

Wilmer Rodríguez Sánchez
&
Universidad Panamericana

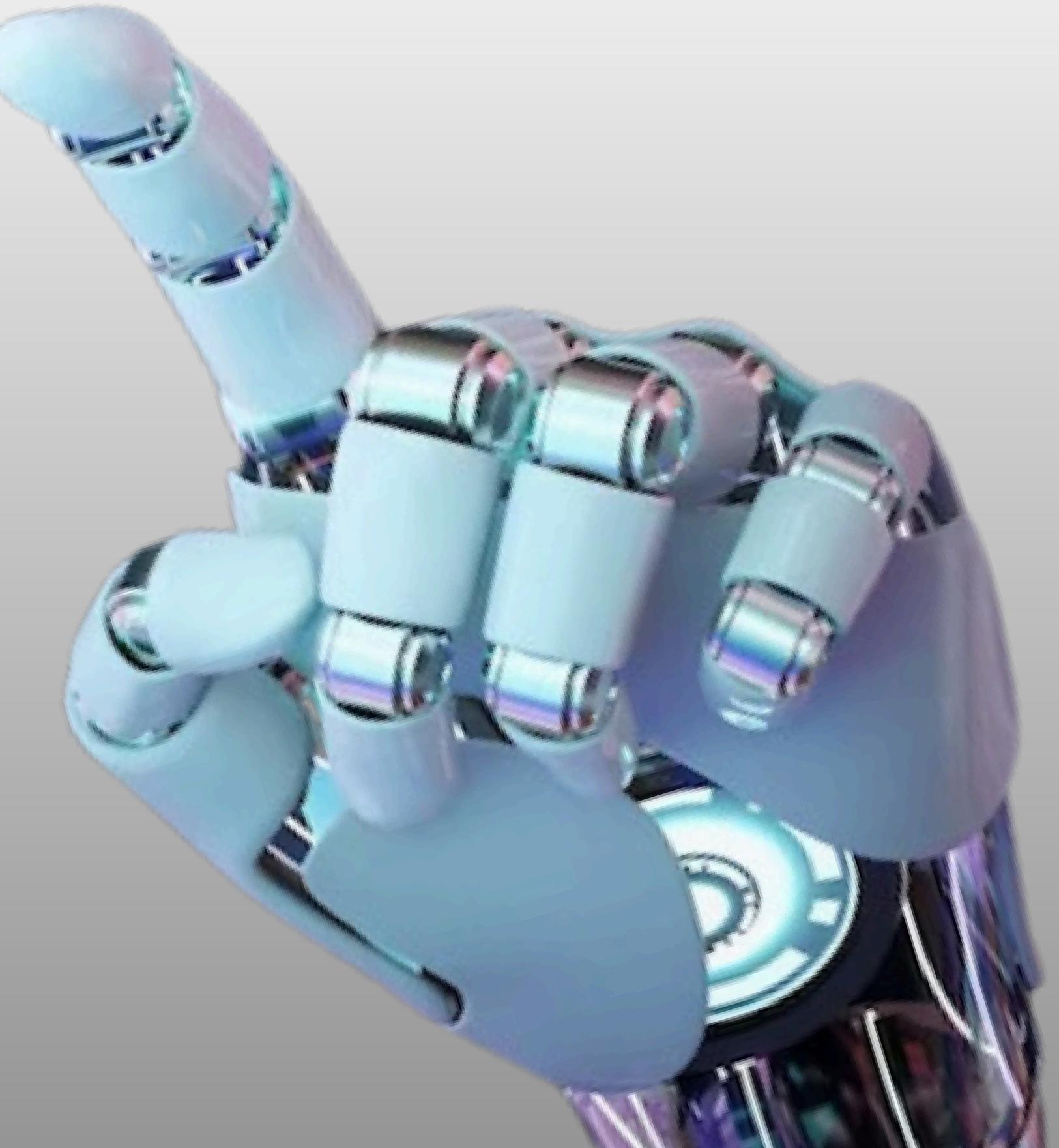
Web Scraping





¿Qué es Web Scraping?

Técnica de extraer datos automáticamente desde páginas web, usando programas que leen el contenido HTML y lo transforman en información útil.

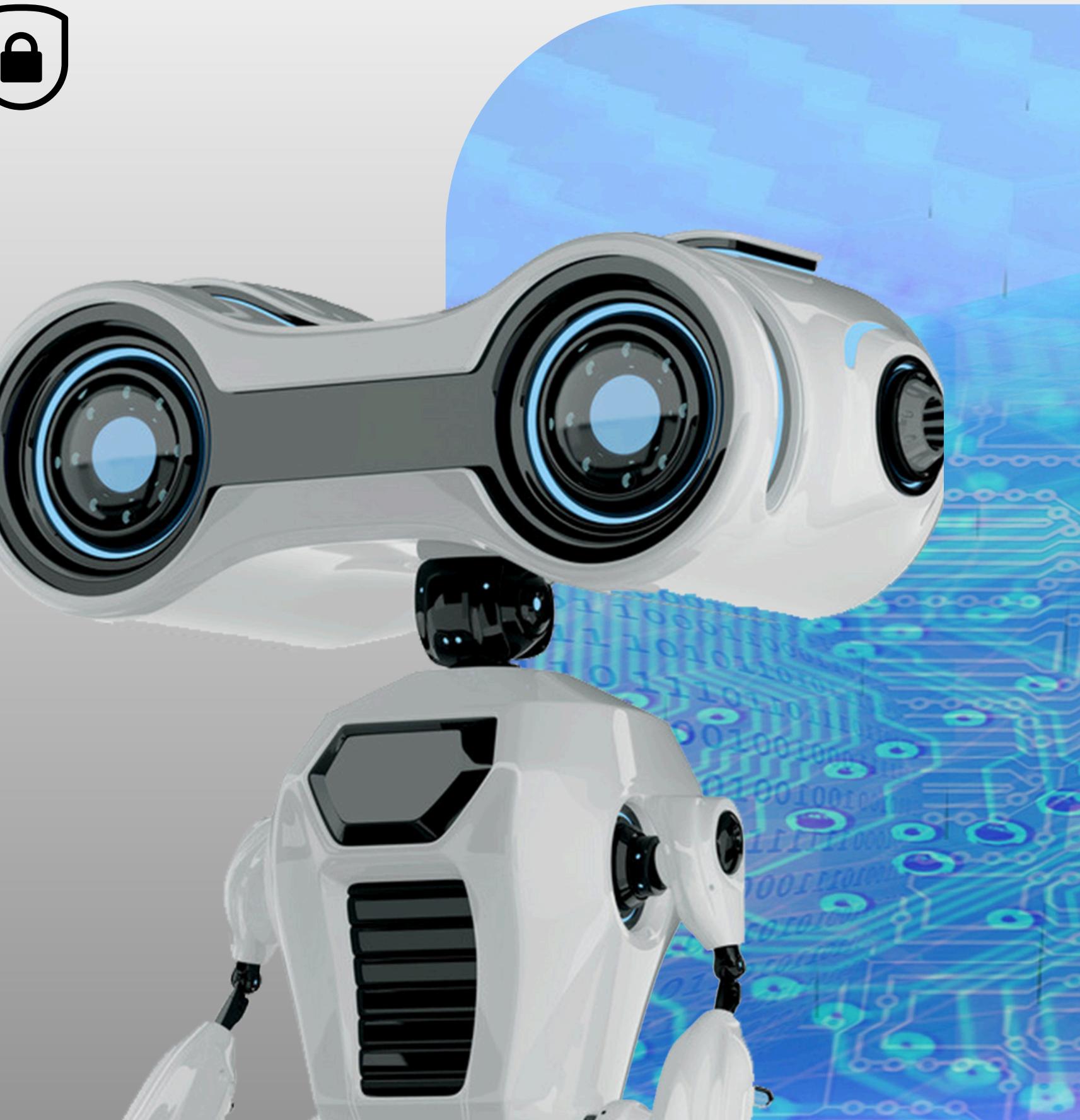


Flujo básico

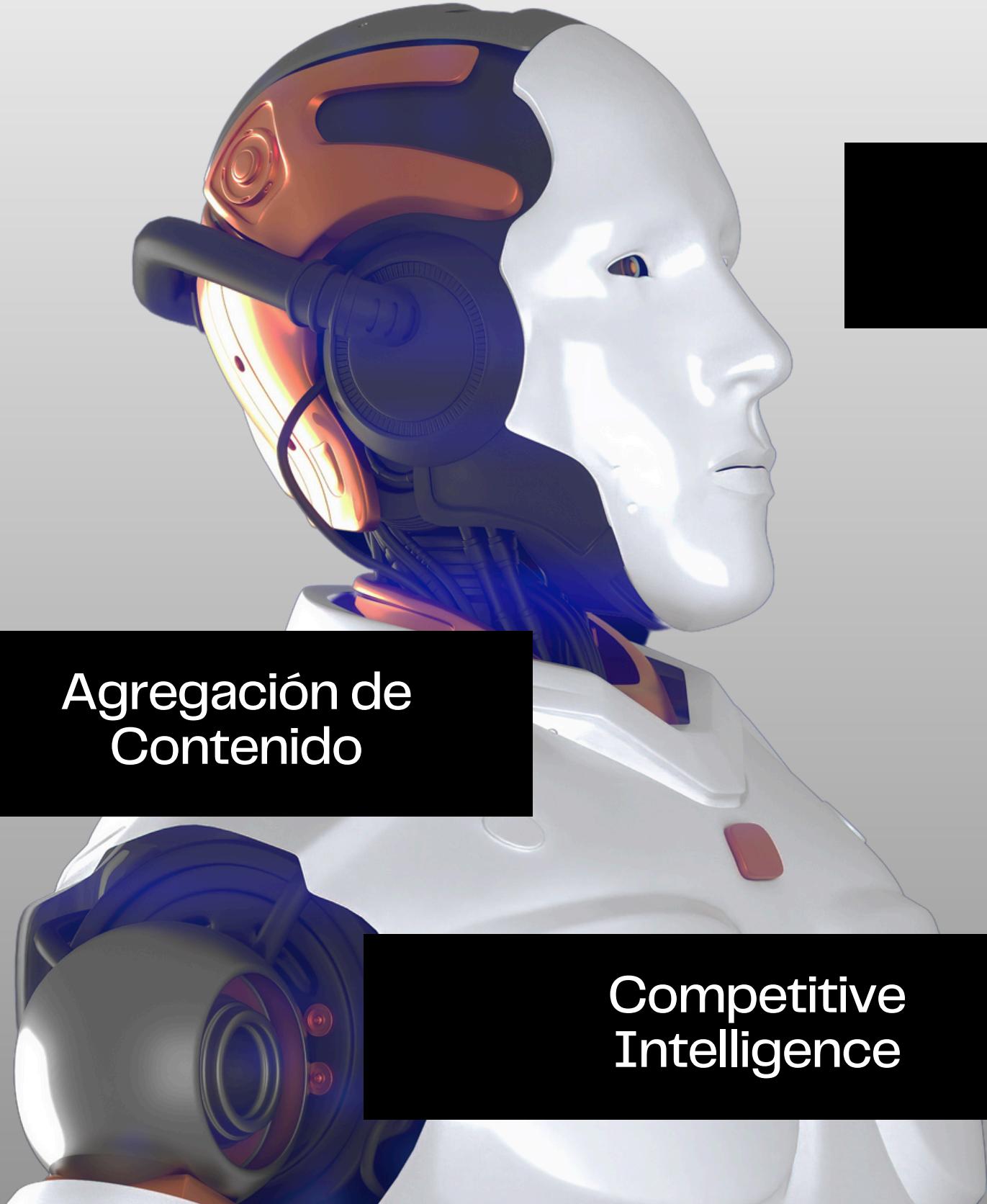


1. Request (Solicitud): El scraper solicita una página web a un servidor (usando HTTP/S).
2. Response (Respuesta): El servidor devuelve la respuesta, generalmente en formato HTML, JSON, o XML.
3. Parsing (Análisis): El scraper lee y analiza el contenido de la respuesta para localizar los datos de interés.
4. Extraction (Extracción): Los datos son extraídos y organizados (ej. en un objeto Python o diccionario).
5. Storage (Almacenamiento): Los datos se guardan en un formato estructurado (CSV, JSON, base de datos).

¡Cada proyecto de Web Scraping es diferente!



Data Science /
Machine Learning



Agregación de
Contenido

Competitive
Intelligence

Monitoreo y
Alerta

Casos de uso



Cada página representa un reto distinto:
estructuras únicas, datos ocultos, protecciones
diversas y formatos impredecibles.
Esa diversidad lo vuelve versátil y desafiante. Al
dominarla, la información extraída se convierte en
ventaja estratégica, conocimiento accionable y
verdadero poder para análisis, decisiones y
automatización.

Límites Legales y Éticos

El scraping no es ilegal per se, pero su implementación puede serlo.

Archivo
`/robots.txt`

Carga sobre el
Servidor
Su scraper debe
ser "amigable".

Términos de
Servicio (ToS)
del sitio web

Protección de
derechos de
autor



Herramientas Clave en Web Scraping



Python

- Requests
 - BeautifulSoup / lxml
 - Scrapy
 - Selenium / playwright
- Production Env
- Docker
 - n8n, zapier, make.
- ...



Se puede Web Scraping con muchas tecnologías y lenguajes de programación.

¡El límite es tu imaginación!



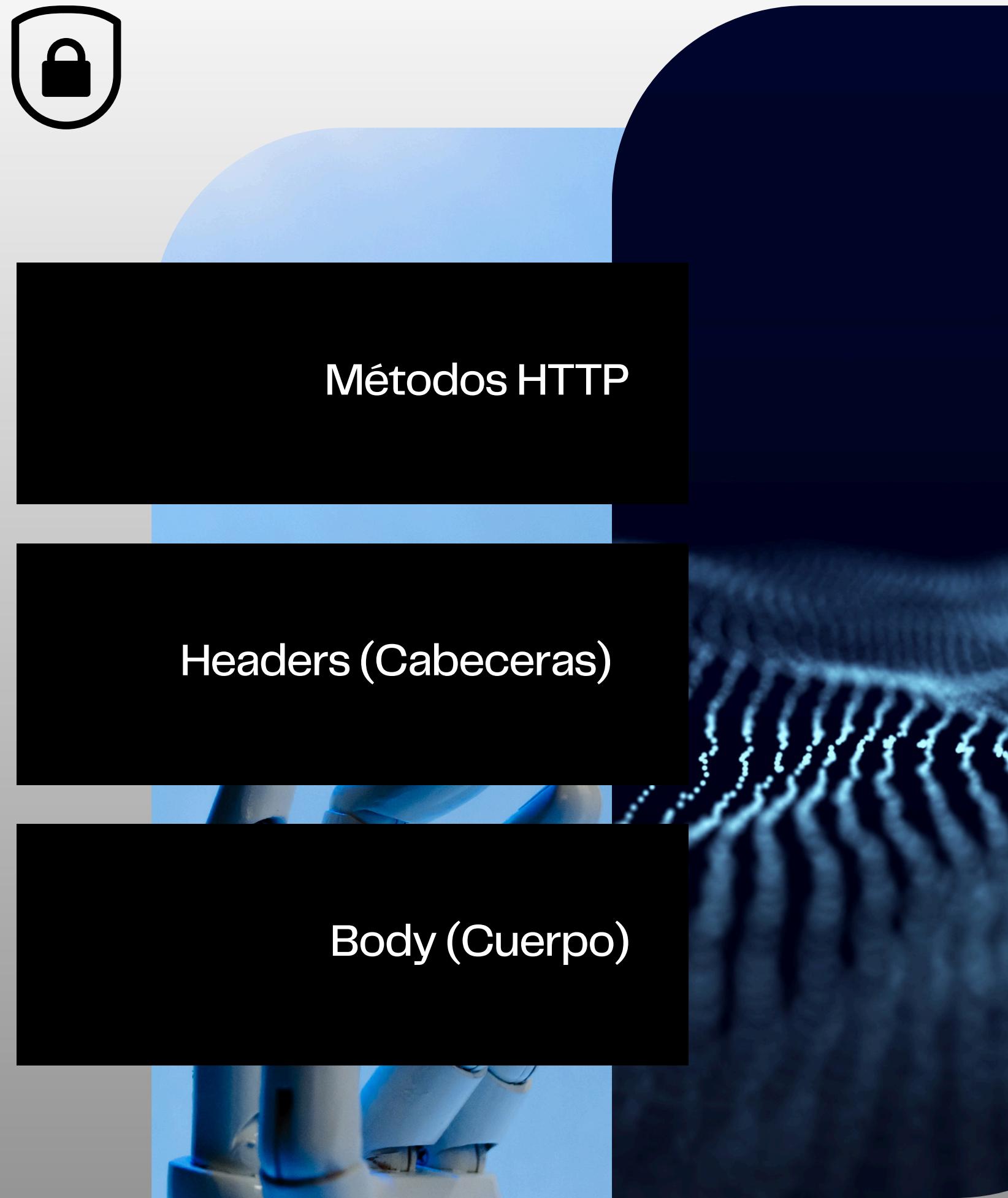
Anatomía de la Request y la Response

Arquitectura: Cliente–Servidor

Métodos HTTP

Headers (Cabeceras)

Body (Cuerpo)





Clonar repositorio de Github

[https://github.com
/wrsbyte/web-scraping-course](https://github.com/wrsbyte/web-scraping-course)



Gracias por su atención



+52 5513324874



wrsbyte.com



me@wesbyte.com



<https://linkedin.com/in/wrsbyte>

¿Preguntas?