Documentación – Scrapping web procesos judiciales Ecuador

Proyecto Scrapping a pagina web de consulta de procesos judiciales de Ecuador: <https://procesosjudiciales.funcionjudicial.gob.ec/busqueda-filtros>

# Objetivo

El presente desarrollo tiene como fin recopilar datos de la pagina de procesos judiciales de Ecuador por medio de la tecnica de Scrapping, de tal forma que se puedan recopilar los datos de las consultas automatizadas en archivos planos en formato csv.

# Herramientas y especificaciones

Este desarrollo fue realizado usando:

* Python 3.12.3
* Biblioteca Selenium WebDriver 4.22.0
* Biblioteca BeautifulSoup 4.12.3

# Funciones:

**función: buscar\_proceso**: Recibe dos argumentos:

* Identificación: Es el número de identificación de la persona o empresa que se quiere consultar, sea como ofendido o como demandado.
* Input\_id: Es el identificador del campo del formulario de la pagina web en el cual será ingresado el número de identificación para hacer la consulta.

Esta función cumple con la tarea de abrir la pagina en el navegador Chrome, busca el campo del formulario donde se ingresará el numero de identificación; para buscar el numero como ofendido el campo esta identificado como mat-input-1 y para la búsqueda como demandado es el campo mat-input-3. Todo esto, usando Selenium.

A su vez crea una lista que guardará los datos recopilados en el scrapping para después plasmarlo en el archivo.csv que será llamado bajo la nomenclatura ofendido\_{identificacion}.cvs o demandado\_{identificacion}.cvs.

El proceso de Scrapping se hace usando BeautifulSoup el cual hace un parse de la estructura del Html, recorre los campos de la tabla resultante en la web y los agrega a la lista creada anteriormente.

Nuevamente Selenium interactúa cliqueando en el botón siguiente de la paginación, ya que la pagina solo arroja tandas de 10 resultados y finaliza el recorrido de todas las paginas hasta que el botón siguiente este deshabilitado.

Si todo fluye correctamente se hace el cierre del navegador, de lo contrario muestra un error capturado en una excepción.

**función: buscar\_detalle\_proceso**: Recibe tres argumentos:

* Numero\_proceso: Es el número de proceso leído desde los archivos CSV resultantes de la ejecución de la función buscar\_proceso.
* Input\_id: Es el identificador del campo del formulario de la página web que para la búsqueda especifica por número de proceso será mat-input-0.
* Driver: el cual tendrá por defecto None, que nos ayudará a mantener el navegador abierto durante el scrapping del detalle sin necesidad de abrir el navegador en cada iteración, optimizando el tiempo empleado en toda la ejecución.

Esta función también abre la página el navegador Chrome, busca el campo donde se ingresará el numero del proceso y luego de escribirlo en el campo mat-input-0, oprime el botón de búsqueda, posteriormente busca el icono del folder que mostrara el detalle de cada proceso y hace el Scrapping de todo el detalle de cada proceso buscado, recopilándolo en una lista para luego ser exportado a otros archivos .CSV independientes.

**función: exportar\_a\_csv:** Esta función recibe dos argumentos que son:

* Datos: es la lista resultante de las dos funciones anteriormente vistas.
* Nombre\_archivo: es el nombre del archivo con el que será creado el .CSV resultante.

Esta función usa el módulo Python llamado CSV para hacer la escritura de texto en dicho formato, agrega un encabezado para los títulos y los organiza en columnas.

**función: main:** Esta función tiene una variable que es el número de identificación de la persona que se va a consultar en la página y ejecuta de manera lógica las funciones anteriormente vistas, agregando algunos print de control en consola para confirmar la finalización exitosa de cada una de estas, además posee dos listas vacías donde guardara el detalle recopilado después de ejecutar la función buscar\_detalle\_proceso.

Adicionalmente, usa nuevamente el modulo CSV para leer los archivos generados en la función buscar\_proceso y recorrerlo para tener uno de los argumentos que necesita la función buscar\_detalle\_proceso, luego llama a la función exportar\_a\_csv tanto para el archivo ofendido como para el archivo demandado.

# Pruebas:

En el marco de las pruebas se logra confirmar que en un ambiente de total funcionamiento por parte de la pagina de procesos judiciales de Ecuador, este desarrollo cumple con las tareas explicadas a cabalidad.

Sin embargo, se ha logrado identificar que la pagina presenta los siguientes problemas en su entorno que ocasionan un comportamiento inesperado del desarrollo:

* En promedio, en uno de cada 20 ejecuciones, la pagina no carga los botones de paginación, ocasionando que el primer archivo CSV resultando solo contenga 10 registros.
* Para el caso de los números de identificación 0968599020001 y 1791251237001, los cuales arrojan demasiados registros, se encontró que en el proceso de scrapping del detalle de cada proceso (buscar\_detalle\_proceso) la pagina puede presentar un error de conexión interrumpiendo así el proceso de ejecución.