

Asignatura: **M2.251 TIPOLOGÍA y CICLO DE VIDA DE LOS DATOS**

Alumnos: **TORRES BOLIVAR, Leidy Liliana**

SOLA VERDÚ, Jose Carlos

Fecha: **09/11/2020**

Dataset: IPHONE 11

Contexto

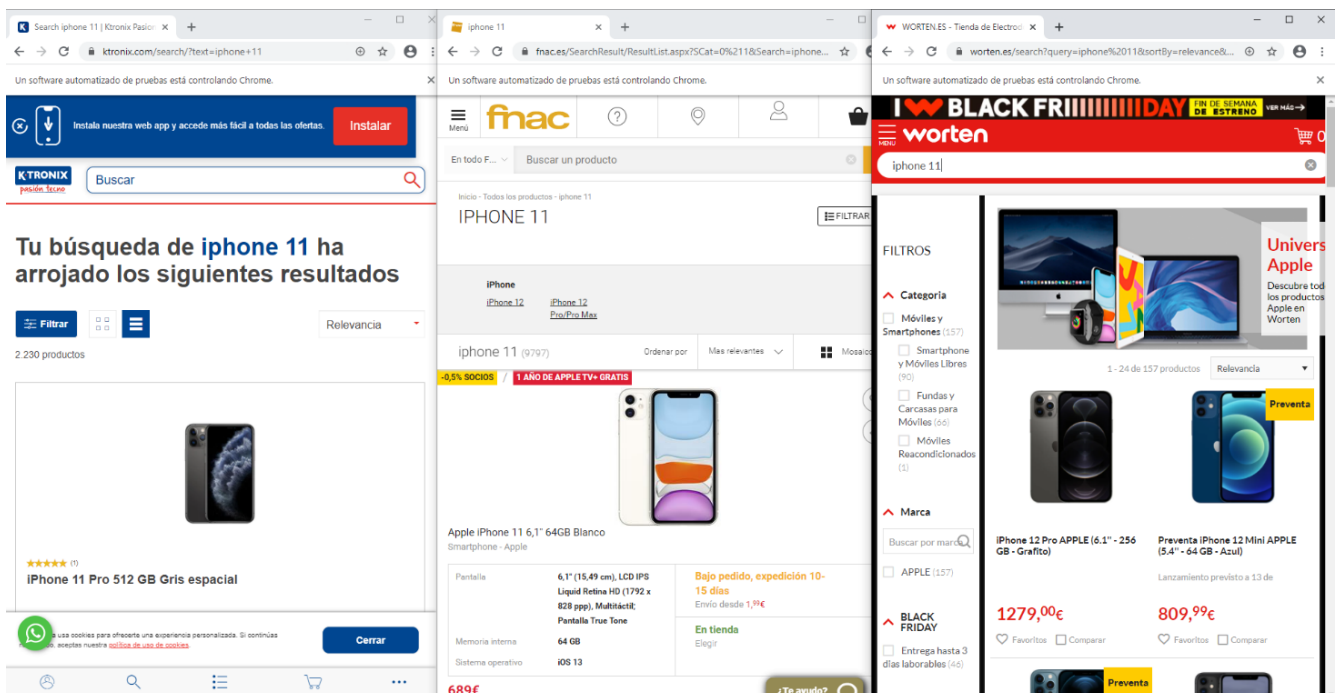
Se han seleccionado distintas webs de Colombia y España para extraer los precios y características de los iPhone 11, con el fin de compararlos de una forma rápida y ver en cuales tiendas de tecnología son más baratos y mejor equipados.

La metodología empleada para la extracción de los datos, es muy interesante porque se puede extrapolar a la búsqueda de otros dispositivos, como ordenadores portátiles, *tablets*, etc. Por tanto, se considera que elaborar un *dataset* que recopile información sobre dispositivos electrónicos puede ser muy interesante porque permite comparar productos de diferentes webs, y conocer en qué precios se está moviendo ese dispositivo, ya sea en Colombia o España.

Descripción del Dataset

El conjunto de datos generado como resultado del desarrollo, agrupa las características y precios de los *smartphones* de la marca "Apple", los iPhone 11, en webs de compra de tecnología, tanto en Colombia como de España. Se recogen las características de los teléfonos, los precios y en qué tiendas se encuentran.

Representación Gráfica



Contenido del *Dataset*

Para cada dispositivo encontrado, el cual se corresponde con un registro en el *Dataset*, se recogen las siguientes características:

- Nombre: nombre del modelo encontrado.
- Val o Precio: precio del dispositivo.
- Característica: recoge las características del modelo, como el almacenamiento, las pulgadas de pantalla, color, etc.
- Tienda: tienda en la que se encuentra el dispositivo.

Agradecimientos

Los datos han sido extraídos de las webs: Ktronix (<https://www.ktronix.com/>), en Colombia, y FNAC (<https://www.fnac.es/>) y Worten (<https://www.worten.es/>) en España.

Inspiración

El conjunto de datos obtenido puede aplicarse en diversos ámbitos. Por un lado, permite conocer en qué precios están oscilando esos dispositivos en el ámbito nacional e internacional, lo que puede ser muy interesante para una nueva tienda *online* que quiere poner precios muy competitivos en el mercado, obteniendo ventaja respecto a su competencia.

Por otro lado, éste conjunto de datos podría ser de gran utilidad para la elaboración de tendencias en la tecnología, en lo que a la venta de dispositivos se refiere. Ya que se podría recoger qué colores se venden más, cuántos GB de almacenamiento busca más la gente, cuánto está dispuesto a gastarse un cliente en la compra de su nuevo dispositivo, etc.

Licencia

La licencia escogida sería la licencia “**Released Under CC BY-SA4.0 License**”, ya que se considera que se puede hacer uso comercial de este desarrollo, que permita el análisis de diferentes webs en busca de los mejores precios en los modelos que una persona quiera comparar, sin la necesidad de ir web a web buscando ese dispositivo.

Las cláusulas detalladas de esta licencia son:

- ❖ El usuario es libre de **compartir**, copiar y redistribuir el material en cualquier medio o formato.
- ❖ Puede **adaptar** y/o transformar el *dataset* a sus necesidades.

Bajo las siguientes condiciones:

- ❖ **Reconocimiento:** El usuario debe reconocer adecuadamente la auditoría, proporcionando un enlace a la licencia e indicar los cambios realizados. Puede realizarlo de forma razonable, pero en ningún caso de forma que parezca que cuenta con el apoyo del licenciador.
- ❖ **Compartir Igual:** Los cambios que realice, deberá difundirlos bajo la misma licencia que la original.

Código fuente y *dataset*

El código fuente y el dataset generado pueden consultarse a través del siguiente enlace: <https://github.com/lilitorres/Practica1WebScraping>

Se ha hecho uso del lenguaje de programación Python y de las técnicas de *Web Scraping*, como el uso de la librería Selenium, con el objetivo de extraer la información almacenada en la web.

Dataset

El *dataset* ha sido publicado en la web Zenodo (<https://zenodo.org/>) cuyo código DOI es: 10.5281/zenodo.4264924

También se puede consultar a través del siguiente enlace:
<https://doi.org/10.5281/zenodo.4264924>

Recursos consultados

Para la elaboración de la práctica, se han consultado los siguientes enlaces:

1. Creative Commons. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. <https://creativecommons.org/publicdomain/zero/1.0/deed.en>
2. Creative Commons. Attribution – Non Commercial – Share A like 4.0 International (CC BY-NC-SA 4.0) <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en>
3. Creative Commons. Attribution – Share A like 4.0 International (CC BY-SA 4.0) <https://creativecommons.org/licenses/by-sa/4.0/deed.en>

Participación de cada Integrante del Grupo

Contribuciones	Firma
Investigación Previa	LLTB, JCSV
Redacción de las Respuestas	LLTB, JCSV
Desarrollo Código	LLTB, JCSV

