



Web Scraping - milanuncios

Master en ciencia de datos

Tipología del ciclo de vida de los datos

Autores:

Adrian Felipe Pinzón Hurtado

Diego Armando Quintero Quiñones

Noviembre/2021

¿Qué es web scraping?

El web scraping consiste en **navegar automáticamente una web y extraer de ella información**. Esto puede ser muy útil para muchísimas cosas y beneficioso para casi cualquier negocio.

Conocido también como: **bot, spider o crawler**



Contexto web scraping Mil Anuncios

PROBLEMA:

- Datos desordenados en páginas web
- Muy dispersos o sin un orden o estructura definida
- Comportamiento del mercado de compra / venta de productos..

SOLUCIÓN:

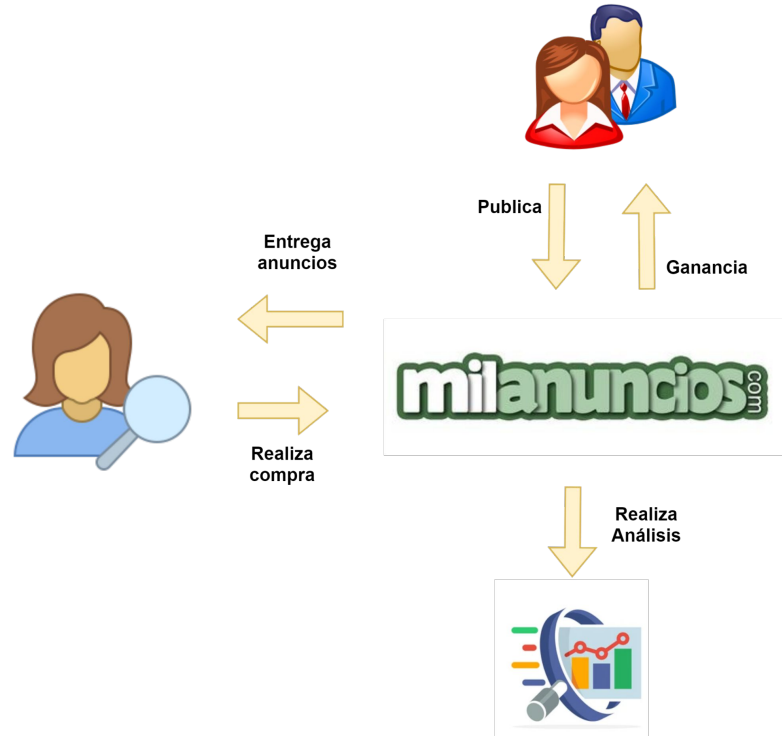
- Tomar las páginas web donde se quiere extraer la información.
- Extraerla, ordenarla y convertir los datos en conocimiento.
- Para el problema específico, validar modelos de negocio, oferta y demanda de productos.



Dominio de los datos

Representación gráfica del dominio de datos en el contexto de las publicaciones online. Aquí se muestran los actores que interactúan:

- Vendedor: Persona que publica el anuncio
- Comprador: Persona que quiere adquirir un producto o servicio
- Intermediario: En este caso realizado por milanuncios.
- Terceros: Capturan y analizan datos.





Ventajas web scraping

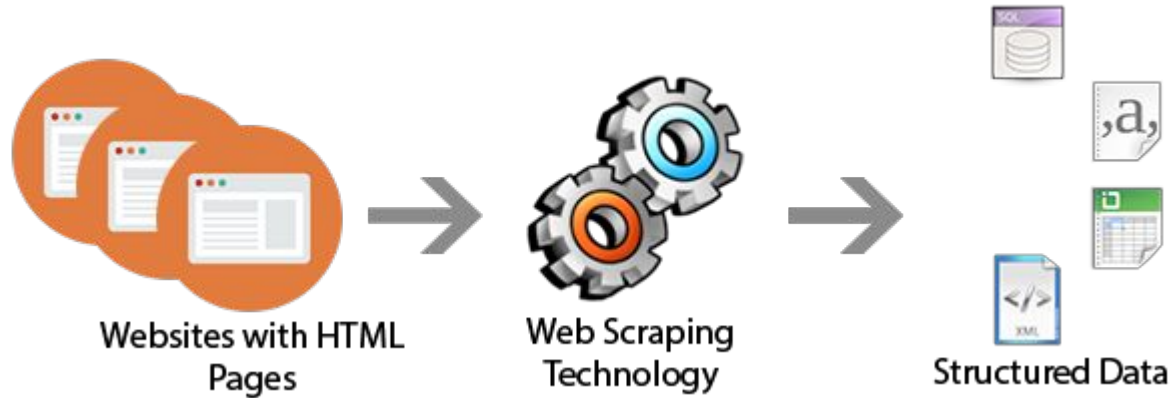
- Generar ideas valiosas de negocio
- Validar ideas
- Ahorrar tiempo y esfuerzo
- Rapidez
- Recopilar datos de fuentes más precisas



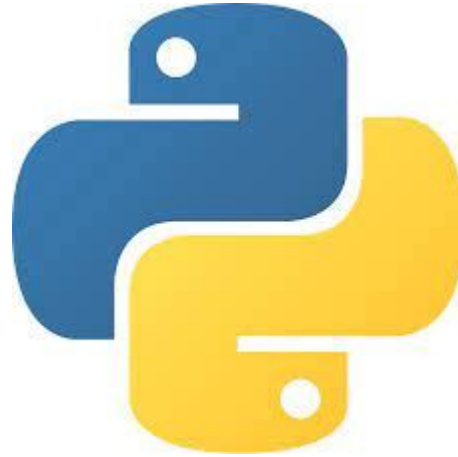
¿Cómo aplicarlo?

- Encuentra las herramientas adecuadas
- Identifica los sitios que deseas analizar
- Analiza las informaciones en busca para lo que quieres

Arquitectura de referencia



Breve explicación del código





Gracias....