

Angel Pablo Hinojosa

**Scraping y Obtención de Datos
para Big (y no tan Big) Data
(Parte II)**

Orígenes de Datos

Open Data (y transparencia)

Orígenes de Datos

CKAN, Datasets y APIs

<http://opendata.ugr.es>

(Y licencias)

Orígenes de Datos

Catálogo nacional:

<http://datos.gob.es/catalogo>

Orígenes de Datos

Catálogo europeo:

<http://open-data.europa.eu/es/data/>

Orígenes de Datos

Catálogo USA:

<http://open-data.europa.eu/es/data/>

Orígenes de Datos

Mapa de orígenes:

<http://eip.lcc.uma.es/opendata/>

(poco actualizado)

Web Scraping

“Rascar” datos de Webs

Con sus cuestiones técnicas

Y sus cuestiones Legales

Web Scraping

HTML

La materia de la que están hechas las webs

<http://www.psicobyte.com/html/curso/>

(Tutorial de HTML)

Web Scraping

Import.io

Rudimentario, pero a veces basta

`https://import.io/`

Web Scraping

Scraper (plugin de Chrome)

<http://www.psicobyte.com/html/curso/>

Web Scraping

Práctica de Scraper

`http://osl.ugr.es`

(...y XPath)

Web Scraping

Caso práctico:

¿Buscamos radares?

`http://www.dgt.es/es/el-
trafico/control-de-
velocidad/granada/`

Web Scraping

Usando Google Docs

<https://docs.google.com>

(Google Spreadsheets, concretamente)

Web Scraping

Usando Google Docs (importar feeds)

=IMPORTFEED("URL")

Espera ¿Qué es un "feed"?

Web Scraping

Usando Google Docs (importar feeds)

<http://osl.ugr.es/feed/>

Web Scraping

Usando Google Docs (importar HTML -listas-)

```
=IMPORTHTML(URL,"list",N)
```

Web Scraping

Usando Google Docs (importar HTML -listas-)

<http://www.dmoz.org/Computers/Internet/>

Web Scraping

Usando Google Docs (importar HTML -tablas-)

```
=IMPORTHTML(URL,"table",N)
```

Web Scraping

Usando Google Docs (importar HTML -tablas-)

`http://www.dgt.es/es/el-traffic/control-de-velocidad/granada/`

Web Scraping

Usando Google Docs (importar HTML -tablas-)

<http://www.dgt.es/es/el-traffic/control-de-velocidad/granada/>

Web Scraping

Usando Google Docs (importar XML)

```
=IMPORTXML(URL,"table",N)
```

(en realidad, HTML con XPath)

Web Scraping

Usando Google Docs (importar XML)

```
http://os1.ugr.es
```

```
//h2
```

```
//a/@href
```

```
//h2/a/@href
```

Ejemplos

Qué sabe Google:

<https://www.google.com/settings/account/history>

Ejemplos

Usando datos propios:

<https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments>

Ejemplos

Usando APIs:

http://tecnologia.elpais.com/tecnologia/2016/04/11/actualidad/1460361324_475730.html

<https://github.com/rinze/spotify-moods>

Ejemplos

Usando APIs:

<http://git.io/top-gh>

<https://github.com/JJ/top-github-users-data>

Gracias

(Ruegos y preguntas)