

Universitat Oberta de Catalunya

Tipología y ciclo de vida de los datos

Práctica 1: web scraping

Sergio Blay
2019/2020

1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.

“Steam¹ es una plataforma de distribución digital de videojuegos desarrollada por Valve Corporation. (...) Steam ofrece gestión digital de derechos, servidores de emparejamiento, transmisiones de vídeo y servicios de redes sociales. También provee al usuario con la instalación y actualización automática de juegos y características de comunidad tales como grupos y listas de amigos, guardado en la nube, voz en el juego y funcionalidad de chat. “

Wikipedia [Consulta: 30 de octubre de 2019] <<https://es.wikipedia.org/wiki/Steam>>

2. Definir un título para el dataset. Elegir un título que sea descriptivo

Precios de los videojuegos de la plataforma Steam

3. Descripción del dataset. Desarrollar una descripción que sea breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).

Al ser la plataforma online más grande del planeta en la que podemos comprar videojuegos, y por sus constantes ofertas y eventos, puede ser interesante hacer un seguimiento de los precios para entre otras cosas:

- Ver las fluctuaciones de precio de un videojuego a través del tiempo.
- Predecir precios.
- Evitar ser engañados con subidas de precio previas a ofertas como sucede en algunos establecimientos.
- Poder adelantarnos a Steam en caso de ser competencia.

¹ <https://store.steampowered.com/>

4. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente.



Figura 1: The Witcher 3: Wild Hunt – Hearts of Stone en Steam²

² https://store.steampowered.com/app/378649/The_Witcher_3_Wild_Hunt_Hearts_of_Stone/

5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

Para cada videojuego se recogen las siguientes características:

- ID: ID que utiliza Steam para identificar el producto.
- Date: fecha de recogida de los datos. Formato dd-mm-aaaa HH:MM.
- Name: nombre del videojuego.
- Total Price: precio total en euros, sin ningún tipo de descuento.
- Current Price: precio en euros al que se encuentra el videojuego en el momento de recogida de los datos.
- Image path: ruta relativa en la que se ha almacenado la imagen correspondiente al videojuego.

El periodo de tiempo es desde la primera ejecución del script asociado a esta memoria. Se han recogido desde la página web de Steam anteriormente mencionada mediante web scraping.

6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).

El propietario de los datos es *Valve Corporation*, entidad de la que se quiere estudiar la evolución de sus precios y agradecemos sus datos.

7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.

Según un informe elaborado en España en 2017 llamado Libro blanco del desarrollo español de videojuegos 2017:

“El videojuego se ha convertido en el principal motor del entretenimiento global. Según la consultora Newzoo, la industria global creció en 2017 el 10,7 % y fue capaz de generar 116.000 millones de dólares a nivel mundial, de los cuales el 87% corresponden a los modelos de negocios digitales. Hoy en día, hay 2.200 millones de jugadores (...). De ellos, mil millones (el 47 %), gastan dinero mientras juegan. El mercado global de videojuegos seguirá creciendo a una tasa anual del 8,2 %, alcanzando los 143.500 millones de dólares en 2020”. (DEV, 2017)

Es debido este auge y a que Steam es la mayor plataforma digital de videojuegos del planeta, por lo que es interesante poder realizar un seguimiento que beneficie tanto a compradores que rebajas o predicciones pueden esperarse de un determinado producto, como a plataformas menores que quieran hacerse un sitio en el mercado.

Dependiendo de a quién beneficie este rastreo de los datos, podemos resolver preguntas como: ¿cuál es el precio mínimo que puedo esperar de un determinado producto? ¿esperar x tiempo

antes de comprar me beneficiará? ¿rebajando este producto puedo mejorar la oferta de Steam y/o anticiparme a ellas?

8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección.

Se ha escogido la licencia **CC BY-SA 4.0 License** debido a que se busca el libre uso del dataset para cualquier persona o entidad, siempre y cuando se reconozca la autoría y se publique bajo la misma licencia para fomentar el uso de este formato.

Recursos

1. Desarrollo Español de Videojuegos (2017). Libro blanco del desarrollo español de videojuegos.
<http://www.dev.org.es/images/stories/docs/libro%20blanco%20dev%202017.pdf>