

Estructuras de Datos

Universidad Nacional de Tres de Febrero

Trabajo Práctico

Fecha de entrega:

17/11/2020

El objetivo de este trabajo práctico, es poner en práctica y sobre todo en código, los conocimientos adquiridos en la materia en un caso práctico.

Se deberá realizar una aplicación que le permita al usuario comparar precios de distintos productos publicados online.

El usuario podrá ingresar una palabra, o una frase y se deberá realizar la búsqueda en al menos 4 tiendas nacionales y generar un archivo csv con todas las respuestas obtenidas ordenadas en primer lugar por la exactitud de la búsqueda (es decir si el resultado contiene la frase o palabra exacta buscada) y como segunda clave de menor a mayor precio.

Para realizar el trabajo se requiere analizar la estructura de al menos 4 sitios web que permitan realizar compras online **de artículos similares**¹, utilizar scrapy para bajar toda la información correspondiente, usando como URL la búsqueda indicada.

Por ejemplo si buscamos freezer:

En Mercado Libre:

[https://listado.mercadolibre.com.ar/freezer#D\[A:freezer,L:undefined\]](https://listado.mercadolibre.com.ar/freezer#D[A:freezer,L:undefined])

En Frávega:

<https://www.fravega.com/l/?keyword=freezer>

En Garbarino:

<https://www.garbarino.com/q/freezer/srch?q=freezer>

En Rodo:

<https://rodo.com.ar/catalogsearch/result/?q=freezer>

Se deberá recuperar todas las publicaciones que presente el sitio, no solo las que presenta en la primera página

Por ejemplo para freezer en Rodo:

<https://rodo.com.ar/catalogsearch/result/?q=freezer>

<https://rodo.com.ar/catalogsearch/result/index/?p=2&q=freezer>

¹ Esto quiere decir que no deberían comparar Frávega con la tienda de COTO digital.

<https://rodo.com.ar/catalogsearch/result/index/?p=3&q=freezer>
<https://rodo.com.ar/catalogsearch/result/index/?p=4&q=freezer>
<https://rodo.com.ar/catalogsearch/result/index/?p=5&q=freezer>
<https://rodo.com.ar/catalogsearch/result/index/?p=6&q=freezer>
<https://rodo.com.ar/catalogsearch/result/index/?p=7&q=freezer>

El archivo de salida deberá tener al menos las siguientes columnas: título de la publicación, categoría, precio, link a la publicación y fecha y hora de recuperación de la web.

El archivo generado tendrá por nombre la búsqueda y la fecha en que se realizó, por ejemplo si se busca freezer vertical
freezer_vertical_20-10-2020.csv

Archivo de configuración: Deberá contener las URL de los sitios que se pueden consultar, directorio donde se guardaran los archivos generados, etc.

Menú: El usuario podrá elegir en cuales marketplace realizar la búsqueda y en el caso que la búsqueda tenga varias palabras podrá optar por buscar:

1. Publicación con la frase exacta
2. Publicación que contenga todas las palabras
3. Publicación que contenga algunas de las palabras.

Se deberán omitir las stopwords en las búsquedas 2 y 3, sólo serán aceptables en el tipo 1 de búsqueda.

Entregables opcionales

El Trabajo Práctico estará aprobado si cumple con los requisitos básicos. Sin embargo, a fin de completar la nota del mismo, se enumeran los siguientes requisitos opcionales que podrán elegirse individualmente o en su totalidad, según vean que pueden llegar a realizarlos o no.

El requisito es el de primer nivel, para darlo por cumplido deberá resolver todos los requisitos secundarios.

1. Como usuario quiero poder especificar el **formato de salida** de mi informe para tener mayor flexibilidad al momento de utilizar estos datos
 - a. Como usuario del informe quiero recibirlo en formato **csv** para poder analizarlo con software de planilla de cálculo
 - b. Como usuario del informe quiero recibirlo en formato **json** para poder utilizarlo como fuente de datos dentro de una aplicación desarrollada en Javascript.
 - c. Como usuario del informe quiero recibirlo en formato tabulado en **HTML**, para poder mostrarlo desde un navegador.
2. Como usuario quiero obtener una **tabla comparativa** de precios en diferentes tiendas para poder elegir el que más me convenga.
 - a. Como comparador de precios, me interesa comparar aquellos productos que sean **exactamente el mismo modelo**, para saber que estoy comprando el producto deseado en la tienda más económica.
 - b. Como comparador de precios quiero que la tabla esté **ordenada** de acuerdo a los productos disponibles en más tiendas, y las tiendas ordenadas de menor precio promedio a mayor precio promedio para todos los productos comparables.
 - c. Como comparador de precios, aquellos productos que estén solamente disponibles en una tienda deseo visualizarlos en una tabla aparte con un formato adecuado para ello.
3. Como usuario quiero obtener una **estadística de tiendas más económicas**, para tener preferencias por las mismas.
 - a. Como buscador de precios quiero obtener un listado de las tiendas ordenado por cantidad de precios menores que posee para el mismo artículo, sobre todas las otras tiendas.
 - b. Como buscador de precios quiero dejar de ver en el informe las tiendas que no tengan mejores precios que ninguna otra tienda.
4. Como usuario quiero utilizar como parámetro un **modelo** de producto en particular para focalizar mi búsqueda
 - a. Como usuario especializado quiero ver las **especificaciones** oficiales del producto elegido antes de realizar el scraping, soportando al menos tres primeras marcas (ejemplo: Samsung, Longvie, Philco).
 - b. Como usuario especializado quiero obtener un informe con dichas especificaciones y fotos del producto para saber qué estoy comprando, junto al listado de precios en todas las tiendas.

Entrega

Se deberá entregar un informe en formato notebook de jupyter, con fragmentos de código embebido. El informe debe incluir diagrama de clases, decisiones de diseño, análisis de las páginas de los sitios seleccionados.

Junto al informe se deberán entregar código fuente, con menú para realizar pruebas y **pruebas unitarias**.

Defensa

La defensa del trabajo consta de dos etapas: el armado de un video explicando y demostrando el sistema, y la defensa en una reunión por Meet con su tutor asignado.

Video explicativo

El video explicativo de menos de 10 minutos, deberá dar cuenta de:

- una demostración de uso
- las decisiones de diseño aplicadas
- las clases utilizadas
- las técnicas de programación que se emplearon

Esto deberá realizarse mediante algún software de captura de pantalla, con explicaciones mediante texto hablado. En pocas palabras: **como un screencast**.

Para simplificar el esfuerzo, se permite que sólo un integrante muestre y hable en el video, ya que compilar los fragmentos implica un esfuerzo adicional que no reviste mayor beneficio.

El video deberá entregarse por YouTube, en forma de video oculto (accesible por dirección) el día de la entrega del trabajo, junto con la entrega anteriormente descrita.

Defensa por Meet

Una vez que los docentes revisen los videos y consideren que el trabajo se encuentra en condiciones de ser defendido, coordinarán con su tutor un momento en el que puedan realizar una videoconferencia para defender su trabajo.

Durante la misma, el tutor realizará preguntas para profundizar sobre las decisiones de diseño, la funcionalidad, y algunas ideas adicionales.

Se evaluará a todos los integrantes del grupo, por lo que es conveniente para esa ocasión tener **micrófono y cámara disponibles**, a modo de hacer la llamada más amena.

La configuración más simple es utilizar el teléfono celular para la llamada, con auriculares.