

SCRAPING

¿Qué es scraping y para qué sirve?

El scraping de datos, de un modo general, se refiere a una técnica en la cual un programa informático extrae datos del resultado generado por otro programa. El scraping de datos se manifiesta normalmente en el scraping web, el proceso de utilizar una aplicación para extraer información valiosa de un sitio web.

El web scraping se refiere al proceso de extracción de contenidos y datos de sitios web mediante software. Por ejemplo, la mayoría de los servicios de comparación de precios utilizan web scrapers para leer la información de precios de varias tiendas online. Otro ejemplo es Google, que rutinariamente scrapea o «rastrea» la web para indexar sitios web.

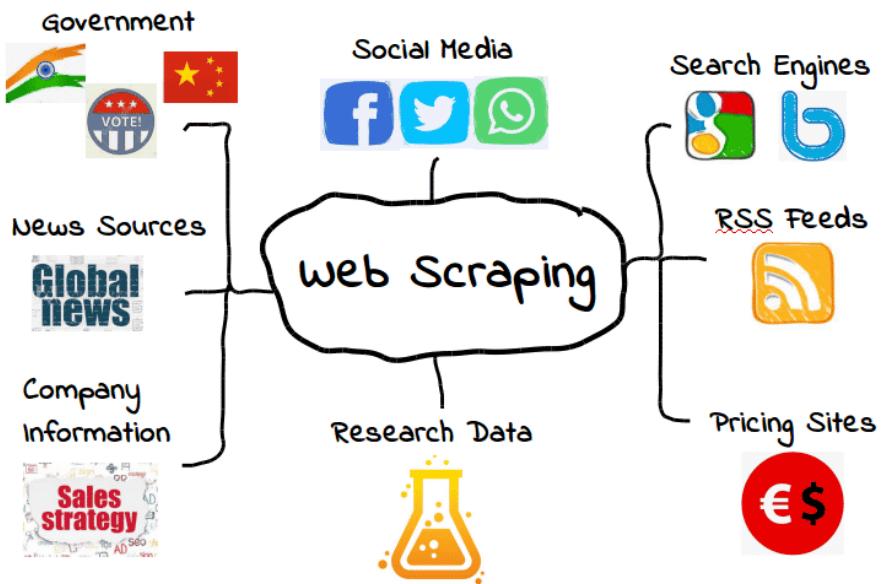
¿Qué podemos scrapear de la web?

Es posible scrapear todo tipo de datos de la web. Desde los motores de búsqueda y los feeds RSS hasta la información gubernamental, la mayoría de los sitios web ponen sus datos a disposición de los scrapers, crawlers y otras formas de recopilación automática de datos.

Sin embargo, eso no significa que estos datos estén siempre disponibles. Dependiendo del sitio web, puede que tengas que emplear algunas herramientas y trucos para obtener exactamente lo que necesitas – suponiendo que los datos sean accesibles en primer lugar. Por ejemplo, muchos scrapers web no pueden extraer datos significativos del contenido visual.

El web scraping se ha convertido en la columna vertebral de muchos procesos basados en datos, desde el seguimiento de las marcas y las comparaciones de precios actualizadas hasta la realización de valiosos estudios de mercado. He aquí algunos de los más comunes.

- **Estudio de Mercado**
- **Automatización del Negocio**
- **Generación de Leads**
- **Seguimiento de Precios**
- **Noticias y Contenidos**
- **Monitorización de la Marca**
- **Inmobiliaria**



¿Cómo se usa scraping?

Recopilar las URL de las páginas de las que desea extraer datos. Realizar una solicitud a estas URL para obtener el HTML de la página. Utilizar localizadores para encontrar los datos en el HTML. Guardar los datos en un archivo JSON o CSV o en algún otro formato estructurado.

Dentro de sus usos legítimos se encuentran los motores de búsqueda web tales como Google, los cuales rastrean los diferentes sitios web, para analizar su contenido y almacenarlo.

¿Cuál es el mejor lenguaje para hacer web scraping?

Python, el popular lenguaje de programación se presta especialmente bien para la creación de programas de web scraping. Puesto que las páginas web han de ser constantemente modificadas y actualizadas, sus contenidos cambian con el tiempo. Puede que cambie su diseño, por ejemplo, o que se les añadan nuevos elementos. Fundamentalmente, por ser un lenguaje intuitivo, fácil de usar y por las librerías y herramientas que han surgido en torno al mismo para llevar a cabo esta técnica.

- Utiliza las librerías *requests* y *Beautiful Soup* de manera conjunta
- Utiliza el framework *Scrapy*. Este framework, además de hacer **scraping** (extraer información de una página) permite hacer **crawling** fácilmente (descubrir los enlaces de una web y navegar a través de ellos).

¿Es Legal el Web Scraping?

Afortunadamente, no hay nada intrínsecamente ilegal en el web scraping. Cuando un sitio web publica datos, normalmente están disponibles para el público y, por tanto, son libres de ser scrapeados.

Por ejemplo, dado que Amazon pone a disposición del público los precios de los productos, es perfectamente legal scrapear los datos de los precios. Muchas aplicaciones de compra populares y extensiones de navegador utilizan el web scraping con este mismo propósito, para que los usuarios sepan que están obteniendo el precio correcto.

Sin embargo, no todos los datos de la web están hechos para el público, lo que significa que no todos los datos de la web son legales para scrapear. Cuando se trata de datos personales y de propiedad intelectual, el web scraping puede convertirse rápidamente en web scraping malicioso, lo que puede dar lugar a sanciones.

En Argentina, no existe una legislación específica que regule expresamente el web scraping. No obstante, podemos distinguir varias leyes y regulaciones que pueden aplicarse a esta técnica, incluyendo:

- **Ley de Protección de Datos Personales.**
- **Ley de Propiedad Intelectual.**
- **Ley de Lealtad Comercial.**
- **Términos de uso del sitio web.**

Fuente:



<https://j2logo.com/python/web-scraping-con-python-guia-inicio-beautifulsoup/#:~:text=El%20web%20scraping%20es%20una,ello%20la%20librer%C3%A3da%20Beautiful%20Soup.>

<https://leponte.com.ar/implicancias-legales-del-web-scraping/>