

Práctica 1. Web Scraping.

El objetivo de esta actividad será la creación de un dataset a partir de los datos contenidos al web. Tenéis que indicar las siguientes características del dataset general:

1. Título del dataset. **Laptops on sale**
2. Subtítulo del dataset. *“Seguimiento y comparación de precios online”*
3. Contexto. Se pretende obtener una base de datos de los ordenadores de la venta de uno de los mayoristas con mayor presencia en el mercado norteamericano, BestBuy (<https://www.bestbuy.com/>)
4. Contenido. Los campos que se pretenden obtener son marca del ordenador, valoración en la web del vendedor, precio y conjunto de especificaciones técnicas: marca, modelo, procesador, RAM, memoria disco duro, color ¿Cuál es el periodo de tiempo de los datos y cómo se ha recogido? El archivo contiene 24 registros extraídos utilizando Python BeautifulSoup y la última actualización fue: 2017-11-13T20:05:36-0800
5. Agradecimientos. ¿Quién es propietario del conjunto de datos? Los datos pertenecen a: Domain Administrator, ubicado en la ciudad de Richfield MN, USA ofrecidos a través del domino “bestbuy.com”
6. Inspiración. ¿Por qué es interesante este conjunto de datos? ¿Qué preguntas le gustaría responder la comunidad? El interés radica en la posibilidad de seguimiento y comparación de precios en línea.
7. Licencia. Seleccionad una de estas licencias y decid porqué la habéis seleccionado. Seleccionaría, **Released Under CC BY-SA 4.0 License**. Porque permite copiar y redistribuir el material en cualquier medio y formato. Permite además remezclarlo, transformarlo y crear material nuevo para cualquier proposito incluso comercial. Para lo cual habría que nombrar al autor y si el contenido se remezcla, transforma o se crea nuevo material se debe distribuir bajo esta misma licencia
8. Código: En adjunto
9. Dataset: En adjunto

Contexto

Esta base de datos fue obtenida de <https://www.bestbuy.com/> como parte de la practica 1 de la asignatura ‘Tipología y Ciclo de vida de los datos’ del Máster Ciencia de Datos de la Universitat Oberta de Catalunya UOC.

El objetivo de la practica pasa por poner en práctica los conocimientos de web scraping aprendidos en dicha asignatura.

Contenido

El archivo contiene 24 registros extraídos utilizando Python BeautifulSoup, los campos extraídos son marca del ordenador, valoración en la web del vendedor, precio y conjunto de especificaciones técnicas: marca, modelo, procesador, RAM, memoria disco duro y color.