

Práctica 1 (35 % nota final)

El enlace a github de esta práctica es: https://github.com/idemas001/PRA1_JdM

Componentes del grupo

Esta PRAC1 ha sido realizada de forma individual por Jordi de Mas, debido a la dificultad para poder encontrar franjas horarias compatibles con otros compañeros de asignatura a causa de mis obligaciones laborales. El hecho de que nuestra profesora enviara una comunicación indicando que se podía hacer individualmente, ha facilitado mucho encontrar una solución al problema.

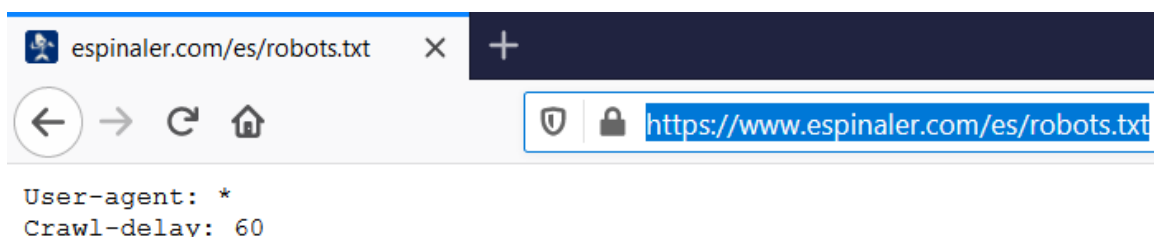
Descripción de la Práctica a realizar

El objetivo de esta actividad será la creación de un dataset a partir de los datos contenidos en una web. Para su realización, se deben cumplir los siguientes puntos:

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

Se pretende recoger el catálogo de productos ofrecido por la marca Espinaler. Se ha elegido extraerlo de la propia web de la empresa ya que contiene la mencionada información y el fichero "robots.txt" correspondiente a esta página web no incluye restricciones para realizar web scraping.

Veamos el contenido del mencionado fichero:



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

Catálogo de productos de conservas de pescado y marisco de la marca Espinaler.

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

El dataset incluye el catálogo de productos de la marca Espinaler correspondiente a las conservas de pescado y marisco.

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



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

El dataset incluye las referencias correspondientes a las conservas de pescado y marisco. El período de tiempo aplicado corresponde al catálogo actual vigente en el momento de la extracción. Se ha recogido mediante su página web <https://www.espinaler.com/es/productos/>

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 la empresa Espinaler. Se trata de una compañía centenaria (creada en 1896) especializada en la selección, fabricación y distribución de conservas de pescado y marisco gourmets. Sus clientes principales son tiendas, restaurantes y vermuterías gourmet, y establecimientos especializados. Su producto más famoso es la Salsa Espinaler, elaborada a partir de una receta propia ideada en 1950.

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

Este conjunto de datos es interesante para aquellas personas, profesionales o no de la cocina, que deseen conocer los productos gourmet de esta marca, sabiendo que se trata de productos de una empresa con un fuerte compromiso medioambiental. Como sabemos, la concienciación ecológica está cada vez más arraigada en todos nuestros actos cotidianos y saber que se dispone de productos que cumplen con estos preceptos, nos permite a los consumidores adquirir productos que son elaborados conforme a estos ideales.

Espinaler es una compañía totalmente respetuosa con el medio ambiente. Tiene un compromiso explícito en sólo utilizar materias primas que procedan de un aprovechamiento respetuoso con el medio ambiente. Los principios incluidos en el compromiso sostenible de la empresa son:

- La adquisición de materias primas que procedan de pesquerías o cultivos marinos sostenibles y responsables.
- La protección del medio ambiente marino, la conservación de sus recursos y de la biodiversidad marina.
- La transparencia y buenas prácticas en procesos, compromisos y acciones.

En su compromiso medioambiental también se debe remarcar que sus conservas incluyen el certificado “Dolphin Safe”, lo que significa que los artes de pesca utilizados en la captura de los túnidos protegen a los delfines.

Y por último, indicar que todos los estuches y latas de todos los productos de Espinaler, son 100% reciclables, contribuyendo así a disminuir el impacto medioambiental.

Con la información extraída se podría responder a la pregunta de si existen productos gourmet respetuosos con el medioambiente. También, resultaría interesante analizar las opiniones de los clientes expresadas en las redes sociales, con el fin de comprobar la satisfacción que producen en los consumidores. En el caso de poseer un negocio de restauración, dichas opiniones nos ayudarían a formar una opinión del producto y, podríamos decidir incluir a Espinaler como uno de nuestros proveedores.

Las opiniones expresadas por los consumidores en las redes sociales podrían ser analizadas mediante análisis de sentimientos, para evaluar la polaridad y subjetividad y así detectar opiniones fraudulentas.

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

- ☐ **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 para la publicación del dataset es “Released Under CC BY-NC-SA 4.0 License”, ya que tiene las cláusulas más adecuadas en relación al trabajo que se ha efectuado:

- Se debe reconocer y citar la obra de la forma especificada por el autor. Esta es una manera de reconocer el trabajo ajeno y las aportaciones realizadas respecto al trabajo original.
- No se permite un uso comercial puesto que se trata de información de productos de una empresa. El dataset generado sólo es para fines informativos. De todos modos, está permitido hacer obras derivadas.
- Se permite la distribución de obras derivadas bajo una licencia idéntica a la licencia que regula la obra original. Esto significa que el trabajo del autor original se sigue distribuyendo bajo los términos que él mismo planteó.

9. Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

```
# -*- coding: utf-8 -*-
"""
```

Editor de Spyder

```
Este es el archivo PRA1_JdM_code.
"""
```

```
# Importamos librerías Python que usaremos.
```

```
import requests
```

```
import csv
```

```
from bs4 import BeautifulSoup
```

```
# Creamos tabla para guardar las url's a acceder
```

```
web_pages = []
```

```
# Grabamos en la tabla la url de las webs que deseamos acceder.
```

```
url_page = 'https://www.espalaler.com/es/producto-espalaler/almeja-blanca'
```

```
web_pages.append(url_page)
```

```
url_page = 'https://www.espalaler.com/es/producto-espalaler/almeja-extra'
```

```
web_pages.append(url_page)
```

```
url_page = 'https://www.espalaler.com/es/producto-espalaler/almeja-natural'
```

```
web_pages.append(url_page)
```

```
url_page = 'https://www.espalaler.com/es/producto-espalaler/atun'
```

```
web_pages.append(url_page)
```

```
url_page = 'https://www.espalaler.com/es/producto-espalaler/berberecho'
```

```
web_pages.append(url_page)
```

```
url_page = 'https://www.espalaler.com/es/producto-espalaler/bonito'
```

```
web_pages.append(url_page)
```

```
url_page = 'https://www.espalaler.com/es/producto-espalaler/caracol-de-mar'
```

```
web_pages.append(url_page)
```

```
url_page = 'https://www.espalaler.com/es/producto-espalaler/chipiron'
```

```
web_pages.append(url_page)
```

```
url_page = 'https://www.espalaler.com/es/producto-espalaler/erizo'
```

```
web_pages.append(url_page)
```

```
url_page = 'https://www.espalaler.com/es/producto-espalaler/langostillo'
```

```
web_pages.append(url_page)
```

```
url_page = 'https://www.espalaler.com/es/producto-espalaler/mejillon'
web_pages.append(url_page)
url_page = 'https://www.espalaler.com/es/producto-espalaler/navaja'
web_pages.append(url_page)
url_page = 'https://www.espalaler.com/es/producto-espalaler/pulpo'
web_pages.append(url_page)
url_page = 'https://www.espalaler.com/es/producto-espalaler/rodaja-de-pulpo'
web_pages.append(url_page)
url_page = 'https://www.espalaler.com/es/producto-espalaler/sardina'
web_pages.append(url_page)
url_page = 'https://www.espalaler.com/es/producto-espalaler/sardinilla'
web_pages.append(url_page)
url_page = 'https://www.espalaler.com/es/producto-espalaler/zamburina'
web_pages.append(url_page)

cabecera = "Producto"
catalog = []
headerCatalog = cabecera
catalog.append(headerCatalog)

for a in web_pages:
    # Hacemos el request a la página web.
    page = requests.get(a).text
    # Procesamos el HTML mediante un objeto BeautifulSoup
    soup = BeautifulSoup(page, 'html.parser')
    for text in soup.find_all('h4'):
        texto = list(text)
        print(texto[0])
        catalog.append(texto[0])

# Abrimos el fichero y grabamos la información.
with open('PRA1_JdM.csv', 'w', newline = '') as csv_file:
    writer = csv.writer(csv_file)
    for text in catalog:
        writer.writerow([text])
```

10. Dataset. Presentar el dataset en formato CSV

Ver el fichero que se entrega en el portal.
A continuación mostramos el contenido del mismo:

Ref. AB ESPINALER OL-120 6/8
Ref. AB ESPINALER OL-120 8/10
Ref. AB ESPINALER OL-120 10/12
Ref. AB ESPINALER OL-120 14/16
Ref. AB ESPINALER OL-120 18/20
Ref. AB ESPINALER OL-120 18-20
Ref. AB ESPINALER OL-120 20/25
Ref. AB ESPINALER OL-120 25-30
Ref. AB ESPINALER RO-280 14/16
Ref. AB ESPINALER RO-280 20
Ref. AB ESPINALER RO-280 30
Ref. AB ESPINALER RO-280 40
Ref. ALMEJA ESPINALER 11/15 RR-125
Ref. ALMEJA ESPINALER 15/20 RR-125
Ref. ALMEJA NATURAL ESPINALER OL-120
Ref. ATUN ESPINALER OL-120 VENTRESCA TROZOS
Ref. ATUN ESPINALER OL-120 VENTRESCA
Ref. ATUN ESPINALER OL-120 ACEITE
Ref. BERBERECHOS ESPINALER RO-120 20/25 E
Ref. BERBERECHOS ESPINALER OL-120 20/25
Ref. BERBERECHOS ESPINALER OL-120 25/30
Ref. BERBERECHOS ESPINALER OL-120 30/40 E
Ref. BERBERECHOS ESPINALER OL-120 25/35
Ref. BERBERECHOS ESPINALER OL-120 35/45
Ref. BERBERECHOS ESPINALER OL-120 45/55
Ref. BERBE ESPINALER OL-120 55/65
Ref. BERBERECHOS ESPINALER OL-120 65/85
Ref. BERBERECHOS PEQUEÑOS ESPINALER OL-120
Ref. BERBERECHOS ESPINALER RO-280 60/70
Ref. BERBERECHOS ESPINALER RO-280 80/90
Ref. BONITO ESPINALER OLIVA OL-120
Ref. BONITO ESPINALER OL-120 ESCABECHE
Ref. VENTRESCA BONITO ESPINALER OL-120 TROZOS
Ref. BONITO ESPINALER OL-120 VENTRESCA ESCABECHE
Ref. BONITO ESPINALER OL-120 VENTRESCA
Ref. BONITO ESPINALER RO-280 VENTRESCA

Ref. BONITO ESPINALER SALSA CATALANA T-200G
Ref. BONITO ESPINALER OLIVA T-200G
Ref. VENTRESCA BONITO ESPINALER T-260g
Ref. MIGAS DE BONITO ESPINALER T-260g
Ref. CARACOL ESPINALER OL-120
Ref. CHIPIRONES ESPINALER OL-120 TINTA
Ref. CHIPIRÓN ESPINALER OL-120 4/6 OLIVA
Ref. CHIPIRONES OL-120 5/8 OLIVA
Ref. CHIPIRÓN ESPINALER OL-120 15/20
Ref. CHIPIRÓN ESPINALER OL-120 10/12 OLIVA
Ref. CHIPIRÓN ESPINALER RO-280 12/14
Ref. HUEVAS DE ERIZO DE MAR OL-120
Ref. LANGOSTILLO ESPINALER OL-120
Ref. MEJI ESPINALER OL-120 4/6
Ref. MEJI ESPINALER OL-120 6/8
Ref. MEJI ESPINALER OL-120 6/8
Ref. MEJI ESPINALER OL-120 7/10
Ref. MEJI ESPINALER OL-120 10/12
Ref. MEJI ESPINALER OL-120 14/16
Ref. MEJI ESPINALER OL-120 20-30
Ref. MEJI ESPINALER OL-120 PICANTE
Ref. MEJI ESPINALER OL-120 GALLEGA
Ref. MEJI ESPINALER RO-120 4/6
Ref. MEJI ESPINALER RO-120 6/8
Ref. MEJI ESPINALER RO-280 25
Ref. MEJI ESPINALER OL-120 12/16
Ref. MEJI ESPINALER RO-280 20
Ref. MEJI ESPINALER RO-280 30
Ref. MEJI ESPINALER OL-120 4/6
Ref. MEJI ESPINALER RO-280 14/16
Ref. MEJI ESPINALER OL-120 AL NATURAL
Ref. NAV ESPINALER RR-125 3/5 SALVORA
Ref. NAV ESPINALER RR-125 5/7 SALVORA
Ref. NAV ESPINALER OL-120 4/6
Ref. NAV ESPINALER OL-120 5/8
Ref. NAV ESPINALER OL-120 OLIVA
Ref. NAV ESPINALER OL-120 3/5
Ref. NAV ESPINALER RO-280 14/16
Ref. NAV ESPINALER RO-280 6/8

Ref. NAV ESPINALER RO-280 8/10
 Ref. PULPO ESPINALER OL-120 OLIVA
 Ref. PULPO ESPINALER GALLEGA OL-120
 Ref. RODAJAS ESPINALER OL-120
 Ref. SARDINAS ESPINALER OLIVA RR-125 3/5
 Ref. SARDINAS ESPINALER ESCABECHE RR-125 3/5
 Ref. SARDINAS ESPINALER TOMATE RR-125 3/5
 Ref. SARDINAS ESPINALER PICANTES RR-125 3/5
 Ref. SARDINILLA ESPINALER RR-125 10/12 OLIVA
 Ref. SARDINILLA ESPINALER RR-125 10/12 PICANTE
 Ref. SARDINILLA ESPINALER RR-125 14/16 OLIVA
 Ref. SARDINILLA ESPINALER RR-125 16/20 OLIVA
 Ref. SARDINILLA ESPINALER RR-125 20/25 OLIVA
 Ref. SARDINILLA ESPI RR-125 30/40 OLIVA
 Ref. ZAMBURIÑAS ESPINALER OL-120

Tabla de contribuciones

Al ser un trabajo individual todas las contribuciones tienen un único participante.

Contribuciones	Firma
Investigación previa	JdM
Redacción de las respuestas	JdM
Desarrollo código	JdM

Formato y fecha de entrega

Durante la semana del 28 de octubre, el grupo podrá entregar al profesor una entrega parcial opcional. Esta entrega parcial es muy recomendable para recibir asesoramiento sobre la práctica y verificar que la dirección tomada es la correcta. Se entregarán comentarios a los estudiantes que hayan efectuado la entrega parcial pero no contará para la nota de la práctica. En la entrega parcial los estudiantes deberán entregar por

correo electrónico (lsubirats@uoc.edu) el enlace al repositorio Github con lo que hayan avanzado.

En referente a la entrega final, hay que entregar un único fichero que contenga el enlace a Github donde haya:

1. Una Wiki donde estén los nombres de los componentes del grupo y una descripción de los ficheros.
2. Un documento PDF con las respuestas a las preguntas y los nombres de los componentes del grupo. Además, al final del documento, debe aparecer la siguiente tabla de contribuciones al trabajo, la cual debe firmar cada integrante del grupo con sus iniciales. Las iniciales representan la confirmación por parte del grupo que el integrante ha participado en dicho apartado. Todos los integrantes deben participar en cada apartado, por lo que, idealmente, los apartados deberían estar firmados por todos los integrantes.

Contribuciones	Firma
Investigación previa	Integrante 1, Integrante 2, ...
Redacción de las respuestas	Integrante 1, Integrante 2, ...
Desarrollo código	Integrante 1, Integrante 2, ...

3. Una carpeta con el código Python o R generado para obtener los datos.
4. El fichero CSV con los datos.

Este documento de la entrega final se tiene que entregar en el espacio de Entrega y Registro de AC del aula antes de las **23:59 del día 11 de noviembre**. No se aceptarán entregas fuera de plazo.