

¿Qué es la técnica de Scraping?

El Web scraping es una técnica mediante la cual es posible obtener información estructurada a partir de un sitio web, para ello se realiza una simulación de la navegación humana con el objetivo de realizar un análisis de datos (Big Data) o realizar procesos Web automatizados.

¿Cuáles son sus principales usos?

La técnica del Web scraping suele ser muy utilizada para obtener un gran volumen de información de manera automatizada. Esta información puede resultar muy útil para analizar las webs de la competencia, mejorar los recursos dedicados al marketing o al SEO, para labores de comercial y ayudar en la toma de decisiones.

Otros posibles fines del Web scraping pueden ser la automatización de tareas como por ejemplo: reserva de productos o servicios web, agregación de información bancaria, iniciación de pagos, verificación de servicios Web.

Ventajas de utilizar web crawling y data scraping

Con web crawling y data scraping los procesos de encontrar y recabar información se automatizan, con ello conseguimos:

- Disminuir la carga de trabajo.
- Abaratar costes de personal.
- Aumentar la velocidad de los procesos.
- Eliminar el error humano.
- Manejar grandes cantidades de datos.
- Conseguir los datos en formatos procesables.

¿Es legal? ¿Existe legislación al respecto? Si existe, citarla.

El “web scraping” suele estar permitido cuando:

- Los datos extraídos son datos disponibles públicamente; y
- La información recopilada no está protegida por un inicio de sesión.

En Argentina, la legalidad del scraping está sujeta a la Ley de Protección de Datos Personales (Ley 25.326), que regula la protección de datos personales. El scraping de datos personales sin el consentimiento adecuado podría infringir esta ley. Además, se debe respetar cualquier legislación de derechos de autor que aplique a los sitios web y los términos y condiciones de uso de dichos sitios.

¿Qué lenguajes de programación pueden utilizarse para la misma?

¿Qué librerías se utilizan?

Python: Es el lenguaje más popular para el raspado web debido a su facilidad de aprendizaje, código abierto y flexibilidad. La biblioteca más utilizada es BeautifulSoup, que

permite una extracción de datos rápida y personalizable. Scrapy, otro framework de Python, es especialmente útil para la minería de datos y la descarga masiva de datos.

JavaScript: Se utiliza cuando los sitios web tienen una representación dinámica de la interfaz de usuario. Sin embargo, no se recomienda para proyectos que requieran la extracción de datos en varios sitios web de manera repetitiva debido a su baja estabilidad en la comunicación.

Node.js: Es útil para recopilar datos de páginas web avanzadas y básicas, especialmente en páginas web dinámicas.

C++: Aunque es un lenguaje con buen rendimiento y códigos complejos, no se recomienda para el raspado web en negocios debido a su complejidad.

PHP: Es capaz de realizar el raspado de muchos sitios web en poco tiempo y es efectivo para extraer datos y programar tareas en diversas páginas web.

En conclusión Python es altamente recomendado debido a su facilidad de uso y las bibliotecas disponibles, mientras que JavaScript es adecuado para casos específicos de sitios web dinámicos. Node.js es útil para sitios web avanzados, mientras que C++ se considera menos adecuado para el raspado web en entornos comerciales. PHP es eficaz para extraer datos y programar tareas en múltiples sitios web.

De tres ejemplos en los que podría utilizarse la técnica.

Comparación de Precios en Comercio Electrónico: Monitorizar y comparar precios de productos en diferentes tiendas en línea.

Análisis de Noticias Financieras: Recopilar noticias y comentarios sobre eventos económicos para la toma de decisiones financieras.

Gestión de Reputación en Redes Sociales: Recopilar comentarios y reseñas de usuarios en redes sociales para evaluar la percepción de la marca

gaston trejo
33.693.178

Bibliografía:

<https://www.amnislabs.com/web-scraping>

<https://datstrats.com/blog/que-es-scraping-usos-aplicaciones/>

<https://www.iubenda.com/es/help/111939-es-legal-el-web-scraping-lo-que-hay-que-saber>

<https://www.areacucuta.com/que-lenguajes-son-empleados-en-web-scraping/>