

Reporte de Trabajo: Web Scraping de Información de Vinos

Candidato: Osiris Ramses Macias Gomez

Proceso Realizado:

Investigación Inicial:

1. Investigé el concepto de web scraping, que es la técnica de extracción automatizada de información de sitios web.
2. Estudié cómo realizar web scraping en el entorno de Node.js y JavaScript.
3. Identifiqué las librerías adecuadas para el análisis y procesamiento de datos web, como Axios para realizar solicitudes HTTP y Cheerio para analizar el HTML.

Obtención de Enlaces y Categorías:

1. Utilicé la librería Axios para realizar una solicitud HTTP a la página web objetivo (<https://www.htfw.com/>) y obtener todos los enlaces presentes en la página.
2. Extraje los enlaces relacionados con las diferentes categorías y tipos de vinos ofrecidos en la página.

Desarrollo de Funciones Scraping:

1. Creé una función llamada `scrapeCategory` para realizar el scraping de información por categoría.
2. Iteré sobre cada producto en la página de una categoría.
3. Dentro de cada producto, extraje detalles como el nombre, precio, descripción, volumen y contenido de alcohol, y la destilería.
4. Almacené estos detalles en un objeto y lo agregué a una matriz de productos.

Scraping de todas las categorías:

1. Creé una función llamada `scrapeAllCategories` para realizar el scraping de todas las categorías.
2. Iteré sobre la lista de categorías previamente obtenidas del web scraping anterior.
3. Utilicé la función `scrapeCategory` para extraer la información de productos de cada categoría.
4. Almacené los datos extraídos en un objeto `allProducts`, utilizando el nombre de la categoría como clave.

Transformación y Almacenamiento de Datos en archivo JSON

1. Convertí el objeto `allProducts` en una cadena JSON con formato adecuado utilizando `JSON.stringify`.
2. Utilicé la librería `fs` para escribir los datos JSON en un archivo llamado 'products.json'.

Transformación y Almacenamiento de Datos en archivo EXCEL

1. Para lograr convertir toda la información en documento de Excel se utilizó la librería `exceljs`

Preparación de la Hoja de Calculo Excel:

1. Se crea un nuevo objeto `Workbook` utilizando la librería `ExcelJS`, que servirá como contenedor para la hoja de cálculo.

Añadir Hoja de Trabajo y Encabezado

- 1.- Se agrega una nueva hoja de trabajo llamada 'Products' al libro.
2. Se añaden encabezados a la hoja de trabajo, indicando las columnas que contendrán la información de los productos: 'Distillery', 'Name', 'Price', 'Description' y 'CL and Alcohol Percentage'.

Llenar Filas con Datos:

- 1.- Se itera a través de cada categoría en la lista de categorías.
- 2.- Para cada categoría, se obtiene la lista de productos correspondientes desde el resultado del web scraping almacenado en `result[category]`.
- 3.- Se itera a través de cada producto en la lista de productos de la categoría.
- 4.- Para cada producto, se añade una nueva fila a la hoja de trabajo con los detalles del producto: destilería, nombre, precio, descripción y porcentaje de contenido de alcohol.

Guardar el Archivo Excel:

- 1.- Después de completar el llenado de datos, se guarda el archivo Excel utilizando el método `writeFile` de la librería `ExcelJS`.
- 2.- Se especifica el nombre del archivo como 'products.xlsx'.
- 3.- Se manejan las promesas para capturar cualquier error que pueda ocurrir durante el proceso de escritura.

Manejo de Errores:

1. Implementé la captura y manejo de errores utilizando el bloque `.catch` en las promesas.
2. En caso de que ocurriera un error durante el proceso de extracción, se imprime un mensaje de error en la consola.