



Documentación Step by Step

Solución Web Scrapping - Idealista

1. Importación de módulos y configuración inicial.
2. Función `get_page_source(url)`:
 - Espera un tiempo y copia la URL al portapapeles.
 - Emula el atajo de teclado para abrir la página utilizando la URL del portapapeles.
 - Espera a que se cargue la página y abre el código fuente.
 - Copia el contenido del código fuente al portapapeles y lo guarda en una variable.
 - Escribe el contenido del código fuente en el archivo `pagesource.txt`.
 - Imprime un mensaje indicando que el contenido de los elementos `<article>` de la URL se ha agregado al archivo.
3. Función `process_page_source()`:
 - Lee el archivo `pagesource.txt`.
 - Realiza algún procesamiento adicional con el contenido del archivo.
 - Imprime un mensaje indicando que el procesamiento adicional se ha completado.
4. Función `stop_program()`:
 - Imprime un mensaje indicando que el programa se ha detenido.
 - Finaliza el programa.
5. Inicio del programa principal:
 - Lee las URLs desde el archivo `urls.txt`.
 - Imprime un mensaje de inicio y espera un tiempo.
 - Activa la ventana del navegador.
 - Establece la tecla de escape (`esc`) para detener el programa.
 - Para cada URL:
 - Invoca `get_page_source(url)` para obtener el código fuente de la página.
 - Invoca `process_page_source()` para procesar el contenido del código fuente.
 - Obtiene los datos de los elementos `<article>` y los almacena en la lista `rows`.
 - Filtra los nuevos datos para evitar duplicados.
 - Guarda los nuevos IDs en el archivo `ids_inmuebles.txt`.
 - Imprime el número total de nuevos datos guardados.
 - Autentica con Google utilizando un archivo JSON de credenciales.
 - Abre la hoja de cálculo de Google por su URL.
 - Selecciona la hoja de cálculo y obtiene la última fila no vacía.
 - Actualiza la hoja de cálculo con los nuevos datos a partir de la última fila.
 - Imprime un mensaje de confirmación de que los nuevos datos se han enviado a la hoja de cálculo en Google.