, son más costosas debido a que se necesita un lugar físico, mayor cantidad de personas, etc. Son más utilizadas para trabajar en páginas chicas.

A diferencia de las HERRAMIENTAS AUTOMATIZADAS, las **ventajas** que tienen la velocidad de acción de caracteriza a todo sistema automatizado. Son escalables, se puede realizar scraping en cualquier momento y lugar, no se necesita una estructura física, incluso se puede realizar desde una maquina local o tener una infraestructura en la nube. Se encargan de actuar ante herramientas “anti-bot”. No es necesario tanto conocimiento de programación, todo está al alcance de un “click”.

**CONCLUSION;** podemos decir que las herramientas, ya sean manuales o automatizadas, tienen distintos usos y/o preferencias. Y se las puede usar de manera particular o en conjunto; es decir, las herramientas automatizadas tienen muchas ventajas, pero falta la conciencia del ser humano. En algunos casos tanta automatización, presenta un margen de error.

### Bibliografía

Koller, S. (17 de 11 de 2021). *¿Qué es el web scraping, para qué sirve y cómo funciona?*  
Obtenido de [seranking.com/es/blog/web-scraping/](https://seranking.com/es/blog/web-scraping/)

Tobella, P. (04 de agosto de 2021). *Introducción a las Técnicas y Herramientas de Web Scraping*. Obtenido de [www.octoparse.es/blog/introduccion-a-las-tecnicas-y-herramientas-de-web-scraping](https://www.octoparse.es/blog/introduccion-a-las-tecnicas-y-herramientas-de-web-scraping)

