

Taller Web Scrapping con Python

ÀREA HACKERS CÍVICS

Vicent Blanes

01/05/2017

¿Por qué Python?

- Python es un lenguaje **interpretado** y con una gran expresividad.
- Dispone de librerías potentes que nos ahorran mucho trabajo.
- Además es muy limpio y fácil de leer, un lenguaje ideal para aprender a programar.



Herramientas para Web Scrapping

- Utilizaremos la librería **urllib** para conectarnos a una dirección de Internet y obtener su contenido.
- Después utilizaremos **beautifulsoup4** para *parsear* el contenido y poder extraer información de forma sencilla
- También utilizaremos la biblioteca **iPython** para trabajar con notebooks, que nos permite ejecutar fragmentos de código de forma independiente y resulta mucho más interactivo.(?)

- Anaconda es una distribución de python que tiene pre-instaladas montones de librerías útiles
- Su instalación es muy sencilla. Podéis encontrarla en <https://www.continuum.io/downloads>.
- Además posee su propio gestor de paquetes. Si te falta alguna librería puedes usar **conda install** y el nombre de esta.
- Para el uso de los *notebooks* tenemos que ejecutar **conda install ipython**

