

# M2.851 Tipología y Ciclo de Vida de los Datos

## Práctica 1: Web Scraping

### 1. Título del dataset. Poned un título que sea descriptivo.

ANEL. Empresas de Economía Social de Navarra.

### 2. Subtítulo del dataset. Agregad una descripción ágil de vuestro conjunto de datos por vuestro subtítulo

Listado de empresas de economía social creadas en Navarra.

### 3. Imagen. Agregad una imagen que identifique vuestro dataset visualmente



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

La economía social está creciendo de forma exponencial en los últimos años, tanto en la creación de nuevas empresas como en los trabajadores que las componen.

La página web de esta asociación, a la cual pertenecemos, ofrece noticias, publicaciones, formación, documentación y un listado de las empresas que forman la asociación. Debido al auge de este tipo de empresas, la información de esta web ([www.anel.es](http://www.anel.es)) es cada vez más dinámica.

El ejercicio se ha basado en una de las secciones, llamada *Empresas Asociadas*, donde se muestra una lista de las empresas que componen la asociación.

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

La lista de empresas tiene una estructura que hemos creído útil para la realización del ejercicio de web scraping.

En primer lugar, indicar que al pulsar en *Empresas Asociadas* (<http://www.anel.es/category/solo-para-empresas/empresas-asociadas/>), no aparecen todas, ya que no caben. En la parte inferior aparece un selector para cambiar de página y poder así visualizar todas. Lo primero que hemos hecho pues es descubrir todos estos links. Luego uno a uno hemos descargado los enlaces, y hemos extraído la información de cada empresa:

- Fecha de la inserción de la noticia
- Nombre de la empresa creada
- Dirección de la empresa creada
- Ciudad de la empresa creada
- Dirección web de la empresa creada
- Correo electrónico de contacto
- Imagen o logotipo. Nombre de la imagen que hemos descargado.

No todas tienen todos los campos. La dirección web y el logotipo no aparecen en todas.

En principio, esta lista irá creciendo con las nuevas incorporaciones, aunque es lógico pensar que no se cambiará frecuentemente. Eso sí, los datos incorporados no tienen caducidad, a no ser que haya cambios o bajas

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

Es ANEL quien difunde estos datos con consentimiento de las empresas. Es más, esta página web pretende ser un escaparte, un medio de difusión para las empresas asociadas.

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

La inspiración en recolectar estos datos viene dada por una inquietud real. Como introducía estamos asociados, en esta época de emprendimiento que estamos llevando a cabo, a ANEL. Y una de las actividades que llevamos a cabo en la asociación son reuniones de las empresas que estamos empezando para, bajo un marco colaborativo, ayudarnos y compartir experiencias.

En estas reuniones, hay un tema recurrente, que es ver cómo interactuar más con las empresas antiguas.

El recopilar este listado, de forma periódica, nos podrá ayudar en este proceso colaborativo entre las nuevas empresas para conocer a las antiguas.

## 8. Licencia. Seleccionad una de estas licencias y decid porqué la habéis seleccionado

Los datos son públicos, y pretender ser una herramienta de difusión para las empresas de la asociación. Por tanto seleccionamos una licencia de dominio público.

## 9. Código: Hay que adjuntar el código con el que habéis generado el dataset, preferiblemente con R o Python, que os ha ayudado a generar el dataset.

Se adjunta a este documento el archivo PRACTICA1.py con el código en lenguaje python

## 10. Dataset

Se adjunta a este documento el archivo ANEL.CSV con el resultado de la ejecución