

Dataset:

Extractor de precios y rating de Sensores en comercio Web

Autor: Ramón Martínez García. Alumno de Tipología y Ciclo de Vida de los Datos en la UOC.

Descripción:

Proyecto simple de extracción de información sobre precio de venta y rating de los sensores disponibles en una web concreta.

Imagen.



Imagen de sensores extraida de Sparkfun.com

4. Contexto. ¿Cuál es la materia del conjunto de datos?

Listado de elementos disponibles en la web así como su precio y puntuación.

5. Contenido. ¿Qué campos incluye? ¿Cuál es el periodo de tiempo de los datos y cómo se ha recogido?

Sensor name: nombre del elemento

Sensor price: precio del producto

Sensor rating: rating si es que existe, de lo contrario dispone de un comentario especificando que no existe puntuación aún.

El periodo de tiempo es el del momento de ejecución del script, y la manera en que se ha recogido es simplemente almacenando los datos tras una petición con urllib de Python y posteriormente recorriendo y buscando los tags en los que estamos interesados con la librería BeautifulSoup.

6. Agradecimientos. ¿Quién es propietario del conjunto de datos? Includ citas de investigación o análisis anteriores.

El propietario de los datos es Sparkfun.com

7. Inspiración. ¿Por qué es interesante este conjunto de datos? ¿Qué preguntas le gustaría responder la comunidad?

Mi interés en concreto es el de analizar productos relacionados con el Internet de las cosas, y entre ellos se encuentran los sensores, una aplicación interesante sería buscar entre diversas webs para hacer un posible comparador de precios para adquirir los productos con coste más reducido para proyectos relacionados.

8. Licencia. Seleccionad una de estas licencias y decid porqué la habéis seleccionado: ☐ Released Under CC0: Public Domain License ☐ Released Under CC BY-NC-SA 4.0 License ☐ Released Under CC BY-SA 4.0 License ☐ Database released under Open Database License, individual contents under Database Contents License ☐ Other (specified above) ☐ Unknown License

La licencia escogida es CC-BY para que se haga mención del autor del código.

9. Código: Hay que adjuntar el código con el que habéis generado el dataset, preferiblemente con R o Python.

El código se adjunta en Github.