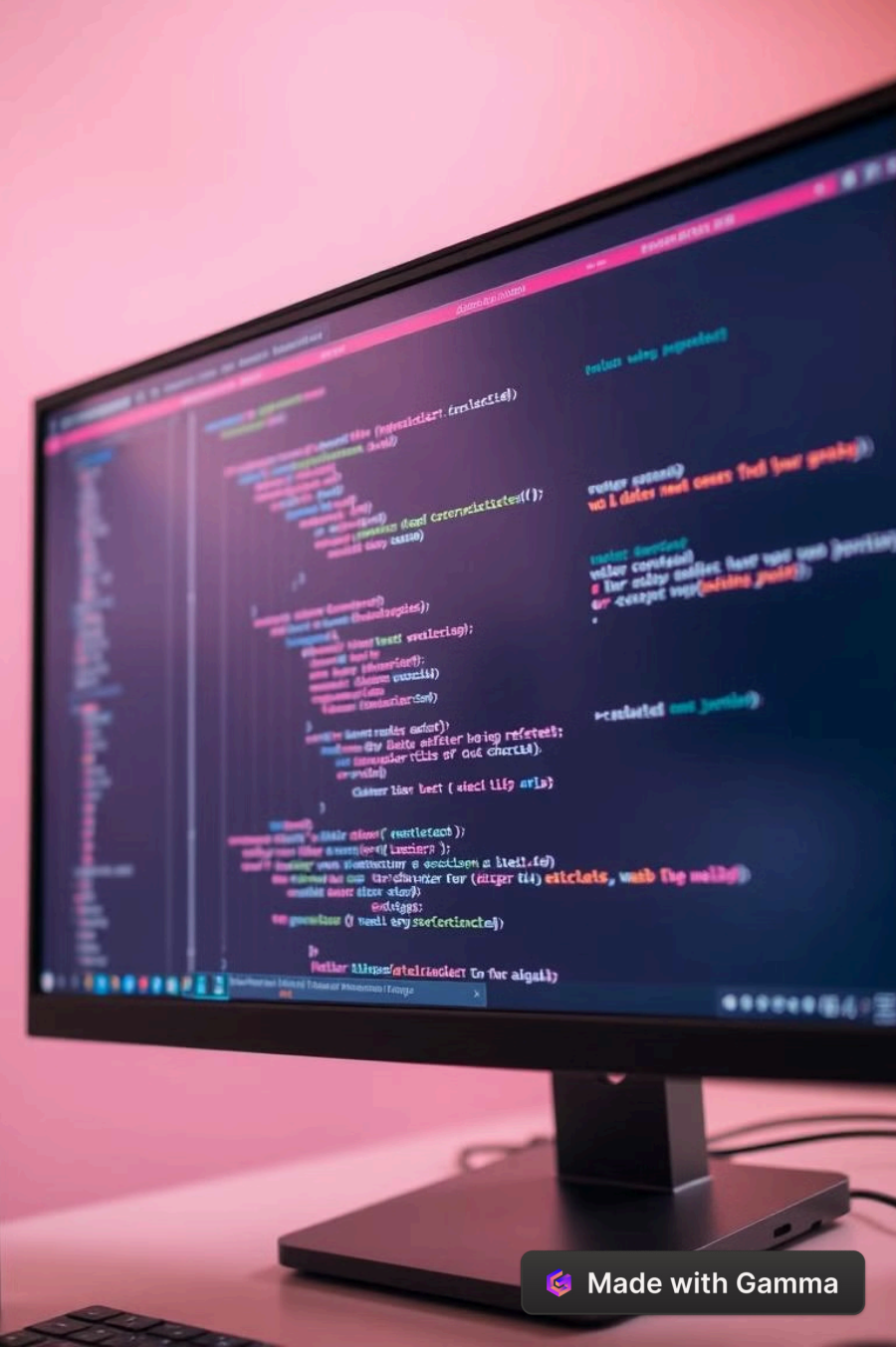


Web Scraping con Python y BeautifulSoup

Descubre el poder del web scraping para extraer datos valiosos de la web. Automatiza la recopilación de información y optimiza tus análisis. Python y BeautifulSoup son tus aliados esenciales en este proceso. ¡Aprende a dominar estas herramientas!



¿Qué es Web Scraping?

El web scraping es una técnica para extraer información estructurada de páginas web. Esta información se convierte en un formato legible. Es útil para monitorear precios, recopilar datos de redes sociales y generar leads.

Monitoreo de precios

Compara precios de televisores en Amazon y Mercado Libre.

Análisis de sentimiento

Extrae tweets y analiza si son positivos o negativos.

Generación de leads

Scraping de perfiles de LinkedIn para encontrar clientes.

Python para Web Scraping

Python es ideal por su sintaxis clara, gran comunidad y librerías potentes. Requests obtiene el HTML, BeautifulSoup lo parsea, y Pandas manipula los datos.



Sintaxis clara

Facilita el desarrollo.



Ampio soporte

Gran cantidad de recursos.



Librerías clave

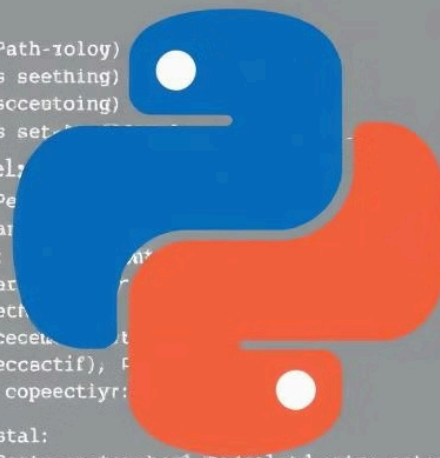
Requests, BeautifulSoup, Pandas.

```
>>>
wastlvilet (ntea.gol;
wasthiles (acrlcanl):

>>>
wetrstie, #1 -"07221.gq"
wastliliet "0-101 = "07222_09"
Wastle "The Totlef pal22-Clker;
matrrollfrier Toc228 - j-65110)"
Wettlef 10 -sible aprphaler-seettin: Marting pythters 2-2710)
wast friet (otker a: accocty" = "0010>,
wastlinat (Blos Perytlmber regabincaner, lote reetler, dartaters/ 3gol)

>>> restritation:
west.forak.(Ccake Path-ology)
west.forat.(seetltes seething)
west.forat (Stake sccentoing)
wastlfctat (eatlers set
nosthore_geti/tiel;
wnst.forke (Stake Pe
what.forat (efeghtar
wast.faract: Fotel:
west.forer (Ccatblat
west.lotat (Tras sech
wnst.forat (Spalt cecet
whatler ef (oter seccactif),
wastlvilet ratiler copectiyr:

>>> precpa/ciaato-estal:
west.friet (Oteer Cottr-anstopehcal medcalstd entra enterrs-perthied)
wastlvriet.(Drigary wribe sequuran bata estaly stras)
wast.friet.(Oteer Coatr.conbshatoge: woth.Ortestin;
wastlvriet (Oteer Fortr according reprinnolate cetibor: tecres")
wast.friek.(Oteer Coatr.abet.golger coths respechlings, serticuts let
wastlvriet (erelloce corpotatiol:
```



Beautiful Soup al Detalle

1

Nombre Elegante

"Beautiful Soup" alude a la limpieza y el orden en el caos del HTML.

2

Parsing Eficiente

Convierte documentos HTML complejos en estructuras navegables y fáciles de analizar.

3

Navegación Sencilla

Permite buscar, modificar y extraer datos con métodos intuitivos y poderosos.

Beautiful Soup al Detalle

Beautiful Soup facilita la extracción de datos de HTML y XML. Navega por el árbol HTML, busca elementos y extrae datos específicos.

Navegación

Por el árbol HTML.

Búsqueda

Por etiqueta y atributo.

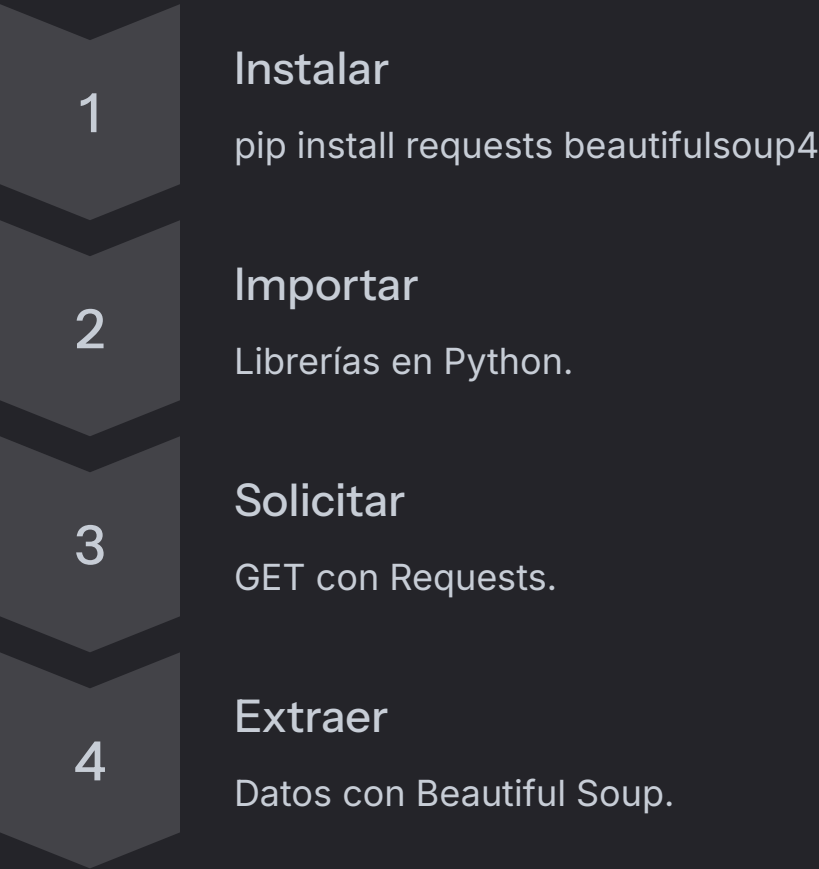
Sintaxis

Sencilla y precisa.

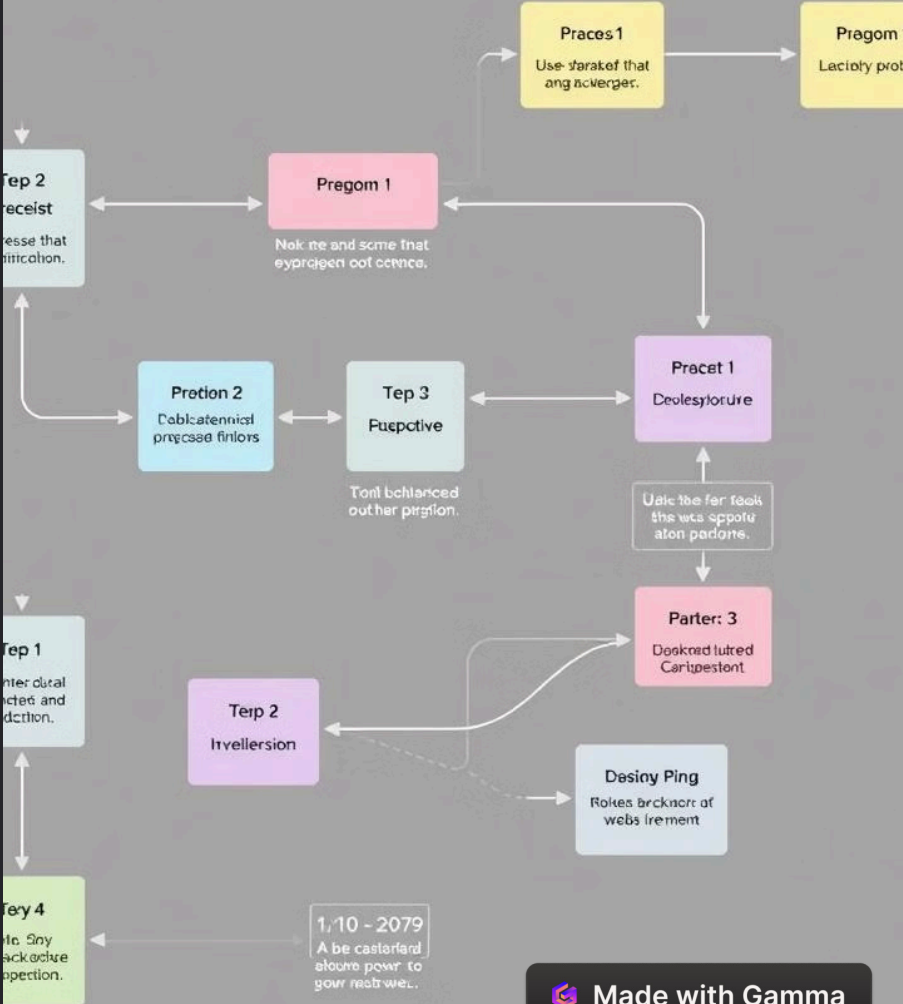


Pasos para Web Scraping

Sigue estos pasos: instala librerías, importa, solicita el sitio con Requests. Crea un objeto BeautifulSoup y extrae los datos deseados. ¡Así de sencillo!



Web Scrapping



Extrayendo Títulos de un Blog

Aplica este código para extraer títulos de artículos. ¡Comienza tu proyecto de web scraping ahora mismo!

```
import requests
from bs4 import BeautifulSoup

url = 'http://ejemplo.com/blog'
response = requests.get(url)
soup = BeautifulSoup(response.content, 'html.parser')

titulos = soup.find_all('h2', class_='titulo-articulo')
for titulo in titulos:
    print(titulo.text)
```



```
4 Pytrent / Pinthe Os, Felleny Nlogs)
Hellcad Kictiont Putlog
Slt Pandass> Applis nunds ac1 Testeert 6 At Combot
Piteblct = Pytting
Sopnter Jany yises'
Copntce Titles: Ittalebacrs Vales:
Copntce Titles: Sttiter eselar
Copntce Titles: Ittelagen siays
Copntce Titles: Sttelabck istale)
Copntce Moler - Pampse:
Copntce Terrasiando Rew donlam- sian
Copntce Reet ssurto siags
Copntce Titles: Tittle later-Doar-Fer soing:
Copnter Titles: Tittelring sine!
>
Mopntce Titles: Datclag tortne tian:
Copntce Titles: Suttlobdilig Peprarts thatie Busst-raclovs.. Ausently copper
Copntce Title: = Thdrag:
Copntce Titles: Ttles Vedl astic gring
Copntce Titles: Colergpion mtaaler
Copntce Titles: Ccterpard: Retler
>
Copnter Titles: Cited Day Peeller ewager
Copntce Titles: Tite. Leersage
Copntce Titles: Totlery badager
Copntce Titles: Tite. Lag. aguer
Copntce Titles: Tittle. siser
Copntce Titles: Tittlag to Canthe Daager
Copntce Titles: Cote. Learninge
Copntce Titles: Titaloackeng tiles
Copnter Titles: Sote. Spæng wable for fetsonaullertiles:
Copntce Tstarepllags
Copntce Witles: Collacy Gonn sase serder and ether
Copntce Titles: Lith. Cedicing the recend oraton:
Copntce Titles: Cote. Spartaoparly Excerssion >=rague. fle theracts)
Pææcer
Copnter Witles: Tlith radical, liade sagrer
Copntce Titles: CotateBedunalarous:
Copntce Titles: Tatlagy Conn. caller to lls, clop)
Copntce Titles: Andf leator
Copntce Titles: Solerss, tn, from Wilees -1000-ticm tattles. Phærcer)
Lopnton Titles: Claies Fato said:
Thome Omethes: Sclodee the-ssor
| = Tondetct - Tflieres:
}
Coprict Pervide lsi Sen for staal)
Inpnter Con sperrats
Copnter Commlindle, stides
Copnter Tittless: Challing Vacte (o lle, læc. 7198)
Copnter Titles: Titterapechnunting berte/Ratles-(21, While Car the nalds)
Sepntce Oitles: Chtenindar-21l has relvas:
(-rititel umponstod)
Inaito Eimorant Whies to packing owest to ain, 3 RetorgyApPley yeth)
Copntce Pernles-Sneetvirginatachie prout.
Snopntce We theract togo
}
```



Legalidad Web Scrapping

¿ES LEGAL EL WEBSCRAPPING?

Tribunal Supremo Español. Demanda de Ryanair contra Atrápalo

Uno de los casos más sonados en España fue el de la demanda que interpuso Ryanair contra Atrápalo y que resolvió finalmente el Tribunal Supremo a favor de Atrápalo.

Ryanair, la aerolínea de vuelos low cost, demandó a Atrápalo, agencia de viajes online, por scrapear sus datos. A través del portal de Atrápalo los usuarios podían realizar una búsqueda de vuelos. Atrápalo buscaba esta información trasladando los términos de la búsqueda al buscador de Ryanair y permitía la contratación. Ryanair demandó a atrápalo calificando a las agencias online de viajes como parásitos del sector.

Sin embargo el Alto Tribunal no estuvo de acuerdo con el demandante y consideró el scraping como una técnica legal conforme dictó en la sentencia Sentencia del 9 de octubre de 2012 número 572/2012.

Legalidad Web Scrapping

¿En qué casos es ilegal el scraping?

Como ya hemos comentado, el scraping no supone una ilegalidad en sí mismo. No obstante su utilización para determinados fines sí que puede suponer una ilegalidad.

Si queremos mantenernos en la legalidad **tenemos que estar atentos y utilizar el scraping de forma ética, mostrando especial interés en no utilizar propiedad intelectual o marcas registradas, no violar derechos de autor, no realizar competencia desleal y cumplir fielmente tanto la LOPD como al nuevo reglamento europeo RGPD. Y, por supuesto, no sobrecargar a los servidores de los sitios scrapeados.**

El objetivo de este post es explicar los conceptos relacionados con la legalidad del scraping a modo general y no puede utilizarse en ningún caso como un asesoramiento personalizado. Aunque este post intenta ser una ayuda, cada caso concreto tiene que estudiarse muy a fondo, ya que no hay unas leyes generales a las que acogerse.

Datos abiertos en España - Referencias

§ Datos.gob.es: el portal oficial del Gobierno de España para el acceso a los datos abiertos del sector público.

§ OpenDataEuskadi: el portal de datos abiertos del Gobierno Vasco, que ofrece datos de diferentes ámbitos como transporte, medio ambiente y estadísticas.

§ OpenDataBarcelona: el portal de datos abiertos del Ayuntamiento de Barcelona, que ofrece información sobre la ciudad en áreas como transporte, cultura, deportes y medio ambiente.

§ OpenDataSevilla: el portal de datos abiertos del Ayuntamiento de Sevilla, que ofrece información sobre la ciudad en áreas como transporte, urbanismo y cultura.

§ OpenDataMálaga: el portal de datos abiertos del Ayuntamiento de Málaga, que ofrece información sobre la ciudad en áreas como transporte, turismo y medio ambiente.

§ OpenDataCanarias: el portal de datos abiertos del Gobierno de Canarias, que ofrece información sobre las islas en áreas como turismo, medio ambiente y transporte.

§ OpenDataAlicante: el portal de datos abiertos del Ayuntamiento de Alicante, que ofrece información sobre la ciudad en áreas como urbanismo, transporte y turismo.

§ DatosAbiertos.gob.es: el portal de datos abiertos del Ministerio de Hacienda y Administraciones Públicas, que ofrece información sobre diferentes ámbitos como economía, industria y empleo.

§ AEMET OpenData: el portal de datos abiertos de la Agencia Estatal de Meteorología, que ofrece información sobre el tiempo y el clima en España.

§ DatosAbiertos.jcyl.es: el portal de datos abiertos de la Junta de Castilla y León, que ofrece información sobre la región en áreas como medio ambiente, turismo y cultura.

Datos abiertos Nivel Global – Referencias

§ Datos abiertos del Banco Mundial: una plataforma que ofrece datos abiertos sobre la economía, el desarrollo y el bienestar en todo el mundo.

§ Datos abiertos de la ONU: una plataforma que ofrece datos abiertos sobre una amplia variedad de temas, como salud, medio ambiente y derechos humanos.

§ Datos abiertos del Fondo Monetario Internacional: una plataforma que ofrece datos abiertos sobre la economía global, incluyendo información sobre el comercio, las finanzas y la deuda.

§ Datos abiertos de la Organización para la Cooperación y el Desarrollo Económico (OCDE): una plataforma que ofrece datos abiertos sobre el desarrollo económico y social en los países miembros de la organización.

§ Datos abiertos de la Unión Europea: una plataforma que ofrece datos abiertos sobre diversos temas, como agricultura, economía y medio ambiente, entre otros.

§ Datos abiertos de la NASA: una plataforma que ofrece datos abiertos sobre la exploración espacial, el clima y la ciencia en general.

§ Datos abiertos de la Agencia de Protección Ambiental de EE. UU.: una plataforma que ofrece datos abiertos sobre el medio ambiente y la salud pública en los Estados Unidos.

§ Datos abiertos de la Biblioteca del Congreso de EE. UU.: una plataforma que ofrece datos abiertos sobre la historia y la cultura de los Estados Unidos.

§ Datos abiertos de la Comisión Europea: una plataforma que ofrece datos abiertos sobre la política y la economía de la Unión Europea.

§ Datos abiertos de la Biblioteca Digital Mundial: una plataforma que ofrece datos abiertos sobre la historia, la cultura y el patrimonio de todo el mundo.