

Dataset: Gunplas a la venta en la web Hobby Search Gundam

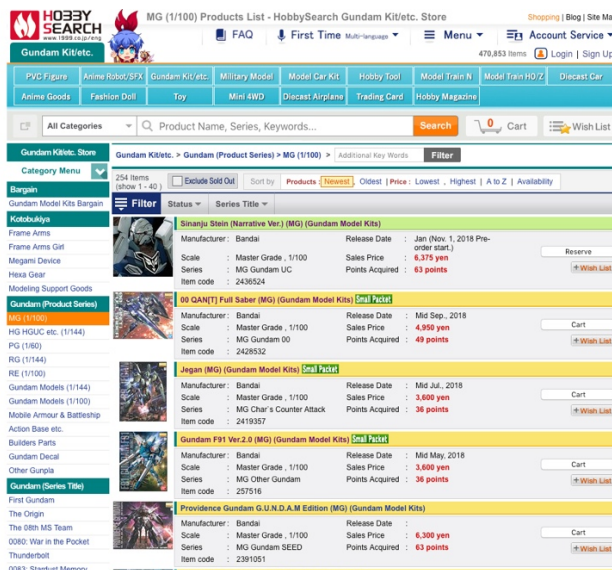
Luis Manuel Martín Guerra

11 de Noviembre de 2018

Descripción

El conjunto de datos resultado de la ejecución del script que comprende esta práctica comprende las principales características de todas las maquetas de la franquicia de animación japonesa, “*Gundam*” fabricadas y puestas a la venta en los últimos 25 años.

Imagen identificativa



1: Captura de pantalla de la página principal de búsqueda de la página web Hobby Search. (2018)

Contexto

Las maquetas plástico de la franquicia de animación japonesa “*Gundam*™”, también conocidos como “Gunplas”, acrónimo de “Gundam Plastic Models”, son una de mis grandes pasiones y aficiones. Al igual que yo, miles de aficionados por todo el mundo disfrutan montando y coleccionando estas maquetas. Sin embargo, resulta complicado tener una base de datos actualizada sobre los Gunplas fabricados, los diferentes tipos de escalas y por supuesto, comparativa entre precios y tamaños.

El siguiente conjunto de datos recoge las principales características de los Gunplas a la venta en la web de *Hobby Search*™, una web japonesa, que actualmente es el referente en los Gunplas en cuanto a cantidad, variedad y precio. Por lo tanto, en el dataset encontraremos información sobre las diferentes escalas en las que se ofrecen, las franquicias de anime en el que están basadas, e incluso el precio de venta recomendado.

Contenido

En este dataset, cada Gunpla se corresponde con un registro o fila, en el que se recogen las siguientes características o atributos:

- Nombre: nombre de la maqueta.
- Fabricante: compañía que ha fabricado la maqueta.
- Escala: escala a la que está realizada la maqueta.
- Serie: franquicia de Gundam en la que aparece la maqueta.
- Original: franquicia base sobre la que se ha inspirado la maqueta.
- Fecha de lanzamiento: mes y año en el que se puso a la venta por primera vez.
- Precio venta: precio en dólares(USD) por el que podemos encontrarla a la venta.
- Precio especial: precio en dólares (USD) por el que lo vende Hobby Search™.
- Código JAN: código de producto identificativo japonés, para verificar su identidad.
- Embalaje: dimensiones de la caja en la que viene el producto.
- Peso: peso en gramos del producto, incluyendo la caja y las instrucciones.

Hobby Search™ lleva desde 1999 vendiendo maquetas y material de modelismo y hoy en día es uno de los sitios de referencia a la hora de consultar “Gunplas” en la red, tanto por sus precios competitivos, como por la cantidad de referencias de las que disponen.

Los datos recogidos en este dataset tiene como fecha de última extracción el 11 de noviembre de 2018.

Agradecimientos

El conjunto de datos que compone este dataset, así como la fotografía que ilustra el mismo son propiedad en exclusiva de HOBBY SEARCH, y se han empleado únicamente con carácter académico y sin ningún propósito de ánimo de lucro.

Antes de llevar a cabo cualquier tipo de prueba sobre los datos propiedad de HOBBY SEARCH, se ha consultado su mapa del sitio y su fichero robots.txt para estar seguros de que la práctica de este data scraping pudiera llevarse a cabo.

Inspiración

La inspiración para llevar a cabo este ejercicio práctico es la de conocer los diferentes modelos y escalas en los que se han producido y, se van produciendo maquetas de Gundam a lo largo de los años. Un estudio interesante sería el de intentar catalogar si hay algún tipo de relación entre el precio de las maquetas y su antigüedad, o cuales son las franquicias de anime más populares en función del número de maquetas que están basadas en ella.

Finalmente el poder averiguar el precio aproximado de una maqueta en función de sus posibles características (escala, precio, etc.) también sería interesante.

Licencia

Dado el carácter privado de los datos, y debido a que el dataset generado es resultado de un ejercicio académico, la licencia seleccionada para este dataset es la:

(CC BY-NC-ND 4.0)

Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional

- Para la protección de la propiedad intelectual e industrial de los datos, se prohíbe el uso comercial de este dataset.
- Para evitar que terceros hagan un uso del script, para la obtención de los datos, de manera que puedan vulnerar la propiedad sobre los mismos de Hobby Search, se prohíbe que se compartan las adaptaciones de la obra.

Por lo tanto, el dataset generado, como los scripts empleados tienen carácter puramente pedagógico y académico, no planteándose ni pretendiendo usar el contenido del dataset o el código generado para fines lucrativos o comerciales.

Código fuente y dataset

El código fuente de los scripts empleados para la extracción de datos, así como el dataset generado, se pueden acceder a través del enlace al repositorio de github en el que se encuentran. Dado el carácter de los datos, este repositorio es privado y únicamente los profesores de la asignatura Tipología y ciclo de vida de los datos tendrán acceso.

Recursos

1. Lawson, R. (2015). *Web Scraping with Python*. Packt Publishing Ltd. Chapter 2. Scraping the Data.
2. Mitchel, R. (2018). *Web Scraping with Python: Collecting Data from the Modern Web (2nd Edition)*. O'Reilly Media, Inc. Chapter 1. Your First Web Scraper.
3. Subirats, L., Calvo, M. (2018). *Web Scraping*. Apuntes de la asignatura Tipología y ciclo de vida de los datos. Universitat Oberta de Catalunya.
4. Lutz, M. (2016). *Learning Python (5th Edition)*. O'Reilly Media, Inc.