

PRÁCTICA 1: WEB SCRAPING

1. Título: “Web scraping Fotocasa”

2. Descripción.

Extrae los precios y características de los diferentes inmuebles en venta de la página web <https://www.fotocasa.es/es/>.

También se extrae la imagen principal del inmueble y se guarda en una carpeta.

3. Imagen.

The screenshot displays the Fotocasa website interface. At the top, there's a navigation bar with the Fotocasa logo, language selection (Español), user account options (Profesional, Mi cuenta), and a button to 'Publica tu anuncio gratis'. Below this, the location 'Lugo' is selected, showing '3.047 Casas de compra en Lugo Provincia'. A search bar contains 'Lugo Provincia' and a 'BUSCAR' button. On the left, there are filters for 'Tipo de vivienda' (piso, apartamento, chalet, etc.) and 'Precio' (Precio desde, Precio hasta). The main area shows three property listings:

- Piso en Centro Finca Palm...**: 52.508 €, 3 habs, 82 m². Includes a photo of a modern building.
- Dúplex en Carril Das Flores**: 155.000 €, 3 habs, 102 m². Includes a photo of a bedroom.
- Piso en Calle Rosalía de C...**: 45.160 €, 1 hab, 48 m². Includes a photo of a multi-story building.

Each listing includes contact information and a 'Contactar' button. The website also features a sidebar with a 'Recibir alertas' button and a 'Ver las promociones de obra nueva en esta ubicación' link.

4. Contexto.

La materia de datos recogidos hace referencia a las características de los inmuebles a la venta en Lugo capital. También se guarda la imagen principal de cada inmueble en una carpeta con el mismo id que el inmueble.

5. Contenido.

Se recogen todos los datos de los anuncios disponibles hasta la fecha. Cuando un inmueble se vende su anuncio se elimina, por lo que no tiene cabida buscar en un rango de fechas.

Se incluyen los siguientes campos:

- Title: título del anuncio.
- Description: descripción del anuncio.
- Price: precio en euros del inmueble.
- hab: número de habitaciones.
- m2: tamaño en m2.
- Timeago: hace cuánto fue publicado.
- Phone: teléfono de contacto.
- url: página del anuncio.
- imgurl: url de la imagen del anuncio.

6. Agradecimientos.

El propietario del conjunto de datos es Fotocasa.

7. Inspiración.

La búsqueda de piso es siempre bastante tediosa. Este web scraping permite tener la información más relevante de manera más sencilla y rápida. Aunque se ha partido de la búsqueda de los inmuebles en Lugo capital, se puede seleccionar otra ciudad y posteriormente copiar la url en el código antes de ejecutarlo.

8. Licencia.

Se ha elegido una licencia pública (Released Under CC0: Public Domain License) ya que los datos son de ámbito público y se quiere que cualquier usuario pueda usarlos.

9. Código.

El dataset ha sido generado con Python usando el módulo BeautifulSoup. El nombre del fichero que contiene el código es 'Fotocasa_scraping.py'.

10. Dataset.

Se ha usado el módulo de pandas de Python para generar el dataset final y se ha guardado como un csv (Fotocasa_scraping.csv).

Las imágenes se guardan en una carpeta llamada 'pictures' que se genera de manera automática en el cwd del usuario que ejecute el código.

La práctica ha sido realizada de manera individual por Eva M^a Ambroa Rey.