

# Buceando en las cloacas con Python

# Sobre Civio

- Periodismo, incidencia, tecnología y litigación en aras de mayor transparencia
- Algunas cosas periodísticas que se nos dan bien:
  - La cobertura del BOE en [El Boe Nuestro de Cada Día](#)
  - [Temas de contratación](#)
  - Aplicaciones de servicio público ([bono social](#) y [de las ayudas por la crisis del Coronavirus](#) y [el indultómetro](#))

[Premio Rey de España al Medio de Iberoamérica 2022](#) 🎉

# Programación *back-end* en un medio de comunicación

*Scripts* sencillos ejecutados en el terminal para temas rápidos.  
Programas más complejos para temas de más recorrido como los [temas de contratación](#).

- Rascar data de HTML y PDF ([requests.py](#), [bs4.py](#), [selenium.py](#), PyPDF) una sola vez, o durante un período de tiempo concreto.
- Análisis de grandes *datasets* ([pandas.py](#))
- Aplicaciones de apoyo a periodistas

# Web scraping en investigaciones largas en Civio

De 25 céntimos a 8 euros por una mascarilla, la guerra de precios que se saldó con millones de euros para los intermediarios. Datos de la Plataforma de contratación

- 16.589 contratos firmados por todas las administraciones públicas (6.445 millones de euros)
- Algunos ayuntamientos se desembolsaron más de cien euros por una garrafa de cinco litros de gel hidroalcohólico

# Usando Python para rascar datos periódicamente en Civio

## La movilidad europea en tiempos de coronavirus

*Scraping* recurrente de tres fuentes de datos:

- Cámaras de tráfico
- Informes de movilidad de Apple (diario)
- Tomtom (cada hora)

# Investigaciones a corto plazo en El Confidencial (Python)

[Metro inaccesible](#). Datos obtenidos gracias a un raspado diario de la web de [incidencias del metro](#)

- Siete de cada diez paradas de la red son accesibles. La mayoría de ellas están alejadas del centro
- Entre el 3 de junio y el 18 de julio, se detectaron 204 estaciones reportando al menos una incidencia en las escaleras y otras 150 estaciones, en ascensores

## Más público que Netflix y miles de reseñas de puteros: así es el gran barrio rojo 'online'

Datos obtenidos gracias a un rascado de los portales [Slumi](#), [pasion.com](#) y [mileroticos](#).

- En torno al 35% de los teléfonos corresponden a más de un anuncio

¿De qué año es tu casa? El mapa que muestra cómo se ha construido España. Datos obtenidos del catastro y fusionados en una capa GEOJSON mediante `geopandas.py`

- El 17% de construcciones de todas las capitales de provincia data de la década del desarrollismo (años 60)
- Busca la edad del edificio en el que vives



# Aplicaciones de apoyo a periodistas

Aplicación de rascado de preguntas de parlamentarios que consultaba cada día si había nuevas preguntas relacionadas con temáticas o grupos parlamentarios concretos.

- [Los suicidios en la Policía cuadruplican a las muertes en acto de servicio tras el fin de ETA](#)
- [La mayoría de casos de acoso sexual en la Guardia Civil acaba en la basura](#)

Autor	Grupo	Título	Concluido
Rodríguez Rodríguez, Alberto (GCUP-EC-GC)	GCUP-EC-GC	<a href="#">Solicitud de informe al Ministerio de Defensa, recabando el protocolo de intenciones entre el Ministerio de Defensa y el Cabildo de Tenerife para desarrollar la enajenación, desalojo y reubicación de determinadas instalaciones militares ubicadas en la isla de Tenerife. (186/000035)</a>	desde 28/08/2019 hasta 04/09/2019
García Sempere, Eva (GCUP-EC-GC)	GCUP-EC-GC	<a href="#">Solicitud de informe a la Administración del Estado, recabando copia del informe de la Subdirección General de Protección del Patrimonio Histórico del Ministerio de Cultura y Deporte favorable a la celebración en julio de 2019 de una prueba del mundial de BMX en el dique de Gamazo (Santander), espacio declarado Bien de Interés Cultural (BIC). (186/000034)</a>	desde 28/08/2019 hasta 04/09/2019
Oramas González-Moro, Ana María (GMx)	GMx	<a href="#">Solicitud de informe al Ministerio de Asuntos Exteriores, Unión Europea y Cooperación, recabando el Acuerdo de la Unión Europea con Mercosur. (186/000033)</a>	desde 28/08/2019 hasta 04/09/2019
Iñárritu García, Jon (GMx)	GMx	<a href="#">Solicitud de informe al Ministerio del Interior, recabando la instrucción interna de Instituciones Penitenciarias sobre detección de maltrato en las prisiones, de la que informó el diario El País el día 03/07/2019. (186/000026)</a>	desde 31/07/2019 hasta 04/09/2019
Iñárritu García, Jon (GMx)	GMx	<a href="#">Solicitud de informe al Ministerio del Interior, recabando los protocolos de la Guardia Civil sobre uso de drones. (186/000025)</a>	desde 26/07/2019 hasta 04/09/2019

El grupo parlamentario Mixto ha presentado 5 iniciativas nuevas en El Congreso:

pregunta	respuesta	descripcion	semejantes	datos
No esta disponible	<a href="#">respuesta</a>	en primer lugar, se prohíbe la entrada para comunicar al menor, por no tener ni el parentesco suficiente (excede el grado de parentesco permitido en este tipo de comunicaciones, es decir, el cuarto grado), ni la autorización por escrito del cónyuge ausente en la comunicación (madre del menor y pareja ...	1	no
<a href="#">pregunta</a>	<a href="#">respuesta</a>	en España no hay presos políticos. En cuanto a los políticos catalanes presos, se indica su situación depende del Poder Judicial y no del Ejecutivo. Así pues, dado estamos en un Estado de Derecho y el Gobierno respeta, como no puede ser de otra forma, la separación de poderes ...	0	no
No esta disponible	<a href="#">respuesta</a>	por parte de la Comisaría Provincial de Granada no existió orden alguna de reiterada de banderas de Andalucía. Por otro lado, cabe indicar los manifestantes contrarios a la celebración del día de la Toma de Granada por los Reyes Católicos solo portaban banderas independentistas y una bandera de la ...	0	si
<a href="#">pregunta</a>	<a href="#">respuesta</a>	la razón del cese del Cónsul General en Edimburgo, Miguel Ángel Vecino, estriba en la pérdida de confianza del Gobierno en su labor, condición indispensable para el desempeño de un puesto de libre designación como el el interesado ocupaba hasta ahora. El Gobierno no contempla en modo alguno el ...	0	no

# Algunos ejemplos de código

- [Scraper de TomTom](#) y [parseador](#)
- Scraper de Carteles de la Guerra (Portal de Archivos Españoles)
  - `fetch-pages.py`
  - `fetch-posters.py`
  - `create-dataset.py`

## ¡Muchas gracias!

Más temas e investigaciones basadas en datos en [civio.es](https://civio.es)