

Descripción de la Solución

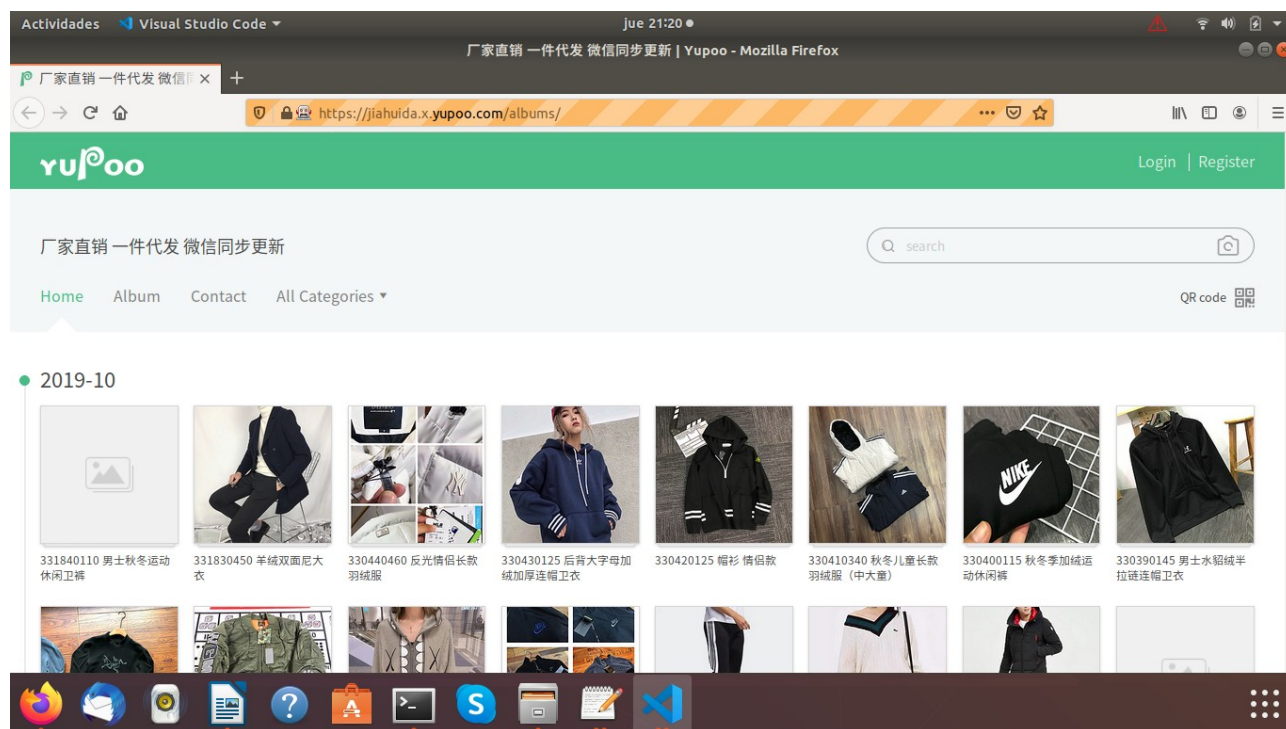
Web Scraper desarrollado en Python configurable mediante el archivo PROPERTIES.py, el script hace uso de las librerías:

- [beautifulsoup4](#)
- [selenium](#)
- navegador web firefox

La solución es compatible con Windows y Linux, en cualquier plataforma se debe de contar con la instalación de las librerías mencionadas, así como de Python 3. Para realizar la instalación de las librerías consultar el [Anexo 1](#)

Instrucciones de uso

Ejecutar el script directamente desde una consola mediante el intérprete de Python, se observará como se abre un navegador firefox y posteriormente se cierra, esto implica que la página está siendo procesada, además en la barra de navegación se mostrará un robot como el de la imagen que se muestra a continuación:



Para poder configurar la correcta ejecución del script se tiene en el archivo PROPERTIES.py:

CSV_PATH_FILE—Es la ruta donde se almacenará el archivo csv de salida, debe incluir el nombre de archivo.

PATH_SELENIUM—Es la ruta donde se encuentra instalada la librería de selenium en el equipo, para el caso de unix/linux es el geckodriver y para Windows es el archivo selenium.exe

CHECK_ONLY_ALBUM—Con esta bandera (True/False) se decide si el script va a ser ejecutado en modo de revisar solo un album o extraer de una página todos los álbumes

True—Sólo extrae las imágenes de un solo album

False—Extrae todas las imágenes dentro de una página, ordenadas por album

DOWNLOAD_URL_ALBUM—Es la url del album en específico, por ejemplo:
<https://jiahuida.x.yupoo.com/albums/74323957?uid=1>

DOWNLOAD_COMPLETE_ALBUMS—Es la url que muestra un listado de todos los albums, por ejemplo: <https://jiahuida.x.yupoo.com/albums/>

Resultados

Se genera un archivo en formato csv con la siguiente cabecera:

name, description, images, referenceAlbum

name—Es el nombre del album

description—Es la descripción por album

images—Lista de imágenes por album

referenceAlbum—Url del album

	A	B	C	D	E
1	name	description	images	referenceAlbum	
3	331840110 男士秋冬运动休闲卫衣	丰富你的日常户外运动选择 各种体型轻松驾驭 重工刺绣小标 简单精致松紧腰围 收口裤脚版型,修身显瘦 细节各方面都一丝不苟! 品质细节非常的到位! 衣柜里必不可少的一条裤子 上身效果特别棒! 给你柔软舒适的体验 经典百搭的款式!			
4	331830450 羊绒双面尼大衣	尺码: L—3XL 复古双排扣设计尽显英伦绅士风度! 高端货, 绝对的自留款! ! 颜色: 黑色 藏青色 尺码: m-xxxl模特180 140上身L码供参考! M: 衣长82 肩宽43.8 胸围102 袖长61 L: 衣长84 肩宽45 胸围106 袖长62.5 XL: 衣长86 肩宽46.2 胸围110 袖长64 XXL: 衣长88 肩宽47.4 胸围114 袖长65.5 XXXL: 衣长89 肩宽48.6 胸围118 袖长66 经过一百二十度高温水洗工艺漂洗以达到去味、杀菌的效果! 简便连帽采用 颜色: 黑色 藏青色		https://liahuida.x.yupoo.com/albums/74323963?uid=1	
5			https://photo.yupoo	https://liahuida.x.yupoo.com/albums/74323957?uid=1	

Cabe mencionar que existen albums que no cumplen con todas estas características, a continuación se presentan algunos ejemplos:

Actividades Navegador web Firefox Jue 21:40 103010115 M-XL 阿迪达斯连帽卫衣 | Yupoo - Mozilla Firefox

103010115 M-XL 阿迪达斯连帽卫衣 | 20

厂家直销 一件代发 微信同步更新

Casa Álbum Contacto Todas Las Categorías

Código QR

103010115 M-XL 阿迪达斯连帽卫衣 | 20

solicitar una descarga masiva compartir

Miniatura Detalle Grande

： 白色
： M - XL
： 加厚加绒软糯面料
衣长 胸围 袖长
67 108 64
69 110 66
71 112 68

： 白色
： M - XL
： 加厚加绒软糯面料
衣长 胸围 袖长
67 108 64
69 110 66
71 112 68

Screenshot_20191017... Screenshot_20191017... 1571293724528.jpg 1571293724528.jpg 1571293724502.jpg 1571293724502.jpg 1571293724478.jpg 1571293724478.jpg

Actividades Navegador web Firefox Jue 21:41 331840110 男士秋冬运动休闲卫裤 | Yupoo - Mozilla Firefox

331840110 男士秋冬运动休闲卫裤 | 0

丰富你的日常户外运动选择
各种体型轻松驾驭
重工刺绣小标
简单精致松紧腰围
收口裤脚版型，修身显瘦
细节各方面都一丝不苟！
品质细节非常的到位！
衣柜里必不可少的一条裤子
上身效果特别棒！
给你柔软舒适的体验
经典百搭的款式！

尺码：L-3XL

solicitar una descarga masiva compartir

Miniatura Detalle Grande

No Hay Imagen Aún

Anexo 1

Instalación de librerías externas

Para poder instalar las librerías se hace uso del manejador de dependencias de Python, pip el cual se puede instalar mediante el comando: `python get-pip.py` (Windows) y `sudo apt-get install python3-pip` (Linux)

BeautifulSoup4

Desde una terminal, y después de haber instalado pip ejecutar el siguiente comando:

```
pip3 install beautifulsoup4
```

Selenium

Desde una terminal, y después de haber instalado pip ejecutar el siguiente comando:

```
pip install selenium
```