

**UNIVERSIDAD POLITÉCNICA DE PUEBLA**  
**Ingeniería en tecnologías de la información**



**Proyecto de Estancia Práctica en  
Desarrollador en Sistemas de Software y  
Administrador de Redes**

“Scraping a Carrefour Francia”

Área temática del CONACYT: VII  
Ingenierías y tecnologías

**Presenta:**

**Miguel Ángel Mitzin Izelo**

**Asesor técnico**

Mtro. César Hernández Rosete

**Asesor académico**

Mtra. Rebeca Rodríguez Huesca

Juan C. Bonilla, Puebla, México.

18 de diciembre de 2020

# Resumen

---

La empresa Novomania requiere hacer un análisis de los productos procesados de Carrefour Francia. Para ello, se optó por obtener la información a través de la técnica Scraping, ya que, es una manera más fácil de obtener la información. Esta técnica consiste en extraer las URL para después recopilar la información alimenticia de cada producto que ofrecen.

Para el desarrollo de este proyecto, se implementará la metodología scrum, se mencionarán los pasos a seguir y resultados que conllevará. También se mencionan las herramientas que se utilizarán, sus características, ventajas y desventajas de utilizarlas. Todo este proceso dará como resultado big data de alimentos procesados, para después, la información recolectada sea analizada, comparada con los clientes de Novomana y los alimentos de México.

# Índice

---

1. Introducción	4
1.1 Descripción de problema.	4
1.2 Justificación.	4
1.3 Objetivo general y específico.	5
2. Metodología y herramientas	6
2.1 Metodología scrum	6
2.2 Herramientas	9
3. Resultados	15
3.1 Fase de iniciación.	15
3.2 Fase de planear y estimar.	17
3.3 Fase de implementación.	19
3.4 Fase de revisión y retrospectiva.	35
3.5 Fase de lanzamiento	40
4. Conclusiones y recomendaciones	42
5. Referencias bibliográficas	43

# 1. Introducción

---

En este primer capítulo se hablará de la necesidad del porqué se está realizando este proyecto, su justificación, los objetivos generales y específicos.

## 1.1. Descripción del problema o necesidad

La empresa Novomania se dedica al análisis de los productos procesados de diferentes países que cuentan con un sistema de venta de alimentos procesados y de sus clientes para determinar qué tan saludables pueden llegar a ser, para así, comparar los alimentos que se ofrecen en México y los de sus clientes. Por ello la empresa necesita extraer la información nutrimental que ofrece cada producto para hacer un análisis completo de todos estos productos de la empresa Carrefour Francia.

## 1.2 Justificación

Para la extracción de la información de todos los productos se aplicará la técnica scraping web la cual consiste en tomar las URL de cada categoría de alimento procesado e ir extrayendo la información de manera más rápida, pues extraerla de manera manual sería muy tardado, ya que son 265 enlaces de los cuales cada enlace contiene más de 65 productos. Se decidió aplicar esta técnica, pues, se pretende optimizar el tiempo en la extracción de datos. Para que, éstos sean analizados y comparados.

Con base en la técnica mencionada, para su implementación implica trabajar en el sistema operativo Ubuntu por compatibilidad con las demás herramientas a utilizar que son: Webdriver, cheerio, npm, selenium, node js, JDK 8 de java y maria db.

## 1.3 Objetivo General y Específicos

### General:

Implementar la técnica scraping web para la extracción de información de los productos procesados de la página <https://www.carrefour.fr/>.

### Específico:

- ✓ Extraer las URL que contengan productos procesados.
- ✓ Diseñar una base de datos, para almacenar las URL extraídas.

- ✓ Ejecutar el código desarrollado app\_Webdriverio.js para convertir los enlaces abstraídos en HTML sin css.
- ✓ Pruebas de funcionalidad para verificar la extracción de información de las url sea correcta.
  - Verificar que los enlaces generados por el código se hayan creado correctamente.
  - Ejecutar el enlace generado para validar que no esté dañado.
- ✓ Almacenar los enlaces HTML obtenidos por el código en la base de datos ya diseñada.
- ✓ Ejecutar el código desarrollado app\_cheerio.js a los enlaces HTML obtenidos.
- ✓ Pruebas de funcionalidad para verificar que la extracción de información de los enlaces HTML sea correcta.
  - Verificar que por cada enlace HTML se obtengan los nombres de los productos, las imágenes de cada producto y su información nutrimental en un enlace.
  - Extraer la información nutrimental de cada producto obtenido.
  - Almacenar la información nutrimental de cada producto en un archivo Excel.

# 2. Metodología y herramientas

---

En este capítulo se hablará de la metodología empleada para la implementación de la técnica de scraping, además se describirá cada herramienta que se utilizará para su implementación y las razones por las cuales se decidió trabajar con éstas.

## 2.1 Metodología scrum

Para el desarrollo de este proyecto, se optó por implementar la metodología scrum, que es una metodología ágil para la gestión de proyectos relacionados con el desarrollo de software.

Scrum trabaja a través de desarrollo en ciclos de trabajos llamados sprints , las cuales se limitan a tiempo y terminan en una fecha específica, aunque no se haya terminado el trabajo, pero derivan una de otra. Su objetivo principal es transformar un conjunto de ítems requeridos por el cliente en un incremento de funcionalidad que será desarrollado para el software. [1]

### 2.1.1 Roles de scrum

Esta metodología define 3 roles, los cuales son:

- ✓ **Scrum Master:** es aquel capaz de trabajar a la par del equipo, sintiéndose parte de él y siendo capaz de colocarse al servicio de este.

#### **Funciones y responsabilidades:**

1. Resolver los conflictos que atrasen el proyecto.
  2. Fomentar la autogestión del equipo e impedir interferencia de terceros.
  3. Garantizar la correcta aplicación de scrum.
- ✓ **El dueño del producto:** persona que autorizada para decidir a sobres las funcionalidades y características que tendrá el producto.

#### **Funciones y responsabilidades:**

1. Convertir el backlog de producto, en software potencialmente entregable.

### **Artefactos o herramientas**

Estos artefactos ayudan a planificar y revisar cada uno de los Sprints. Por ello se mencionan los siguientes:

- ✓ **Backlog producto:** es creado y modificado por el dueño del producto. Además, mantendrá una lista actualizada de los requerimientos funcionales para el software.

Para crear la lista de los requerimientos de debe de tomar en cuenta:

1. El grado de prioridad.
2. Esfuerzo de demanda.

Para tomar en cuenta la priorización de los requerimientos de debe de tomar en cuenta:







1. Beneficios.
2. Valor con respecto a productos de la competencia.

- ✓ **Backlog de sprint:** lista reducida de ítems del backlog de producto que han sido negociados con el dueño del producto. Se genera al comienzo de cada sprint y representa que actividades desarrollara durante la iteración actual.

- ✓ **Scrum taskboard:** los ítems incluidos en el backlog se actualizan diariamente en:

1. Tares pendientes.
2. En curso.
3. Terminadas.

Las cuales se visualizan mediante una tabla como la que se muestra en la **FIGURA 1**.

PENDIENTES	EN CURSO	TERMINADAS
		
		
		

**FIGURA 1. Scrum taskboard.**

Al finalizar cada sprint, se hará entrega de un incremento de funcionalidad para el proyecto. [1]

## 2.1.2 Fases

A continuación, se muestran las fases y procesos con las que cuenta scrum, las cuales se describen de acuerdo con la **Figura 2**. [8]



**Figura 2. Procesos y fases de scrum**

### ✓ De iniciación

Se crea la visión del Proyecto que sirve de enfoque y dirección. También se crean e identifican roles claves del proyecto como el Scrum Master, Product Owner, interesados, equipo del proyecto. Así mismo, se define la lista de prioridades o el Product Backlog la cual sirve de base para la elaboración del plan de lanzamiento y tamaño de cada Sprint.

#### **Procesos:**

1. Crear la visión del proyecto.
2. Identificar al Scrum Master y a los interesados o socios del proyecto.
3. Formación del equipo Scrum.
4. Desarrollo de épica.



5. Creación de la lista priorizada de pendientes del producto (Product Backlog).
6. Realizar el plan de lanzamiento.
- ✓ **Planificación y estimación**

Se definen los Sprints de acuerdo con las historias de usuarios, se alinean a todo lo que genera valor a la organización y se hacen las estimaciones de tiempo y esfuerzo para cumplirlas, los cuales se traducen en listas de tareas cuyos tiempos de desarrollo se definen en reuniones de equipo correspondientes, así como el proceso de definición del Sprint Backlog que contiene todas las tareas que deben completarse en el Sprint.

**Procesos:**

1. Elaborar historias de usuario.
2. Aprobar, estimar y asignar historias de usuarios.
3. Elaboración de tareas.
4. Estimar tareas.
5. Elaboración de la lista de pendientes del Sprint (Sprint Backlog).

✓ **Implementación**

Se trabaja en las tareas del Sprint Backlog para crear Sprint Deliverables, para ello se utiliza un Scrumboard para realizar el seguimiento del trabajo y de actividades que se llevan a cabo. Se realizan las Daily Standup Meeting que son reuniones cortas y eficientes en tiempo donde el equipo da el estatus de sus actividades diarias y manifiesta cualquier inconveniente que pueda tener. Igualmente se actualiza o revisa la lista de prioridades de pendientes del producto.

**Procesos:**

1. Crear entregables (Create Deliverables),
2. Llevar a cabo el standup diario (Conduct Daily Standup)
3. Mantenimiento de la lista priorizada de pendientes del producto (Groom Prioritized Product Backlog)

✓ **Revisión y retrospectiva**

Se realizan reuniones que permitan discutir y revisar avances, dependencias e impedimentos en el desarrollo del proyecto. También se lleva a cabo la demostración del Sprint Deliverable al Propietario del producto y a los Socios relevantes en un Sprint Review Meeting. Igualmente, el Scrum Master y el Equipo Scrum se reúnen para discutir las lecciones aprendidas a lo largo del Sprint.

### **Procesos:**

1. Convocar Scrum de Scrums (Convene Scrum of Scrums)
2. Demostración y validación del Sprint (Demonstrate and Validate Sprint)
3. Retrospectiva de Sprint (Retrospect Sprint)

### ✓ **Lanzamiento:**

Se desarrolla el proceso donde los Accepted Deliverables se les entregan a los Socios relevantes. Un acuerdo formal llamado Working Deliverables Agreement documenta la finalización con éxito del Sprint. Del mismo modo, se realizan actividades de retrospectiva que permite identificar mejoras y lecciones aprendidas del proyecto.

### **Procesos:**

1. Envío de entregables (Ship Deliverables).
2. Retrospectiva del proyecto (Retrospect Project).

## **2.1.3 Ventajas y desventajas de scrum [10]**

### Ventajas:

- ✓ Adaptabilidad: El control del proceso empírico y desarrollo iterativo hacen que los proyectos sean adaptables y abiertos a la incorporación del cambio.
- ✓ Transparencia: Todos los indicadores de información tal como un Scrumboard y Sprint Burndown Chart son compartidos, lo que lleva a un ambiente de trabajo abierto.
- ✓ Retroalimentación continua: Se proporciona a través de los procesos llamados Standup diario, demostración y validación del sprint.
- ✓ Entrega continúa de valor: Los procesos iterativos permiten la entrega continua de valor tan frecuentemente como el cliente lo requiere.
- ✓ Entrega anticipada de alto valor: El proceso de creación de la lista priorizada de pendientes del producto asegura que los requisitos de mayor valor del cliente sean los primeros en cubrirse.
- ✓ Resolución de problemas de forma más rápida: Colaboración y colocación de equipos multifuncionales conducen a la resolución de problemas con mayor rapidez.
- ✓ Entregables efectivos: El proceso de creación de la lista priorizada de pendientes del producto y revisiones periódicas después de la creación de entregables, asegura entregas efectivas para el cliente. Se enfoca en el cliente y éste se encuentra siempre involucrado en todo el proyecto.

### Desventajas:

- ✓ Se requiere de un experto en la toma de decisiones durante las reuniones.
- ✓ El proyecto se puede desviar mucho si el product owner no tiene claro lo que quiere.
- ✓ Poca documentación.
- ✓ Dificultad para estimar presupuestos en base a desarrollo del proyecto.

## 2.2 Herramientas

Para aplicar la técnica de scraping, las herramientas que se van a utilizar son las siguientes:

### 2.2.1 Ubuntu [5]

Es un sistema operativo de software libre y código abierto. Orientada tanto al usuario de escritorio como a servidor

Ventajas:

- ✓ Es gratuito a comparación de las versiones de Windows
- ✓ Alta seguridad.
- ✓ Disponibilidad de código fuente para poder editar.
- ✓ Derecho a redistribuir mejoras y reusar otros códigos libres
- ✓ Realiza tareas de mayor eficiencia que Windows.

Desventajas:

- ✓ Al trabajar con Ubuntu estas abierto con la terminal para la instalación de programas lo cual no se ve en Windows.
- ✓ Al ser código libre pueden llegar a haber problemas de patentes de software.
- ✓ Instalar programas y controladores, puede llegar a ser muy difícil, ya que las empresas crean más productos para Windows.

### 2.2.2 Node js [6]

Es una plataforma construida encima del entorno de ejecución de JavaScript para construir aplicaciones de red.

Ventajas:

- ✓ Al estar basado en Javascript el backend y el frontend estarían programados con una misma tecnología.
- ✓ Además de ser Open Source, con amplia evolución y constante actualización por parte de la comunidad, nos da la posibilidad de crear aplicaciones sin tantos requerimientos y al ser modularizado nos permite importar sólo los módulos necesarios para correr la aplicación.

Desventajas:

- ✓ Cuenta con una API inestable, ya que, entre versiones se tiende a romper la compatibilidad y se deben hacer cambios en tus aplicaciones para mantener todo funcionando.
- ✓ Cualquier operación con uso intensivo de la CPU anula todos los beneficios de rendimiento que Node.js ofrece con su modelo de E / S sin bloqueo controlado por eventos porque todas las solicitudes entrantes se bloquearán mientras el hilo esté ocupado con su procesamiento de números.

### 2.2.3 NPM

Node package manager, proporciona dos funciones principales: Repositorios en línea de paquetes / módulos de node.js. NPM los proporciona en una sola línea de código. [4]

Ventajas:

- ✓ Evita que tengas que descargar manualmente todas las dependencias de un paquete,
- ✓ Puedes publicar y actualizar una librería en la comunidad npm con una sola línea de comando.
- ✓ Administra múltiples versiones de código y dependencias de código.

Desventajas:

- ✓ Debe ser específico con la versión de la librería a utilizar o de lo contrario algunos módulos no serán compatibles con lo que se está desarrollando. [9]

### 2.2.4 Selenium [3]

Es una herramienta de código abierto que se utiliza para automatizar las pruebas realizadas en los navegadores web.

Ventajas:

- ✓ Impulsa el navegador de manera mucho más efectiva y supera las limitaciones de Selenium 1.x, que afectó la cobertura de prueba funcional, como la carga o descarga de archivos, ventanas emergentes y barreras de diálogo.
- ✓ WebDriver supera la limitación de la política de origen de host único de Selenium RC.

Desventajas:

- ✓ No permite ningún tipo de organización, más allá de agrupar en módulos
- ✓ Es necesario tener el navegador arrancado

- ✓ Las pruebas tienen que ser locales; no se pueden lanzar en remoto.

### 2.2.5 Java [7]

Es un lenguaje de programación orientado a objetos el cual se puede ejecutar en diversas plataformas con diferentes sistemas operativos como Windows IOS y Linux.

Ventajas:

- ✓ Su JDK permite correr una maquina virtual, permitiendo ahorrar la reserva de memoria.
- ✓ Recursos disponibles como librerías y su propia documentación
- ✓ Es multiplataforma para desarrollar aplicaciones en web, escritorio y móvil.

Desventajas:

- ✓ Al tratarse de un lenguaje interpretado, el rendimiento en la ejecución de programas suele ser un poco menor .
- ✓ Solo se pueden ejecutar los programas de java con el JDK, sin este, no se pueden ejecutar los programas.
- ✓ No se recomienda que personas que programen de manera estructural programen en este, pues el ser orientado a objetos lo hace un nuevo paradigma.

### 2.2.6 Maria DB [2]

Es un sistema de gestión de bases de datos, es decir, un conjunto de programas que permiten modificar, almacenar, y extraer información de una base de datos que se deriva de MySQL.

Ventajas

- ✓ Proporciona estadísticas de índices y tabla.- pueden ayudar para la optimización de la base de datos.
- ✓ La precisión para tipo de datos TIME, DATETIME y TIMESTAMP.- ampliada al microsegundo.
- ✓ Columnas dinámicas.- proporcionan al usuario columnas virtuales en las tablas.

Desventajas

- ✓ Suele tener un retardo al liberar la versión estable, equivalente en nomenclatura a la de MySQL, para poder implementar sus mejoras y realizar las pruebas pertinentes.

Se decidió trabajar con estas herramientas por el fácil uso que se le puede dar, ya que, existen otras herramientas para aplicar la técnica de scraping como

puede ser Python, pero, debido a la falta de conocimiento de programar en Python y la falta de compatibilidad de Ubuntu con este tipo de herramientas se decidió optar por las ya mencionadas.

# 3. Resultados

---

## Resumen

En este capítulo abarcara los roles de trabajo, la historia de usuario y los pasos que se siguen para obtener la información de los productos procesados.

### 3.1.- Fase de iniciación

#### 3.1.1.- Visión del proyecto.

Extraer una gran cantidad de información de productos procesados de Carrefour Francia en corto tiempo, a través de la técnica scraping para su respectivo análisis de alimentos saludables.

#### 3.1.2.- Formación del equipo scrum.

En la **TABLA 1**, se muestra la definición de roles para la realización de este proyecto.

Equipo scrum		
Scrum master	Dueño del producto	Equipo de desarrollo
Ricardo Marciel	César Hernández Rosete	Miguel Ángel Mitzin Izelo.

**TABLA 1. Equipo scrum.**

#### 3.1.3.- Historia de usuario

En la **TABLA 2** se presenta en la historia de usuario donde se muestra la necesidad del cliente.

Historia de usuario	
<b>Numero:</b> HU1	<b>Usuario:</b> César Hernández Rosete.
<b>Nombre de la historia:</b> Información nutrimental de productos procesados.	
<b>Prioridad en negocio:</b> Alta	<b>Riesgo en desarrollo:</b> Alta
<b>Como:</b> usuario	
<b>Quiero:</b> obtener toda la información nutrimental de la página Carrefour Francia	
<b>Para:</b> hacer un análisis de los productos procesados.	
<b>Observaciones:</b> No aplica.	

**TABLA 2. Historia de usuario.**

### 3.1.4.- Creación del product backlog.

De acuerdo con la **TABLA 3**, se muestran las actividades de todos los sprints que se comenzarán a realizar.

En curso	Terminadas
Extraer los enlaces de productos procesados.	Ninguna.
Diseñar la base de datos.	Ninguna.
Almacenar las URL.	Ninguna.
Obtener los enlaces HTML de acuerdo con las URL extraídas.	Ninguna.
Almacenar los enlaces HTML.	Ninguna.
Obtener los enlaces de producto de acuerdo con los enlaces HTML extraídos.	Ninguna.
Almacenar los enlaces de producto en la base de datos.	Ninguna.
Obtener los enlaces HTML de acuerdo con los enlaces de producto extraídos.	Ninguna.
Extracción de información nutrimental de acuerdo con los enlaces HTML de producto obtenidos.	Ninguna.

**TABLA 3. Product backlog.**

### 3.1.5.- Sprints

En las **TABLAS 4 a la 8**, se muestran los sprints para la obtención de la información de los productos procesados.



Sprint 1	
Número de historia de usuario	Tareas
1	Extraer los enlaces de productos procesados de Carrefour Francia.
1	Diseñar la base de datos para almacenar las URL extraídas, los enlaces HTML, los enlaces de producto y los enlaces HTML de producto.
1	Almacenar las URL de Carrefour Francia.

**TABLA 4. Sprint 1.**

Sprint 2	
Número de historia de usuario	Tareas
1	Obtener los enlaces HTML de acuerdo con las URL extraídas.
1	Almacenar los enlaces HTML.

**TABLA 5. Sprint 2.**

Sprint 3	
Número de historia de usuario	Tareas
1	Obtener los enlaces de producto de acuerdo con los enlaces HTML extraídos.
1	Almacenar los enlaces de producto en la base de datos.

**TABLA 6. Sprint 3.**

Sprint 4	
Número de historia de usuario	Tareas
1	Obtener los enlaces HTML de acuerdo con los enlaces de producto extraídos.

**TABLA 7. Sprint 4.**

Sprint 5	
Número de historia de usuario	Tareas
1	Extracción de información nutrimental de acuerdo con los enlaces HTML de producto obtenidos.

**TABLA 8. Sprint 5.**

### 3.2.- Planear y estimar

De acuerdo con las tareas que se asignaron a cada uno de los sprints. En la **TABLA 9** se muestran las fechas de lanzamiento y fin de cada tarea.

Plan de entregas				
Tarea	Sprint	Prioridad	Fecha de inicio	Fecha final
Extraer las URL de productos procesados de Carrefour Francia.	1	Baja.	21 de septiembre del 2020.	22 de septiembre del 2020.
Diseñar la base de datos	1	Baja.	23 de septiembre del 2020.	24 de septiembre del 2020.
Almacenar las URL de Carrefour Francia.	1	Baja.	23 de septiembre del 2020.	24 de septiembre del 2020.
Obtener los enlaces HTML de acuerdo con las URL extraídas.	2	Alta.	25 de septiembre del 2020.	23 de octubre del 2020.
Almacenar los enlaces HTML.	2	Baja.	24 de octubre del 2020.	24 de octubre del 2020.
Obtener los enlaces de producto de acuerdo con los enlaces HTML extraídos y almacenarlos en la base de datos.	3	Media.	26 de octubre del 2020.	6 de noviembre del 2020.
Obtener los enlaces HTML de acuerdo con los enlaces de producto extraídos.	4	Alta	10 de noviembre del 2020.	23 de noviembre del 2020.
Extracción de información nutrimental de acuerdo con los enlaces HTML de producto obtenidos.	5	Alta	24 de noviembre del 2020.	10 de diciembre del 2020.

**TABLA 9. Plan de entregas.**

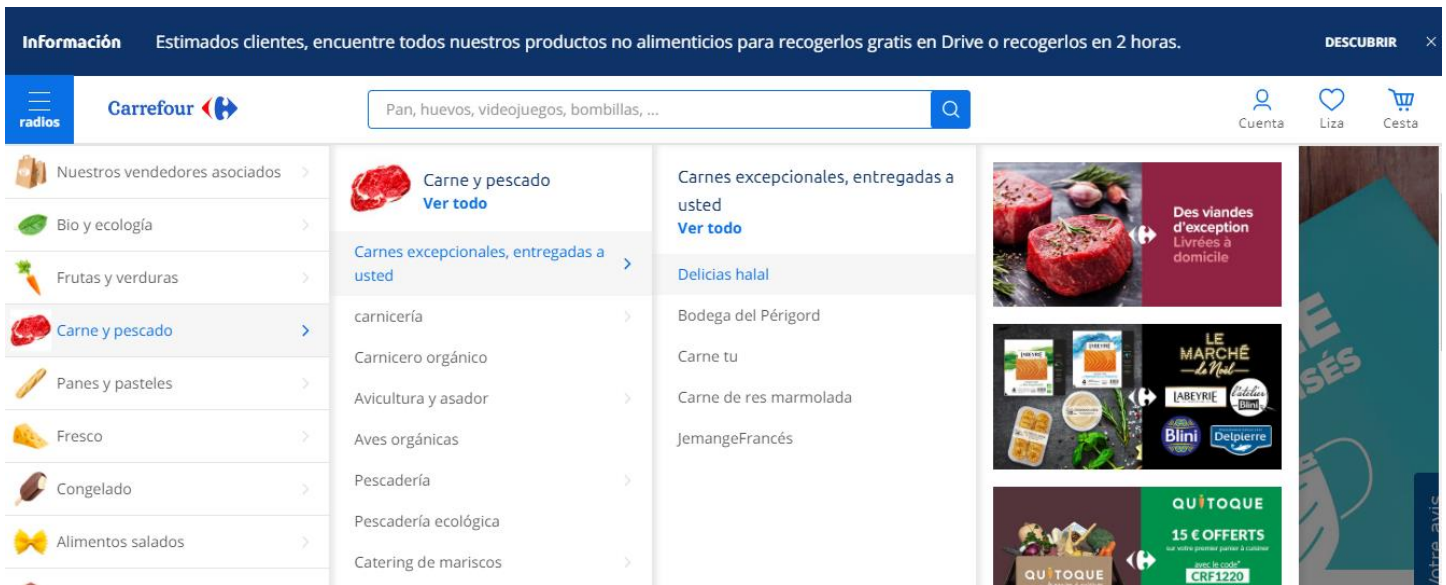
### 3.3.- Implementación

En esta sección se presentará el trabajo realizado en cada uno de los sprints.

#### 3.3.1.- Desarrollo Sprint 1

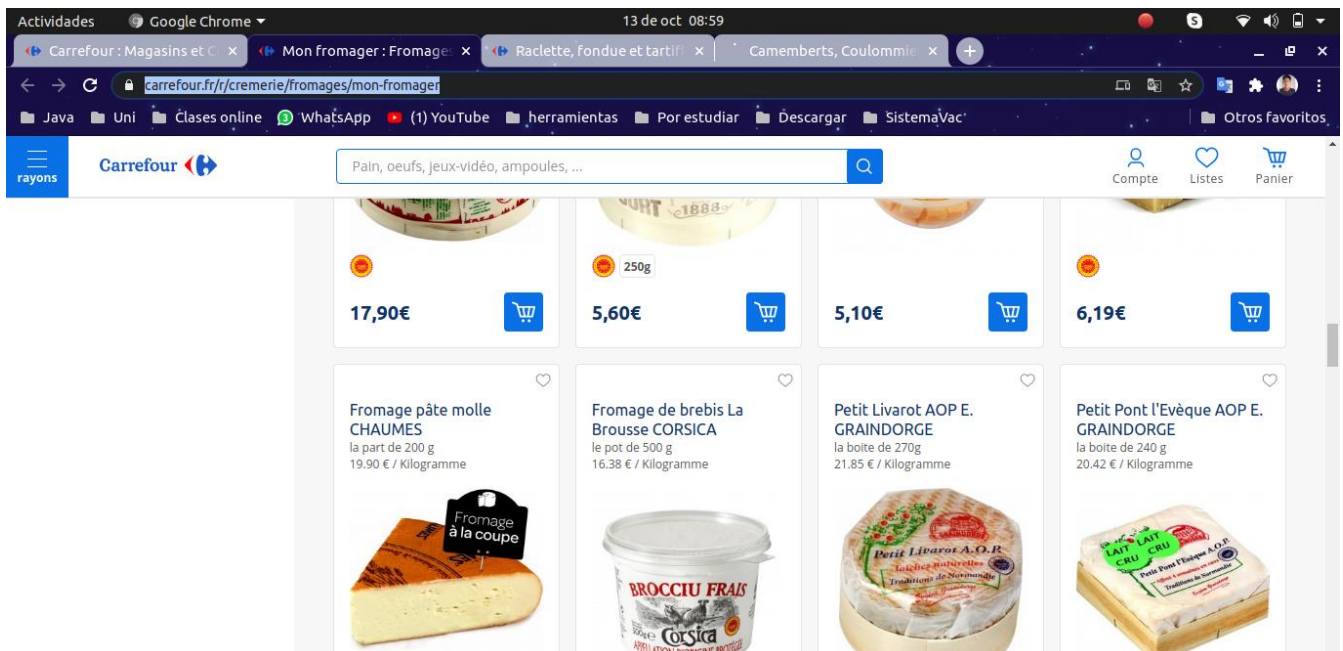
##### Tarea 1: Extracción de las URL

Se ingresó a la página Carrefour Francia, en la cual se seleccionaron las categorías, donde los productos eran procesados, ya que, algunas de éstas no contienen productos procesados. De acuerdo con la **FIGURA 3** se muestra las secciones de productos, las cuales cuentan con 2 subcategorías.

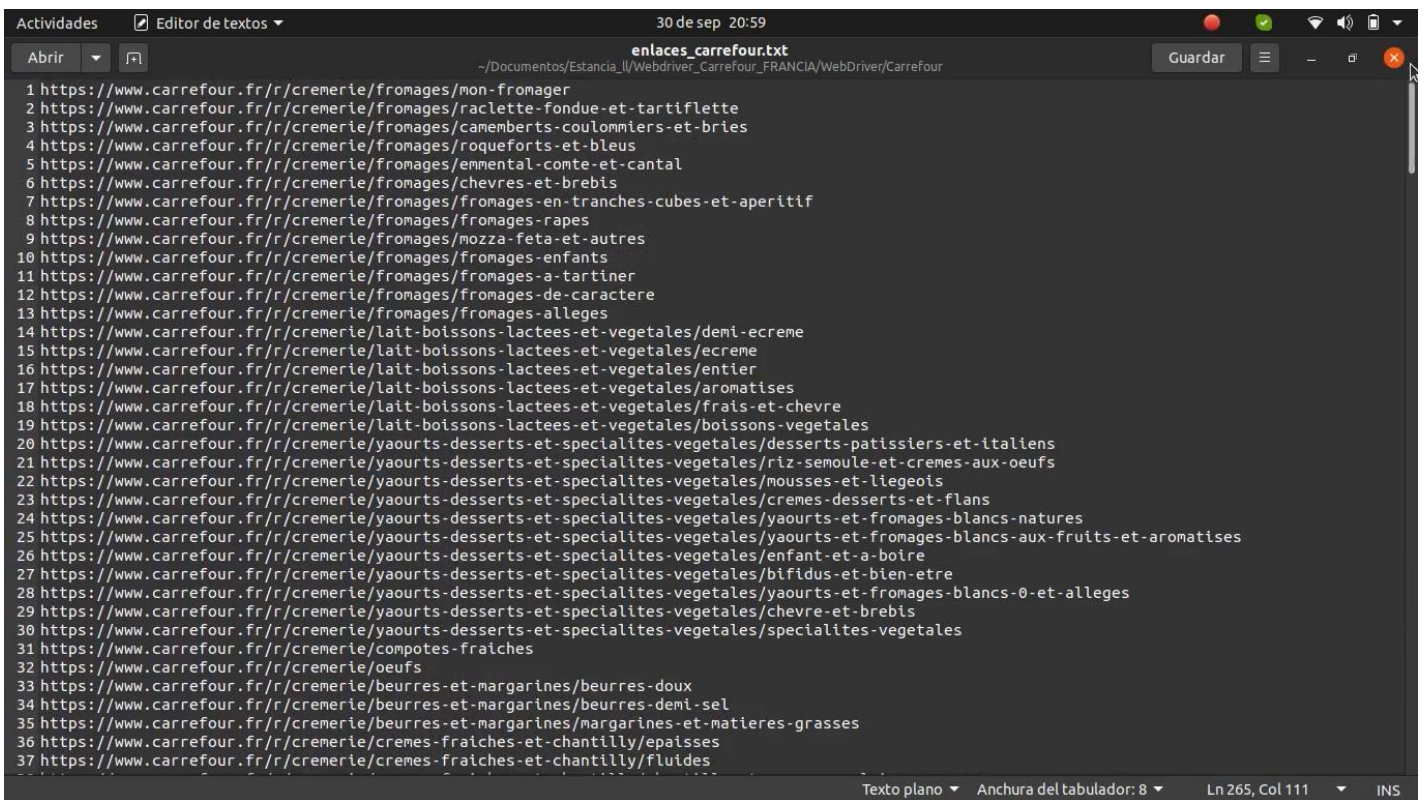


**FIGURA 3. Categorías y subcategorías de Carrefour Francia.**

De acuerdo con la **FIGURA 4**, al acceder a las 2 subcategorías, se muestran los productos, entonces, se copia el enlace y se almacena en el txt **FIGURA 5**.



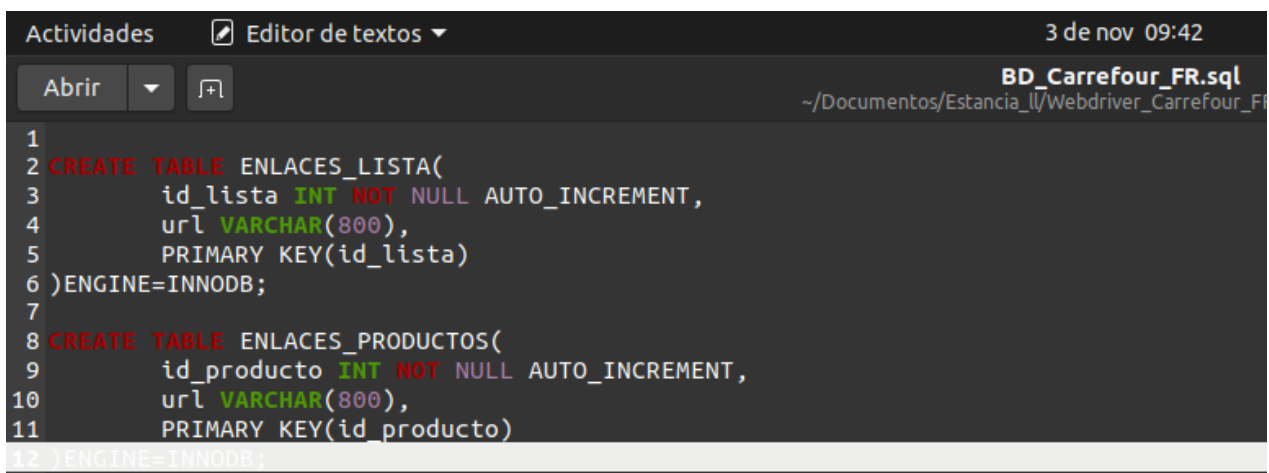
**FIGURA 4. Extracción de enlace de productos.**



**FIGURA 5. Almacenamiento de la URL en un archivo txt.**

## Tarea 2: Diseño de la base de datos

En la **FIGURA 6** se muestra el script de la base de datos, la cual se llama BD\_Carrefour\_FR, donde en la tabla ENLACES\_PRODUCTOS, se almacenarán las URL.



**FIGURA 6. Script de la base de datos.**

De acuerdo con la **FIGURA 7** se muestra la implementación de la base de datos en mariadb.

```
Actividades Terminal 28 de sep 21:34
angelmitzin@angel: ~
MariaDB [(none)]> create data base BD_Carrefour_FR
-> ;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MariaDB server version for the right syntax to use
near 'data base BD_Carrefour_FR' at line 1
MariaDB [(none)]> create data base BD_Carrefour_FR;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MariaDB server version for the right syntax to use
near 'data base BD_Carrefour_FR' at line 1
MariaDB [(none)]> CREATE DATABASE BD_Carrefour_FR;
Query OK, 1 row affected (0.000 sec)

MariaDB [(none)]> use BD_Carrefour_FR;
Database changed
MariaDB [BD_Carrefour_FR]> CREATE TABLE ENLACES_LISTA(
-> id_lista INT NOT NULL AUTO_INCREMENT,
-> url VARCHAR(800),
-> PRIMARY KEY(id_lista)
-> )ENGINE=INNODB;
Query OK, 0 rows affected (0.389 sec)

MariaDB [BD_Carrefour_FR]> CREATE TABLE ENLACES_PRODUCTOS(
-> id_producto INT NOT NULL AUTO_INCREMENT,
-> url VARCHAR(800),
-> PRIMARY KEY(id_producto)
-> )ENGINE=INNODB;
Query OK, 0 rows affected (0.208 sec)

MariaDB [BD_Carrefour_FR]>
MariaDB [BD_Carrefour_FR]> show tables;
+-----+
| Tables_in_BD_Carrefour_FR |
+-----+
| ENLACES_LISTA             |
| ENLACES_PRODUCTOS        |
+-----+
2 rows in set (0.001 sec)

MariaDB [BD_Carrefour_FR]>
```

**FIGURA 7. Implementación de la base de datos.**

### Tarea 3: Almacenar las URL de Carrefour Francia.

Lo siguiente es almacenar las URL en la base de datos, por lo cual de acuerdo con la **FIGURA 8**, se muestra el scrip llamado app\_Insert\_BD.js el cual almacenará las URL en la base de datos.

```
app_Insert_BD.js — ~/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour — Atom
File Edit View Selection Find Packages Help
Telemetry Consent app_Insert_BD.js
1 const fs = require('fs');
2 //Modulos necesarios <-
3 //Conexion con la BD ->
4 var mysql = require('mysql');
5 var c = mysql.createConnection({
6   host   : 'localhost',
7   user   : 'mitzin',
8   password : '1212',
9   database : 'BD_Carrefour_FR'
10 });
11 c.connect();
12 //Conexion con la BD <-
13 //Leer enlaces ->
14 const enlaces = fs.readFileSync("enlaces (copia).txt", "utf-8").split("\n").filter(Boolean); //Leer enlaces de un archivo .txt y convertirlo a un arreglo filtrando los e
15 //Leer enlaces <-
16 //En Javascript puedes usar [for of], que es algo mas practico que el tipico
17 // for(i = 0; i < enlaces.lenght; i++)
18 // console.log(enlaces[i]);
19 // }
20 //Con for(url of enlaces) la variable url es exactamente igual a enlaces[i]
21 //Insertar enlaces en BD ->
22 for(url of enlaces){
23   c.query("SET NAMES utf8");//Es importante codificar bien los enlaces
24   c.query("INSERT INTO ENLACES_LISTA VALUES (NULL, '"+url+"')")
25 }
26 //Insertar enlaces en BD <-
27 //Esto solo es un ejemplo
28 //Escribir en un archivo .txt ->
29 fs.writeFileSync("enlacesHTMLRespaldo.txt", enlaces.join("\n"), "utf-8");
30 //Escribir en un archivo .txt <-
31 //Cerrar la conexion con MySQL ->
32 c.end();
~/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/app_Insert_BD.js 1:1 LF UTF-8 JavaScript GitHub Git (0)
```

**FIGURA 8. Script app\_insert\_BD.js.**

## Scrum diario

La **TABLA 10** describe lo que ocurrió en cada reunión diaria de acuerdo con las tareas de el sprint 1.

Fecha	¿Qué hice ayer?	¿Qué hice hoy?	Responsable	¿Qué obstáculos obtuve?
21/09/2020	Revise que todo estuviera instalado correctamente.	Extraer los URL de Carrefour fr.	Miguel Ángel Mitzin Izelo.	Ninguno.
22/09/2020	Extraje las URL de Carrefour fr.	Extraer los URL de Carrefour fr.	Miguel Ángel Mitzin Izelo.	Ninguno.
23/09/2020	Extraje los URL de Carrefour fr.	Diseñar e implementar la base de datos y almacenar las URL de Carrefour fr en la base de datos.	Miguel Ángel Mitzin Izelo.	Ninguno.

**TABLA 10. Scrum diario.**

### 3.3.2.- Desarrollo Sprint 2

#### Tarea 4: Obtener los enlaces HTML de acuerdo con las URL extraídas.

De acuerdo con la **FIGURAS 9, 10 y 11**, se muestra el scrip app\_Webdriverio.js el cual leerá las URL en la base de datos, abrirá la URL y extraerá el enlace HTML.

```
3  const webdriverio = require('webdriverio');
4  const fs = require('fs');
5  const cheerio = require('cheerio');
6  //const Client = require('mariasql');
7  //Modulos necesarios <-
8
9  /*Modulos opcionales ->
10 //const exec = require('child_process').exec, child;
11
12 /* Este modulo puede ejecutar comandos de terminal dependiendo el sistema operativo.
13 /* Puede ejecutar un audio cuando finalice el Script.
14 */
15 //Modulos opcionales <-
16
17 var mysql = require('mysql');
18 var c = mysql.createConnection({
19   host    : 'localhost',
20   user    : 'mitzin',
21   password : '1212',
22   database : 'BD_Carrefour_FR'
23 });|
24 //Conexion con webdriver ->
25 const options = {
26   desiredCapabilities: {
27     browserName: 'chrome'
28   }
29 };
```

**FIGURA 9. Script app\_Webdriverio.js.**

```

30  const client = webdriverio.remote(options);
31  //Conexion con webdriver <-
32
33  let enlace = 0;
34  let delay = 12000;
35
36  // Iniciar navegacion con Webdriver ->
37  const init = async (array_url) => {
38      await client.init();
39
40      for (const r_url of array_url) {
41          //r_url contiene el id y el url de la tabla
42          await client.url(r_url.url);
43          console.log(r_url.url);
44          await client.execute(() => {
45              const styleElement = document.createElement('style');
46              document.head.appendChild(styleElement);
47              document.readyState === 'complete';
48              styleElement.sheet.insertRule('::-webkit-scrollbar { display: none; }');
49          });
50          console.log(delay);
51          await client.pause(delay); //Delay entre cada enlace
52          if(delay == 12000){
53              delay=8000;
54          }
55

```

**FIGURA 10. Script app\_Webdriverio.js.**

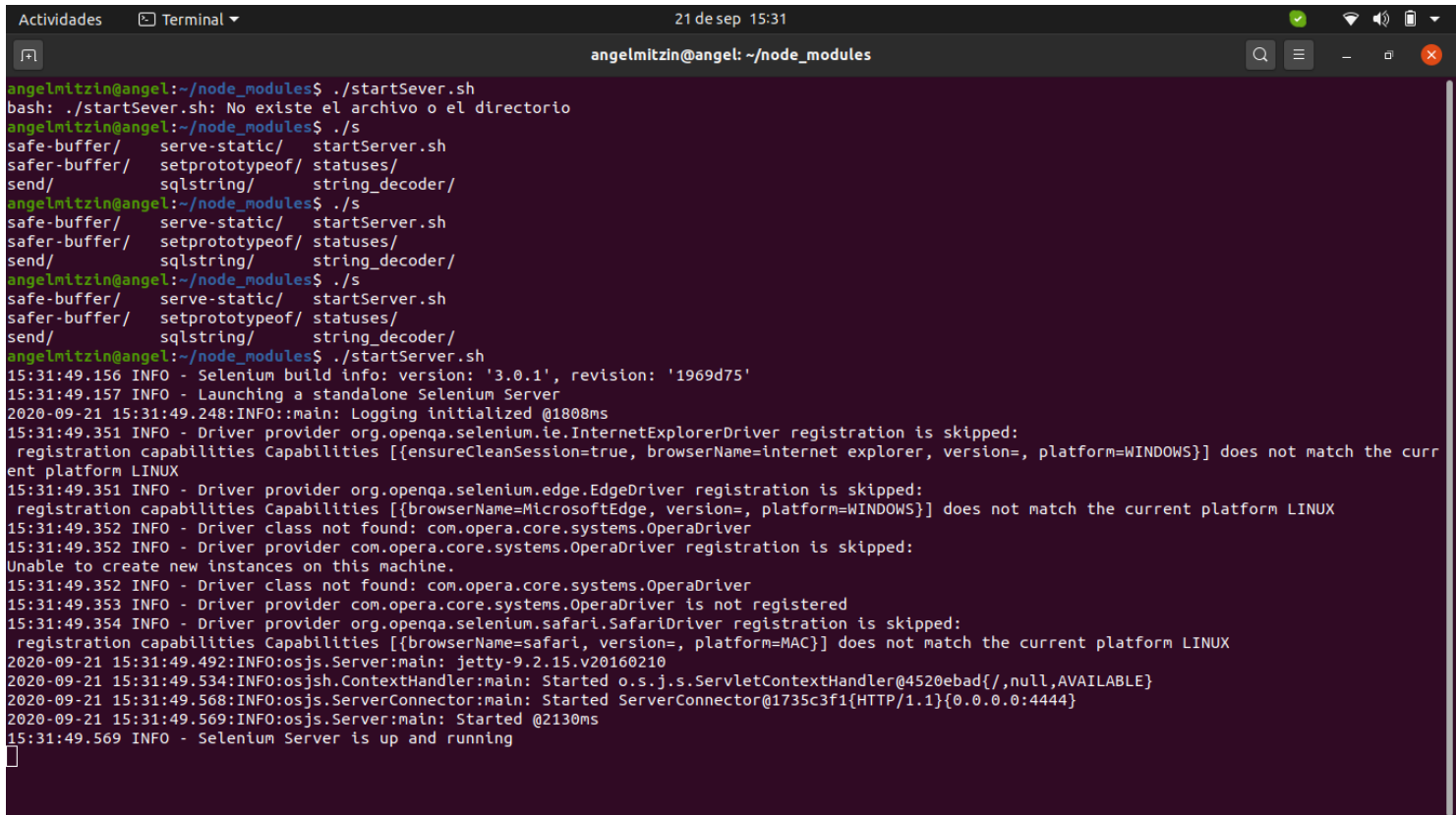
```

56  //Obtencion del codigo HTML ->
57  await client.getSource().then(function (code){
58      if(code.length < 70000 && delay == 8){
59          console.log("El enlace esta defectuoso");
60          // Es necesario terminar el Script para verificar si el enlace es defectuoso
61          // Ó hay algún problema al acceder al sitio Web.
62          //En todo caso, se puede eliminar el enlace manualmente.
63          process.exit(1);
64      }
65      else{
66          enlace = r_url.id;
67          //Detectar si la lsita de productos tiene mas de 65 productos ->
68          const $ = cheerio.load(code);
69          const numero_productos = $(".product-list-header__count span span").text().replace(" resultats", "");
70          if(parseInt(numero_productos) > 65){
71              console.log("El enlace: ["+enlace+".html] tiene más de 65 productos");
72          }
73          //Detectar si la lsita de productos tiene mas de 65 productos <-
74
75          fs.writeFileSync("EnlacesCarrefourHTML/enlace"+enlace+".html",code,"utf-8");
76          //c.query("DELETE FROM enlacesParaProds WHERE id = "+(enlace));
77          c.end();//Esto a veces es opcional, dependiendo el modulo de MySQL instalado.
78          enlace++;
79      }
80  });
81  //Obtencion del codigo HTML <-
82  }
83  return client.end();
84  };

```

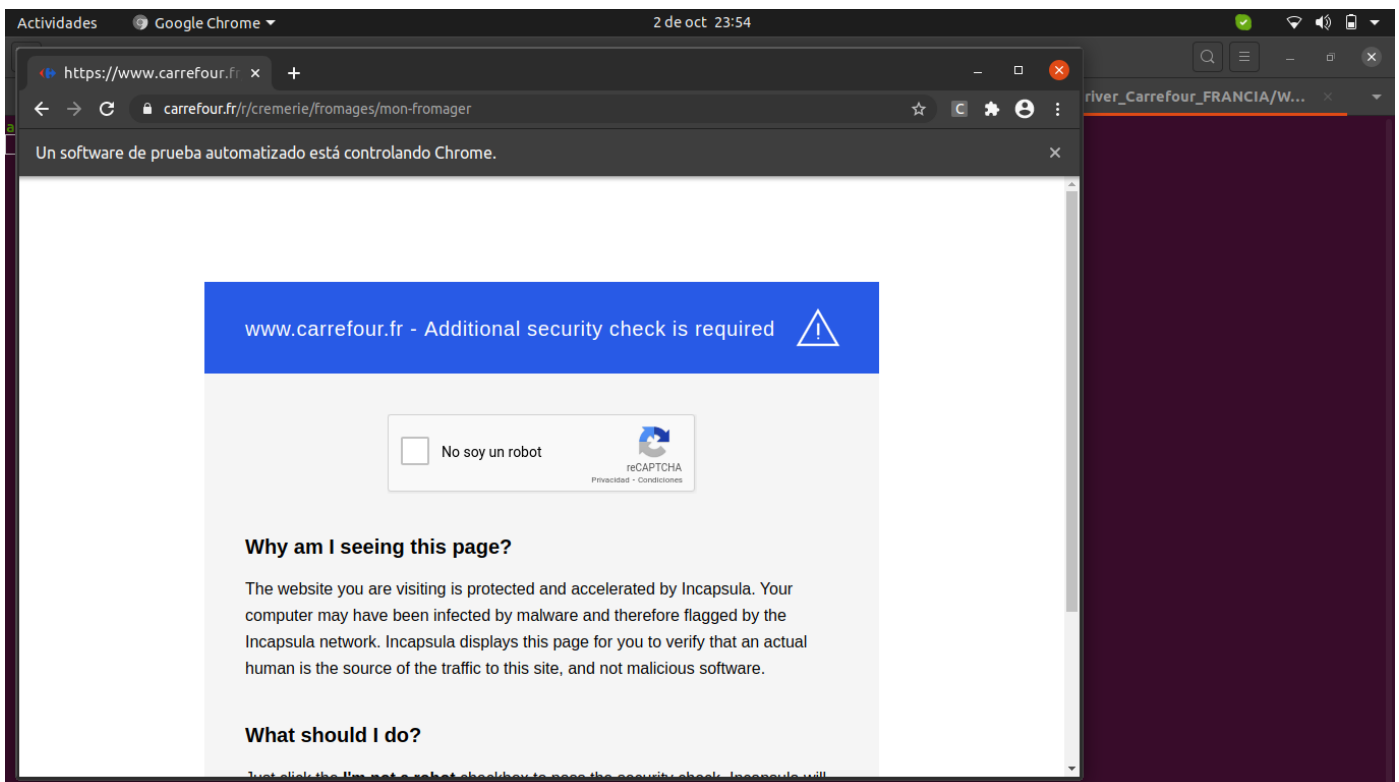
**FIGURA 11. Script app\_Webdriverio.js.**

Para ejecutar el script `app_Webdriverio.js`, primero se ejecuta el servidor selenium con el comando `./StartServer.sh` **FIGURA 12**. Después se ejecuta el script `app_Webdriverio.js` que obtendrá el enlace, abrirá Google Chrome con la URL seleccionada, se tendrá que omitir el captcha para poder acceder a los productos **FIGURA 13** y así, el script pueda obtener el enlace HTML cuando se haya omitido el captcha **FIGURA 14**. El enlace HTML se almacena en una carpeta llamada `EnlacesCarrefourHTML` **FIGURA 15**.



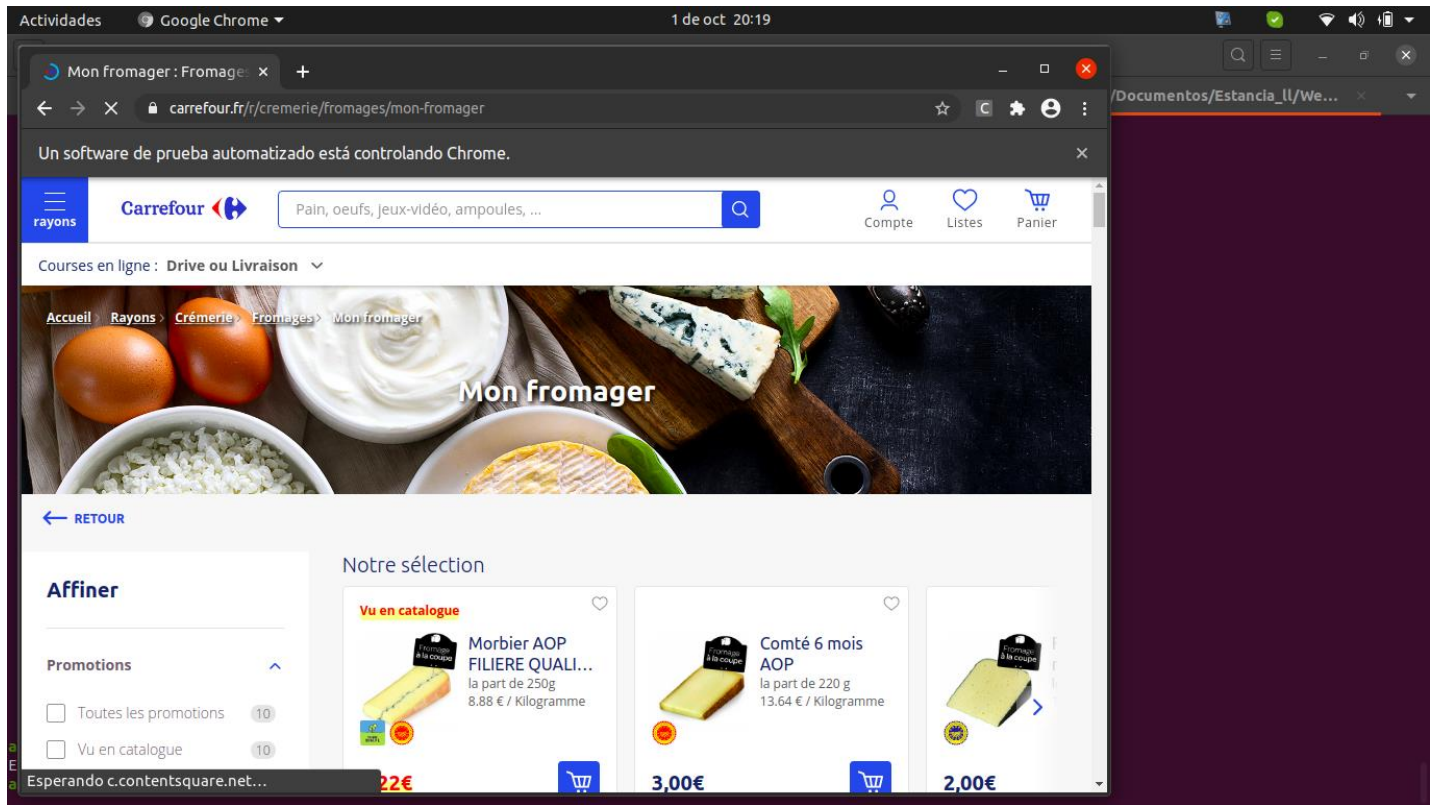
```
angelmitzin@angel:~/node_modules$ ./startSever.sh
bash: ./startSever.sh: No existe el archivo o el directorio
angelmitzin@angel:~/node_modules$ ./s
safe-buffer/      serve-static/    startServer.sh
safer-buffer/     setprototypeof/ statuses/
send/             sqlstring/       string_decoder/
angelmitzin@angel:~/node_modules$ ./s
safe-buffer/      serve-static/    startServer.sh
safer-buffer/     setprototypeof/ statuses/
send/             sqlstring/       string_decoder/
angelmitzin@angel:~/node_modules$ ./s
safe-buffer/      serve-static/    startServer.sh
safer-buffer/     setprototypeof/ statuses/
send/             sqlstring/       string_decoder/
angelmitzin@angel:~/node_modules$ ./startServer.sh
15:31:49.156 INFO - Selenium build info: version: '3.0.1', revision: '1969d75'
15:31:49.157 INFO - Launching a standalone Selenium Server
2020-09-21 15:31:49.248:INFO::main: Logging initialized @1808ms
15:31:49.351 INFO - Driver provider org.openqa.selenium.ie.InternetExplorerDriver registration is skipped:
 registration capabilities Capabilities [{"ensureCleanSession=true, browserName=Internet explorer, version=, platform=WINDOWS}] does not match the curr
ent platform LINUX
15:31:49.351 INFO - Driver provider org.openqa.selenium.edge.EdgeDriver registration is skipped:
 registration capabilities Capabilities [{"browserName=MicrosoftEdge, version=, platform=WINDOWS}] does not match the current platform LINUX
15:31:49.352 INFO - Driver class not found: com.opera.core.systems.OperaDriver
15:31:49.352 INFO - Driver provider com.opera.core.systems.OperaDriver registration is skipped:
Unable to create new instances on this machine.
15:31:49.352 INFO - Driver class not found: com.opera.core.systems.OperaDriver
15:31:49.353 INFO - Driver provider com.opera.core.systems.OperaDriver is not registered
15:31:49.354 INFO - Driver provider org.openqa.selenium.safari.SafariDriver registration is skipped:
 registration capabilities Capabilities [{"browserName=safari, version=, platform=MAC}] does not match the current platform LINUX
2020-09-21 15:31:49.492:INFO:osjs.Server:main: jetty-9.2.15.v20160210
2020-09-21 15:31:49.534:INFO:osjs.ContextHandler:main: Started o.s.j.s.ServletContextHandler@4520ebad[/,null,AVAILABLE]
2020-09-21 15:31:49.568:INFO:osjs.ServerConnector:main: Started ServerConnector@1735c3f1{HTTP/1.1}{0.0.0.0:4444}
2020-09-21 15:31:49.569:INFO:osjs.Server:main: Started @2130ms
15:31:49.569 INFO - Selenium Server is up and running
```

**FIGURA 12. Ejecución del servidor selenium .**

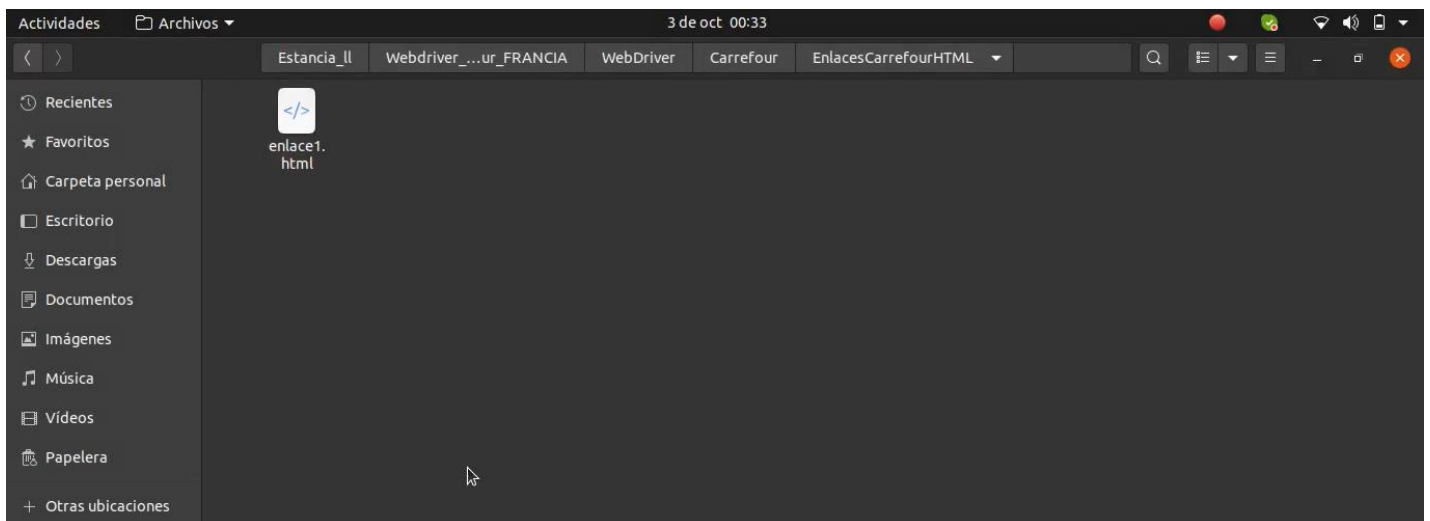


**FIGURA 13. Ejecución de Chrome y omisión del captcha .**





**FIGURA 14. Capcha omitido.**



**FIGURA 15. Obtención y almacenamiento del enlace HTML.**

### **Tarea 5: Almacenar los enlaces HTML.**

Una vez obtenidos los enlaces, estas se almacenarán en la base de datos **FIGURA 16**, utilizando el mismo scrip `app_Insert_BD.js`.

```

Actividades Terminal 3 de nov 09:56
angelmitzin@angel: ~/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour
angelmitzin@angel: ~/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/W...
angelmitzin@angel: ~/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/W...
ENLACES_LISTA
ENLACES_PRODUCTOS
-----
2 rows in set (0.001 sec)
MariaDB [BD_Carrefour_FR]> SELECT * FROM ENLACES_PRODUCTOS;
-----
id_producto | url
-----
1 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace1.html
2 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace2.html
3 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace3.html
4 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace4.html
5 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace5.html
6 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace6.html
7 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace7.html
8 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace8.html
9 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace9.html
10 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace10.html
11 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace11.html
12 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace12.html
13 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace13.html
14 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace14.html
15 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace15.html
16 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace16.html
17 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace17.html
18 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace18.html
19 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace19.html
20 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace20.html
21 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace21.html
22 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace22.html
23 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace23.html
24 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace24.html
25 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace25.html
26 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace26.html
27 | file:///home/angelmitzin/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace27.html

```

**FIGURA 16. Enlaces HTML insertados en la base de datos.**

## Scrum diario

La **TABLA 11** describe lo que ocurrió en cada reunión diaria de acuerdo con las tareas de el sprint 2.

Fecha	¿Qué hice ayer?	¿Qué hice hoy?	Responsable	¿Qué obstáculos obtuve?
25/09/2020	Preparar los códigos para la extracción de los enlaces HTML	Extraer 14 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó.
28/09/2020	Extraje 14 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
29/09/2020	Extraje 14 enlaces HTML de las URL obtenidas.	Extraer 14 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
30/09/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
01/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
02/10/2020	Extraje 14 enlaces HTML de las URL obtenidas.	Extraer 14 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .

05/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
06/10/2020	Extraje 14 enlaces HTML de las URL obtenidas.	Extraer 14 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
07/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
08/10/2020	Extraje 14 enlaces HTML de las URL obtenidas.	Extraer 14 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
09/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
12/10/2020	Extraje 14 enlaces HTML de las URL obtenidas.	Extraer 14 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
13/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
14/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
15/10/2020	Extraje 14 enlaces HTML de las URL obtenidas.	Extraer 14 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
16/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
19/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
20/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
21/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
22/10/2020	Extraje 13 enlaces HTML de las URL obtenidas.	Extraer 13 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
23/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Extraer 12 enlaces HTML de las URL obtenidas.	Miguel Ángel Mitzin Izelo.	Al tratar de obtener otro enlace, el capcha me bloqueó .
24/10/2020	Extraje 12 enlaces HTML de las URL obtenidas.	Almacenar los enlaces HTML en la base de datos.	Miguel Ángel Mitzin Izelo.	Ninguno.

**TABLA 11. Scrum diario.**

### 3.3.3.- Desarrollo Sprint 3

**Tarea 6 : Obtener los enlaces de producto de acuerdo con los enlaces HTML extraídos e inserción de los enlaces en la base de datos.**

Para obtener los enlaces de cada producto, se utilizó el scrip app\_Cheerio.js **FIGURA 17 y 18.**

```
4  const cheerio = require('cheerio');
5  const fs = require('fs');
6
7  var mysql = require('mysql');
8  var c = mysql.createConnection({
9    host    : 'localhost',
10   user    : 'mitzin',
11   password : '1212',
12   database : 'BD_Carrefour_FR'
13 })
14
15 const CARPETA_ENLACES="EnlacesHTML";
16
17 fs.readdir(CARPETA_ENLACES, (err, files) => {
18   for (f of files) {
19     const i = 1;
20     console.log(f);
21     //const codigo = fs.readFileSync("EnlacesCarrefourHTML/enlace1.html", "utf-8");
22     const codigo = fs.readFileSync(f, "utf-8");
23     //console.log("llego hasta aqui");
24
25     const $ = cheerio.load(codigo);
26     const URL_BASE = "https://www.carrefour.fr";
27     const enlaces = [];
28     //const imagenes = [];
29
30     var lista_productos = $(".ds-product-card--vertical div div div a").each(function () {
31       //Extraer nombre del producto ->
32       const nombre = $(this).find("h2").text();
33     });
34   }
35 }
~/Documentos/Estancia_II/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/app_Cheerio.js 5:26
```

**FIGURA 17. Script app\_Cheerio.js**

```
31 //Extraer nombre del producto ->
32 const nombre = $(this).find("h2").text();
33 console.log(nombre);
34
35 //Extraer nombre del producto <-
36
37 //Extraer enlace del producto ->
38 const enlace_extraido = URL_BASE+$(this).attr("href");
39 if(!enlaces.includes(enlace_extraido)){
40   enlaces.push(enlace_extraido);
41   console.log(enlace_extraido);
42
43   //console.log("enlace del producto");
44 }
45 //Extraer enlace del producto <-
46
47 //Extraer url imagen del producto ->
48 /*const url_imagen_extraida = URL_BASE+$(this).find("img").attr("src");
49 if(!imagenes.includes(url_imagen_extraida)){
50   imagenes.push(url_imagen_extraida);
51   console.log(url_imagen_extraida);
52   //console.log("url de la imagen del producto");
53 }*/
54 //Extraer url imagen del producto <-
55
56 //Insertar URL EN LA BASE DE DATOS
57 c.query("INSERT INTO ENLACES_LISTA VALUES (NULL, '"+enlace_extraido+"'");
58 console.log("Enlace"+i+"insertado");
59 i++;
~/Documentos/Estancia_II/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/app_Cheerio.js 5:26
```

**FIGURA 18. Script app\_Cheerio.js**

Al ejecutar el script app\_Cheerio.js, se extrae el nombre del producto y su respectivo URL **FIGURA 19**. Además, el mismo scrip, va almacenando en la base de datos la URL del producto **FIGURA 20**.

```

angelmitzn@angel: ~/Documentos/Estancia_II/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour
Bleu d'Auvergne AOP FILIERE QUALITE CARREFOUR
https://www.carrefour.fr/p/bleu-d-auvergne-aop-filiere-qualite-carrefour-3276559928226
Enlace_insertado
Enlace_insertado
Cantal entre-deux AOP FILIERE QUALITE CARREFOUR
https://www.carrefour.fr/p/cantal-entre-deux-aop-filiere-qualite-carrefour-3276558502328
Enlace_insertado
Enlace_insertado
Cantal vieux AOP FILIERE QUALITE CARREFOUR
https://www.carrefour.fr/p/cantal-vieux-aop-filiere-qualite-carrefour-3276550203926
Enlace_insertado
Enlace_insertado
Fourme d'Ambert AOP FILIERE QUALITE CARREFOUR
https://www.carrefour.fr/p/fourme-d-ambert-aop-filiere-qualite-carrefour-3276553147890
Enlace_insertado
Enlace_insertado
Fromage de chèvre Rocamadour AOP FILIERE QUALITE CARREFOUR
https://www.carrefour.fr/p/fromage-de-chevre-rocamadour-aop-filiere-qualite-carrefour-3245414506718
Enlace_insertado
Enlace_insertado
Fromage de chèvre Chavignol AOP FILIERE QUALITE CARREFOUR

```

**FIGURA 19. Extracción del nombre y enlace de cada producto.**

```

MariaDB [BD_Carrefour_FR]> SELECT * FROM ENLACES_LISTA;
+-----+-----+
| id_llsta | url
+-----+-----+
277 | https://www.carrefour.fr/p/bleu-d-auvergne-aop-filiere-qualite-carrefour-3276559928226
278 | https://www.carrefour.fr/p/bleu-d-auvergne-aop-filiere-qualite-carrefour-3276559928226
279 | https://www.carrefour.fr/p/cantal-entre-deux-aop-filiere-qualite-carrefour-3276558502328
280 | https://www.carrefour.fr/p/cantal-entre-deux-aop-filiere-qualite-carrefour-3276558502328
281 | https://www.carrefour.fr/p/cantal-vieux-aop-filiere-qualite-carrefour-3276550203926
282 | https://www.carrefour.fr/p/cantal-vieux-aop-filiere-qualite-carrefour-3276550203926
283 | https://www.carrefour.fr/p/fourme-d-ambert-aop-filiere-qualite-carrefour-3276553147890
284 | https://www.carrefour.fr/p/fourme-d-ambert-aop-filiere-qualite-carrefour-3276553147890
285 | https://www.carrefour.fr/p/fromage-de-chevre-rocamadour-aop-filiere-qualite-carrefour-3245414506718
286 | https://www.carrefour.fr/p/fromage-de-chevre-rocamadour-aop-filiere-qualite-carrefour-3245414506718
287 | https://www.carrefour.fr/p/fromage-de-chevre-chavignol-aop-filiere-qualite-carrefour-3245413705556
288 | https://www.carrefour.fr/p/fromage-de-chevre-chavignol-aop-filiere-qualite-carrefour-3245413705556
289 | https://www.carrefour.fr/p/fromage-de-chevre-chavignol-bleu-aop-filiere-qualite-carrefour-3245413705501
290 | https://www.carrefour.fr/p/fromage-de-chevre-chavignol-bleu-aop-filiere-qualite-carrefour-3245413705501
291 | https://www.carrefour.fr/p/mont-d-or-aop-igp-filiere-qualite-carrefour-3560070936526
292 | https://www.carrefour.fr/p/mont-d-or-aop-igp-filiere-qualite-carrefour-3560070936526
293 | https://www.carrefour.fr/p/morbier-aop-filiere-qualite-carrefour-3560070459049
294 | https://www.carrefour.fr/p/morbier-aop-filiere-qualite-carrefour-3560070459049
295 | https://www.carrefour.fr/p/morbier-aop-filiere-qualite-carrefour-3276558387307
296 | https://www.carrefour.fr/p/morbier-aop-filiere-qualite-carrefour-3276558387307
297 | https://www.carrefour.fr/p/ossau-iraty-aop-filiere-qualite-carrefour-3276555435599
298 | https://www.carrefour.fr/p/ossau-iraty-aop-filiere-qualite-carrefour-3276555435599
299 | https://www.carrefour.fr/p/plateau-de-fromages-assortiment-igp-aop-filiere-qualite-carrefour-3245414468702
300 | https://www.carrefour.fr/p/plateau-de-fromages-assortiment-igp-aop-filiere-qualite-carrefour-3245414468702
301 | https://www.carrefour.fr/p/reblochon-de-savoie-fermier-affine-6-semaines-aop-filiere-qualite-carrefour-3276559934142
302 | https://www.carrefour.fr/p/reblochon-de-savoie-fermier-affine-6-semaines-aop-filiere-qualite-carrefour-3276559934142
303 | https://www.carrefour.fr/p/reblochon-fermier-de-savoie-5-semaines-d-affinage-aop-filiere-qualite-carrefour-3276559934159
304 | https://www.carrefour.fr/p/reblochon-fermier-de-savoie-5-semaines-d-affinage-aop-filiere-qualite-carrefour-3276559934159
305 | https://www.carrefour.fr/p/tomme-de-savoie-igp-filiere-qualite-carrefour-3245390217462
306 | https://www.carrefour.fr/p/tomme-de-savoie-igp-filiere-qualite-carrefour-3245390217462
307 | https://www.carrefour.fr/p/fromage-au-bouchon-3222110025194
308 | https://www.carrefour.fr/p/fromage-au-bouchon-3222110025194

```

**FIGURA 20. Enlaces de producto insertados en la base de datos.**

## Scrum diario

La **TABLA 12** describe lo que ocurrió en cada reunión diaria de acuerdo con las tareas de el sprint 3.

Fecha	¿Qué hice ayer?	¿Qué hice hoy?	Responsable	¿Qué obstáculos obtuve?
26/10/2020	Prepare el código de app_cheerio.js	Obtener 311 enlaces por producto de acuerdo con los enlaces HTML	Miguel Ángel Mitzin Izelo.	En la base de datos obtuve enlaces duplicados.
27/10/2020	Obtuve 311 enlaces por producto de acuerdo con los enlaces HTML	Obtener 311 enlaces por producto de acuerdo con los enlaces HTML	Miguel Ángel Mitzin Izelo.	Ninguno.
28/10/2020	Obtuve 311 enlaces por producto de acuerdo con los enlaces HTML	Obtener 311 enlaces por producto de acuerdo con los enlaces HTML	Miguel Ángel Mitzin Izelo.	Ninguno.
29/10/2020	Obtuve 311 enlaces por producto de acuerdo con los enlaces HTML	Obtener 311 enlaces por producto de acuerdo con los enlaces HTML	Miguel Ángel Mitzin Izelo.	Ninguno.
30/10/2020	Obtuve 311 enlaces por producto de acuerdo con los enlaces HTML	Obtener 311 enlaces por producto de acuerdo con los enlaces HTML	Miguel Ángel Mitzin Izelo.	Ninguno.
02/11/2020	Obtuve 312 enlaces por producto de acuerdo con los enlaces HTML	Obtener 312 enlaces por producto de acuerdo con los enlaces HTML	Miguel Ángel Mitzin Izelo.	Ninguno.
03/11/2020	Obtuve 311 enlaces por producto de acuerdo con los enlaces HTML	Obtener 311 enlaces por producto de acuerdo con los enlaces HTML	Miguel Ángel Mitzin Izelo.	Ninguno.
04/11/2020	Obtuve 311 enlaces por producto de acuerdo con los enlaces HTML	Obtener 311 enlaces por producto de acuerdo con los enlaces HTML	Miguel Ángel Mitzin Izelo.	Ninguno.

05/11/2020	Obtuve enlaces producto de acuerdo con los enlaces HTML	311 por de los	Obtener enlaces producto de acuerdo con los enlaces HTML	311 por de los	Miguel Ángel Mitzin Izelo.	Ninguno.
06/11/2020	Obtuve enlaces producto de acuerdo con los enlaces HTML	311 por de los	Obtener enlaces producto de acuerdo con los enlaces HTML	311 por de los	Miguel Ángel Mitzin Izelo.	Ninguno.

**TABLA 12. Scrum diario.**

### 3.3.4.- Desarrollo Sprint 4

#### Tarea 7: Obtener los enlaces HTML de acuerdo con los enlaces de producto extraídos.

Ya almacenadas los enlaces en la base de datos, lo siguiente es extraer su HTML, para ello, en las **FIGURAS 21 y 22** , se muestra el script app\_server.js.

```

1 //Dependencias ->
2 const fs = require('fs');
3 //const Client = require('mariasql');
4 //Dependencias <-
5 //Constantes ->
6 const SERVER_PORT = 7000;
7 const LIMITE_DE_ENLACES = 16;
8 const CARPETA_ENLACES = "PRODUCTOS";//En esta carpeta se guardara el codigo descargado
9 //Constantes <-
10
11 //Cargar server_socket ->
12 eval(fs.readFileSync("server_socket.lib", "utf-8"));
13 //Cargar server_socket <-
14
15 //Conexion con la BD ->
16 var mysql = require('mysql');
17 var c = mysql.createConnection({
18   host    : 'localhost',
19   user    : 'mitzin',
20   password : '1212',
21   database : 'BD_Carrefour_FR'
22 });
23 //Conexion con la BD <-
24 function getEnlaces(ind, callback) {
25   var resultado = [];
26
27   c.query('SELECT * FROM ENLACES_LISTA LIMIT ' + LIMITE_DE_ENLACES, function(err, rows) {
28     if (!err) {
29       //Colocar todos los enlaces en un arreglo
30       for (i = 0; i < rows.length; i++) {
31         resultado.push({
32           "url": rows[i].url

```

**FIGURA 21. Script app\_Server.js.**

```
33     });
34   }
35   if (callback) {
36     callback(resultado);
37   }
38 }
39 });
40 }
41 function saveFile(code, url){
42   fs.readdir(CARPETA_ENLACES, (err, files) => {
43     const Num_ENLACE = files.length + 1;
44     console.log("#####");
45     console.log("Guardando enlace: "+Num_ENLACE+" -> url: "+url);
46     console.log("#####");
47
48     //Guardar enlace en la carpeta
49     fs.writeFileSync(CARPETA_ENLACES+"/enlace"+Num_ENLACE+".html", code, "utf-8");
50
51     //Borrar el enlace en la BD
52     c.query("DELETE FROM ENLACES_LISTA WHERE url = '"+url+"'");
53   });
54 }
```

**FIGURA 22. Script app\_Server.js.**

Lo primero por hacer es activar la vpn para que nos detecten que estamos accediendo desde Francia, para ello se abre una terminal, donde se encuentra un archivo .ovpn. Se ejecuta el .ovpn, se coloca el usuario, contraseña **FIGURA 23** y nos mostrará un mensaje que se cumplió, **FIGURA 24**.

```
root@angel: /home/angelmitzin/Descargas/VPNBook.com-Op...
angelmitzin@angel:~/Descargas/VPNBook.com-OpenVPN-FR1$ sudo su
[sudo] contraseña para angelmitzin:
Lo sentimos, vuelva a intentarlo.
[sudo] contraseña para angelmitzin:
root@angel: /home/angelmitzin/Descargas/VPNBook.com-OpenVPN-FR1# openvpn vpnbook-fr1-tcp80.ovpn
Tue Nov 24 17:20:59 2020 OpenVPN 2.4.7 x86_64-linux-gnu [SSL (OpenSSL)] [LZO]
[LZ4] [EPOLL] [PKCS11] [MH/PKTINFO] [AEAD] built on Sep 5 2019
Tue Nov 24 17:20:59 2020 library versions: OpenSSL 1.1.1f 31 Mar 2020, LZO 2.10
Enter Auth Username: vpnbook
Enter Auth Password: *****
```

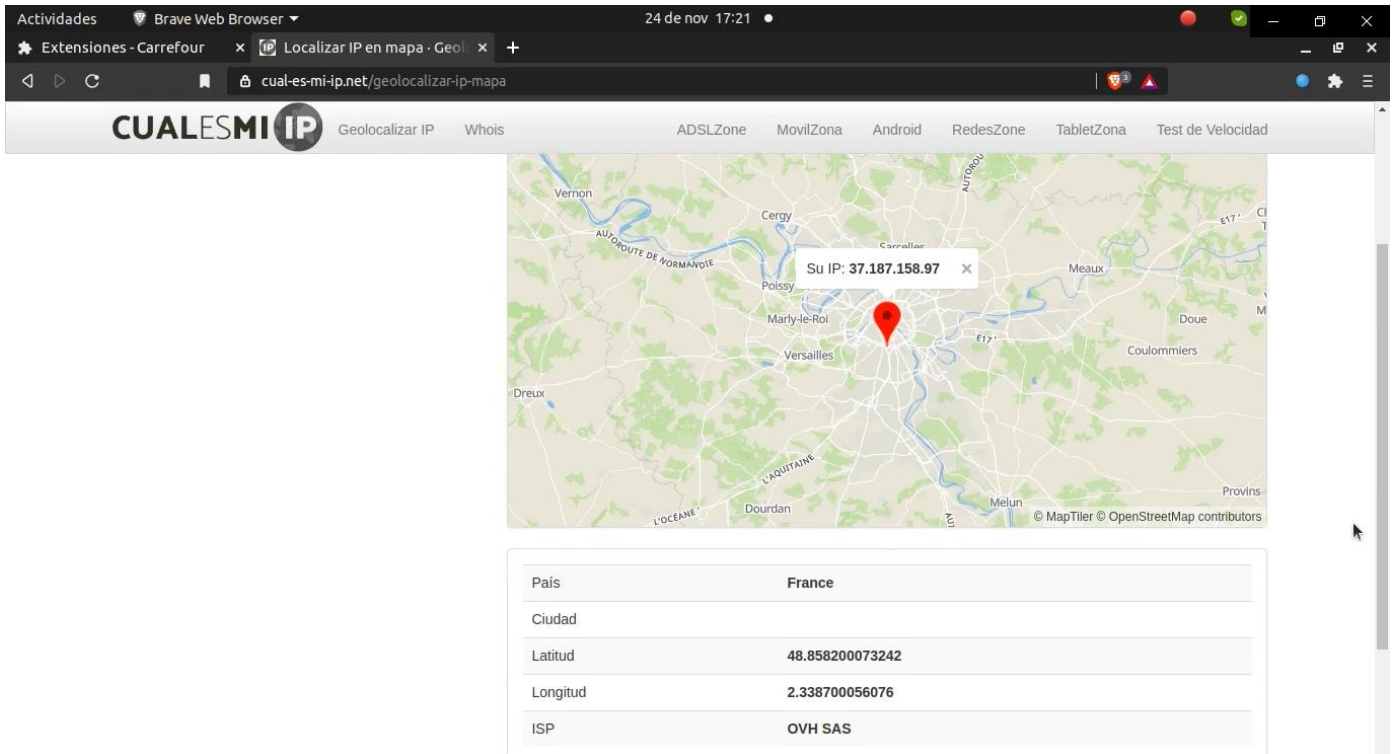
**FIGURA 23. Usuario y contraseña para VPN.**

```
root@angel: /home/angelmitzin/Descargas/VPNBook.com-Op...
Tue Nov 24 17:21:31 2020 OPTIONS IMPORT: route options modified
Tue Nov 24 17:21:31 2020 OPTIONS IMPORT: --ip-win32 and/or --dhcp-option options
modified
Tue Nov 24 17:21:31 2020 OPTIONS IMPORT: peer-id set
Tue Nov 24 17:21:31 2020 OPTIONS IMPORT: adjusting link_mtu to 1627
Tue Nov 24 17:21:31 2020 OPTIONS IMPORT: data channel crypto options modified
Tue Nov 24 17:21:31 2020 Data Channel: using negotiated cipher 'AES-256-GCM'
Tue Nov 24 17:21:31 2020 Outgoing Data Channel: Cipher 'AES-256-GCM' initialized
with 256 bit key
Tue Nov 24 17:21:31 2020 Incoming Data Channel: Cipher 'AES-256-GCM' initialized
with 256 bit key
Tue Nov 24 17:21:31 2020 ROUTE_GATEWAY 192.168.1.254/255.255.255.0 IFACE=wlp2s0
HWADDR=10:63:c8:76:79:a1
Tue Nov 24 17:21:31 2020 TUN/TAP device tun3 opened
Tue Nov 24 17:21:31 2020 TUN/TAP TX queue length set to 100
Tue Nov 24 17:21:31 2020 /sbin/ip link set dev tun3 up mtu 1500
Tue Nov 24 17:21:31 2020 /sbin/ip addr add dev tun3 local 10.12.0.70 peer 10.12.
0.69
Tue Nov 24 17:21:33 2020 /sbin/ip route add 37.187.158.97/32 via 192.168.1.254
Tue Nov 24 17:21:33 2020 /sbin/ip route add 0.0.0.0/1 via 10.12.0.69
Tue Nov 24 17:21:33 2020 /sbin/ip route add 128.0.0.0/1 via 10.12.0.69
Tue Nov 24 17:21:33 2020 /sbin/ip route add 10.12.0.1/32 via 10.12.0.69
Tue Nov 24 17:21:33 2020 Initialization Sequence Completed
```

**FIGURA 24. Conexión de VPN completada.**

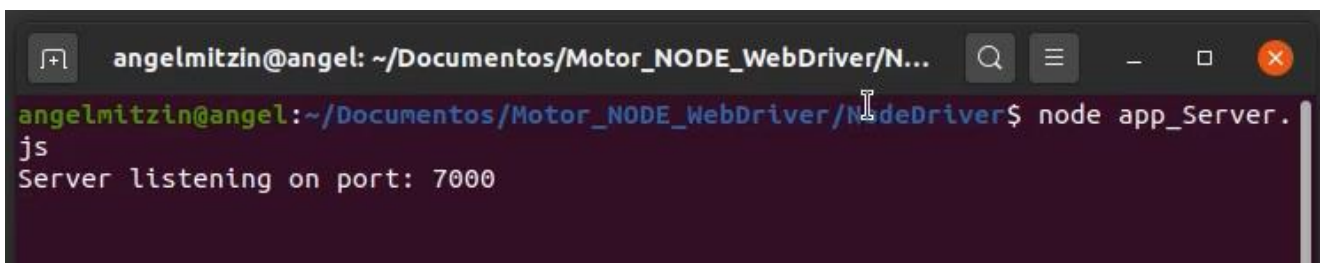


Lo siguiente es comprobar que se está conectando desde Francia, para ello se utiliza la página [www.cual-es-mi-ip.net/geolocalizar-ip-mapa](http://www.cual-es-mi-ip.net/geolocalizar-ip-mapa). **FIGURA 25.**



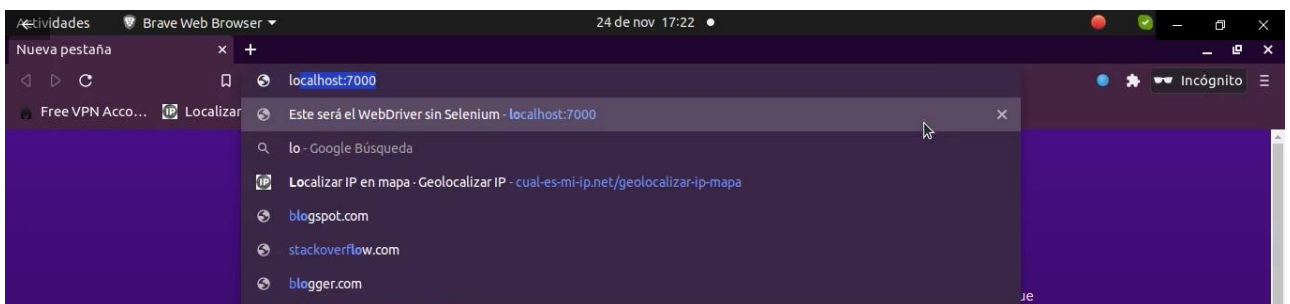
**FIGURA 25. Geolocalización de IP.**

Lo siguiente es ejecutar el script `app_server.js` en la terminal. Al ejecutarlo, el script notifica que puerto se está utilizando **FIGURA 26.**



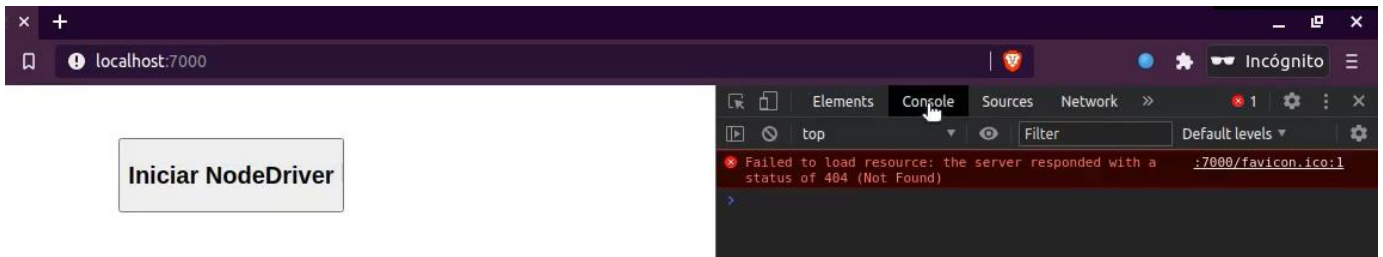
**FIGURA 26. Ejecución del script `app_server.js`**

Después se abre el navegador brave con una pestaña de incógnito y se escribe `localhost:7000` para acceder al servidor **FIGURA 27.**



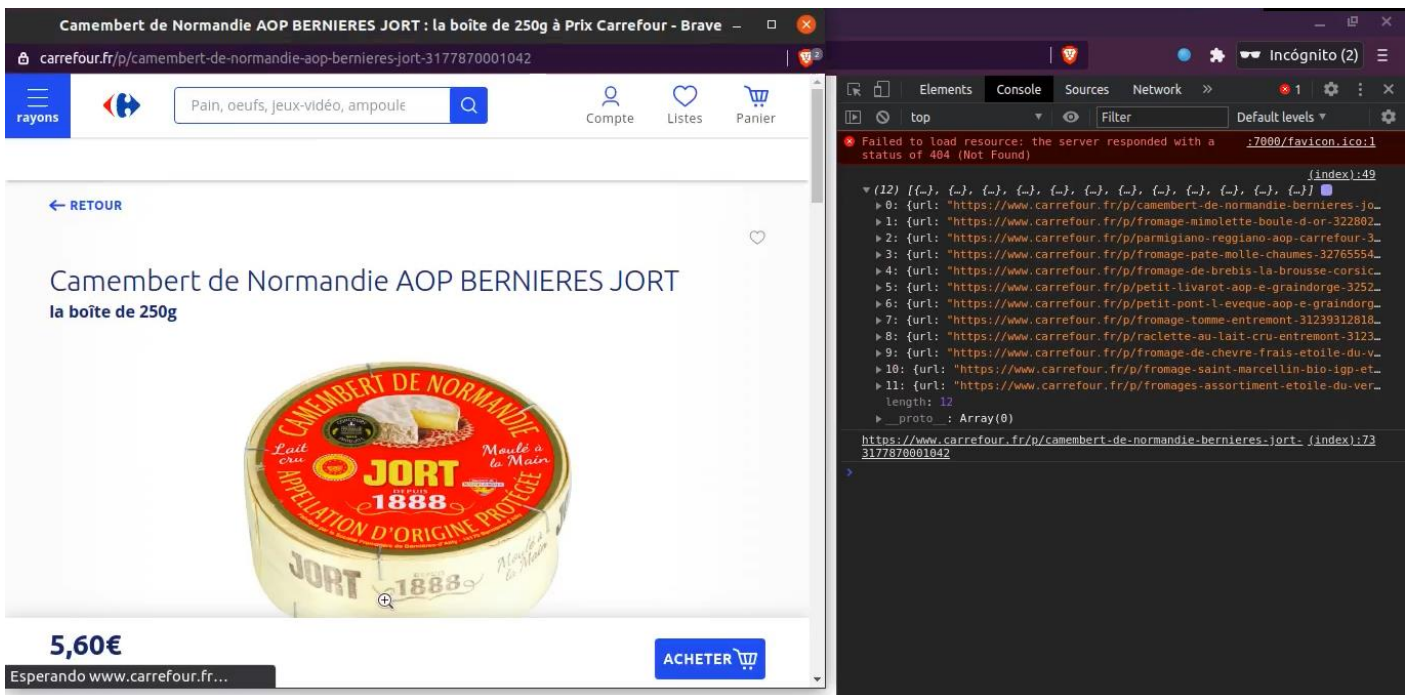
**FIGURA 27. Acceso al puerto 7000 en brave.**

Al acceder nos muestra un botón. Antes de dar clic en el botón abrimos la ventana de inspeccionar, en el apartado consola **FIGURA 28**.



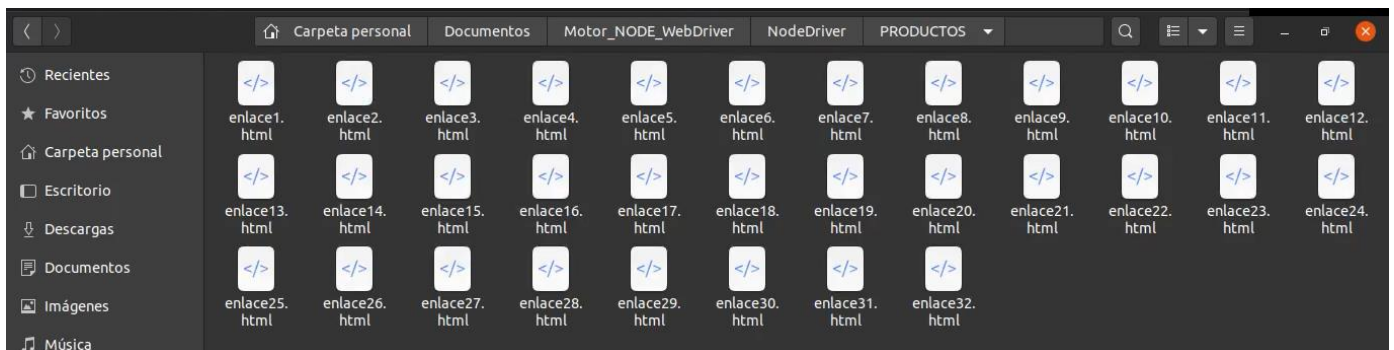
**FIGURA 28. Botón inicio y consola.**

Al dar clic en el botón comenzará a abrir una ventana emergente con el enlace obtenido del producto de la base de datos y del lado de la consola se muestra el array de los enlaces que va a obtener el enlace HTML **FIGURA 29**.



**FIGURA 29. Extracción del enlace a HTML.**

Al finalizar el array, iremos a la carpeta PRODUCTOS, donde se generaron los enlaces HTML **FIGURA 30**.



**FIGURA 30. Enlaces HTML extraídos.**

## Scrum diario

La **TABLA 13** describe lo que ocurrió en cada reunión diaria de acuerdo con las tareas de el sprint 3.

Fecha	¿Qué hice ayer?	¿Qué hice hoy?	Responsable	¿Qué obstáculos obtuve?
10/11/2020	Almacenar los enlaces de producto en la base de datos.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
11/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
12/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
13/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
16/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
17/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
18/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
19/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
20/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
23/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
24/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	El servidor falló, por lo cual, no cargaba los enlaces.
25/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
26/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.

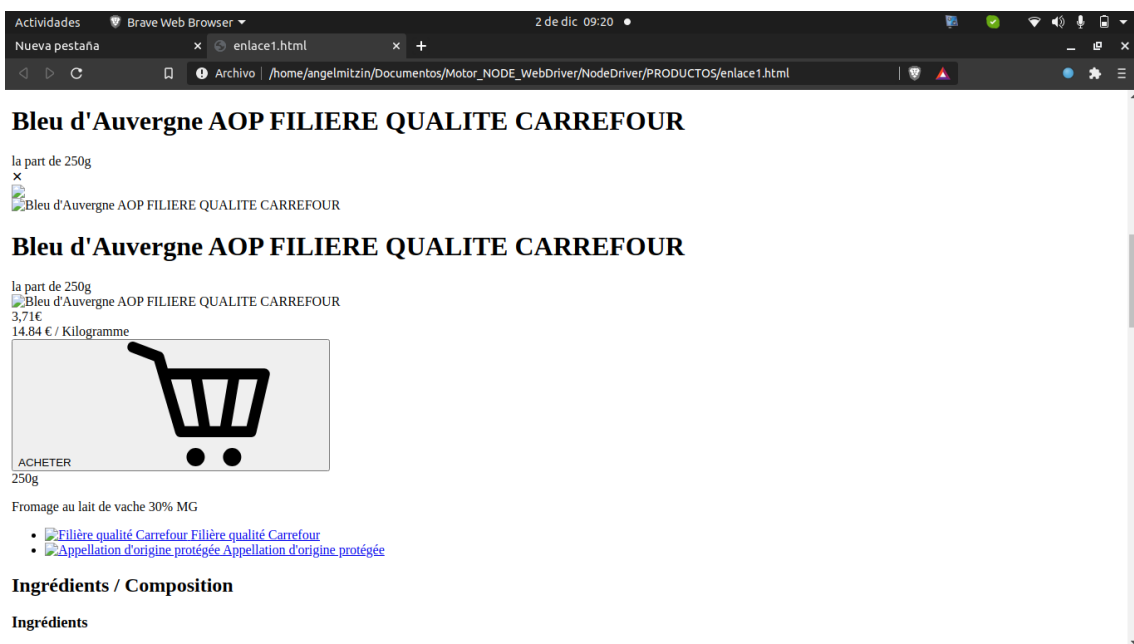
27/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.
30/11/2020	Obtuve 217 enlaces HTML de cada producto.	Obtener 217 enlaces HTML de cada producto.	Miguel Ángel Mitzin Izelo.	Ninguno.

**TABLA 13. Scrum diario.**

### 3.3.5.- Desarrollo Sprint 5

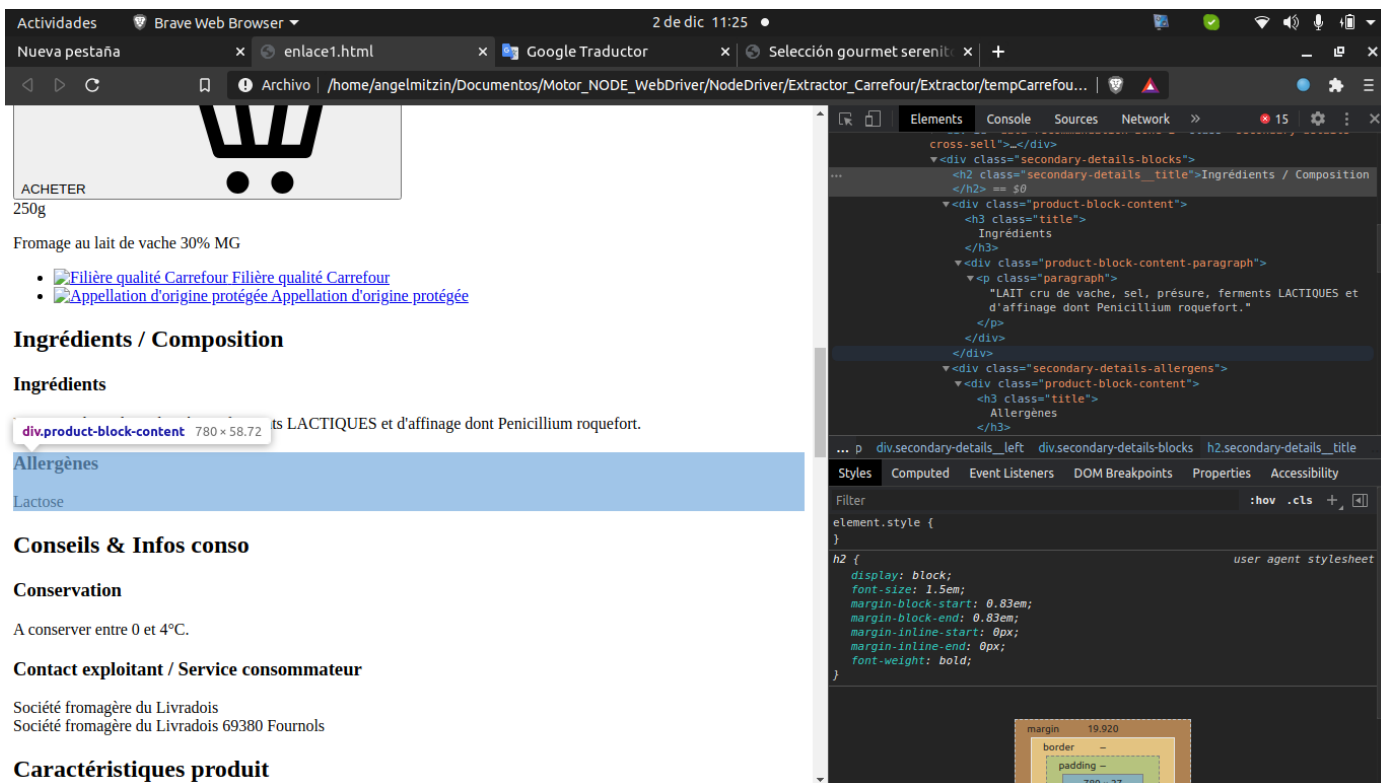
#### Tarea 8: Extracción de información nutrimental de acuerdo con los enlaces HTML de producto obtenidos.

Lo primero por hacer es abrir un enlace HTML de producto **FIGURA 31**

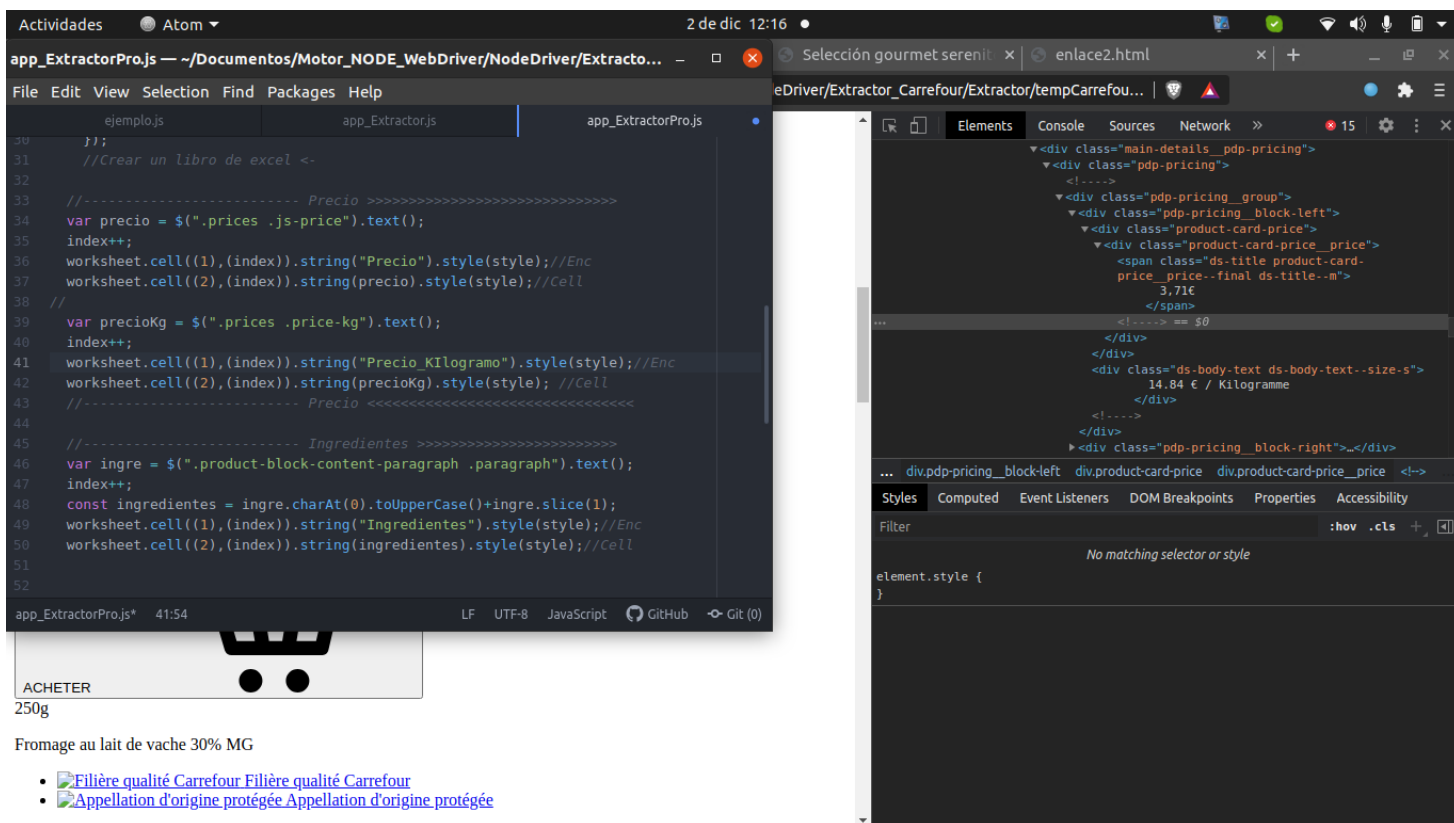


**FIGURA 31. Ejecución de enlace HTML.**

Lo siguiente es abrir la consola del navegador y teclear control más shift más C para identificar los div que son padres para poder extraer la información del enlace y así colocarlo en el script app\_Extractor.js. **FIGURAS 32 y 33**



**FIGURA 32.- Ejecución de enlace HTML y apertura de consola del navegador.**



**FIGURA 33.- Script app\_Extractor.js y consola de navegador.**

Una vez identificado el id o un div o un spam, dependiendo de donde se encuentre la información. Se ejecuta el script app\_Extractor.js. **FIGURA 34**

```

Actividades Terminal 3 de dic 12:06
angelmitzln@angel: ~/Documentos/Motor_NODE_WebDriver/NodeDriver/Extractor_Carrefour/Extractor
angelmitzln@angel:~/Documentos/Motor_NODE_WebDriver/NodeDriver/Extractor_Carrefour/Extractor$ node app_ExtractorPro.js
5,60€
Valeur énergétique (kJ)Valeur énergétique (kcal)Matières grassesDont acides gras saturésGlucidesDont sucresProtéinesSel
Valeurs nutritionnelles

pour 100 g

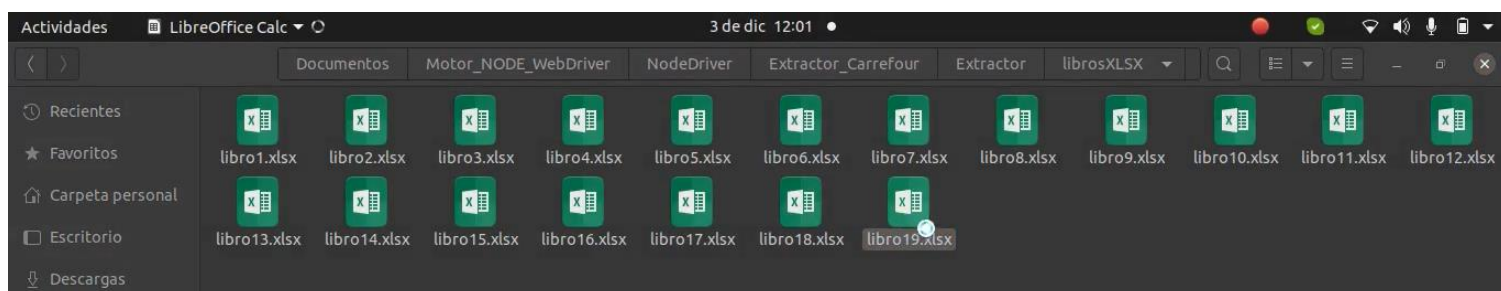
Valeurs nutritionnelles

pour 100 g
Valeur énergétique (kJ) 1150 kJ / 100 g Valeur énergétique (kcal) 277 kcal / 100 g Matières grasses 21.9 g / 100 g Dont acides gras saturés 15.5 g / 100 g Glucides 0 g / 100 g Dont sucres 0 g / 100 g Protéines 20 g / 100 g Sel 1.55 g / 100 g
Valeur énergétique (kJ) 1150 kJ / 100 g Valeur énergétique (kcal) 277 kcal / 100 g Matières grasses 21.9 g / 100 g Dont acides gras saturés 15.5 g / 100 g Glucides 0 g / 100 g Dont sucres 0 g / 100 g Protéines 20 g / 100 g Sel 1.55 g / 100 g
Valeur énergétique (kJ) 1150 kJ / 100 g Valeur énergétique (kcal) 277 kcal / 100 g Matières grasses 21.9 g / 100 g Dont acides gras saturés 15.5 g / 100 g Glucides 0 g / 100 g Dont sucres 0 g / 100 g Protéines 20 g / 100 g Sel 1.55 g / 100 g
Valeur énergétique (kJ) 1150 kJ / 100 g Valeur énergétique (kcal) 277 kcal / 100 g Matières grasses 21.9 g / 100 g Dont acides gras saturés 15.5 g / 100 g Glucides 0 g / 100 g Dont sucres 0 g / 100 g Protéines 20 g / 100 g Sel 1.55 g / 100 g
Valeur énergétique (kcal) 277 kcal / 100 g Valeur énergétique (kcal) 277 kcal / 100 g Matières grasses 21.9 g / 100 g Dont acides gras saturés 15.5 g / 100 g Glucides 0 g / 100 g Dont sucres 0 g / 100 g Protéines 20 g / 100 g Sel 1.55 g / 100 g
Valeur énergétique (kcal) 277 kcal / 100 g Valeur énergétique (kcal) 277 kcal / 100 g Matières grasses 21.9 g / 100 g Dont acides gras saturés 15.5 g / 100 g Glucides 0 g / 100 g Dont sucres 0 g / 100 g Protéines 20 g / 100 g Sel 1.55 g / 100 g
Matières grasses
Matières grasses 21.9 g / 100 g 21.9 g / 100 g
Dont acides gras saturés
Dont acides gras saturés 15.5 g / 100 g 15.5 g / 100 g
Glucides
Glucides 0 g / 100 g 0 g / 100 g
Dont sucres
Dont sucres 0 g / 100 g 0 g / 100 g
Protéines
Protéines 20 g / 100 g 20 g / 100 g

```

**FIGURA 34.- Ejecución de Script app\_Extractor.js.**

Después haber terminado de ejecutarse, nos creará un archivo Excel con la información del producto. **FIGURAS 35 y 36**



**FIGURA 35.- Extracción de información en archivo Excel.**

The image shows the LibreOffice Calc interface with a spreadsheet open. The spreadsheet has columns labeled A through L. Row 1 contains the following text: 'Precio', 'Contenido', 'Contenido', 'Contenido', 'Contenido', 'Contenido', 'Contenido', 'Contenido', 'Contenido', 'Contenido', 'Contenido', 'Contenido', 'Contenido'. Row 2 contains: '3,47€', 'LACT cru de', 'Unité éne', 'Unité éner', 'Valeur éner', 'Matières gr', 'Acides gras', 'Glucides 1 g', 'Dont sucre', 'Protéine 23', 'Sel 1.3 g / 1', 'LactoseLac'. The formula bar shows '= 3,38€'. The window title is 'libro19.xlsx - LibreOffice Calc'.

**FIGURA 36.- Apertura de en archivo Excel.**

## Scrum diario

La **TABLA 14** describe lo que ocurrió en cada reunión diaria de acuerdo con las tareas de el sprint 5.

Fecha	¿Qué hice ayer?	¿Qué hice hoy?	Responsable	¿Qué obstáculos obtuve?
24/11/2020	Obtener 217 enlaces HTML de cada producto.	Identificar un div padre o un id para identificar la información de los productos.	Miguel Ángel Mitzin Izelo.	No encontré la información en un div.
25/11/2020	Obtuve 217 enlaces HTML de cada producto.	Identificar un div padre o un id para identificar la información de los productos.	Miguel Ángel Mitzin Izelo.	No encontré la información en un div.
26/11/2020	Obtuve 217 enlaces HTML de cada producto.	Identificar un div padre o un id para identificar la información de los productos.	Miguel Ángel Mitzin Izelo.	No encontré la información en un div.
27/11/2020	Obtuve 217 enlaces HTML de cada producto.	Identificar un div padre o un id para identificar la información de los productos.	Miguel Ángel Mitzin Izelo.	No encontré la información en un div.
30/11/2020	Obtuve 217 enlaces HTML de cada producto.	Identificar un div padre o un id para identificar la información de los productos.	Miguel Ángel Mitzin Izelo.	No encontré la información en un div.
01/12/2020	Obtuve 217 enlaces HTML de cada producto.	Extraer la información en un archivo excel	Miguel Ángel Mitzin Izelo.	Ninguno.
02/12/2020	Obtuve 217 enlaces HTML de cada producto.	Extraer la información en un archivo excel	Miguel Ángel Mitzin Izelo.	Ninguno.
03/12/2020	Obtuve 217 enlaces HTML de cada producto.	Extraer la información en un archivo excel	Miguel Ángel Mitzin Izelo.	Ninguno.
04/12/2020	Obtuve 217 enlaces HTML de cada producto.	Extraer la información en un archivo excel	Miguel Ángel Mitzin Izelo.	Ninguno.

**TABLA 14. Scrum diario.**





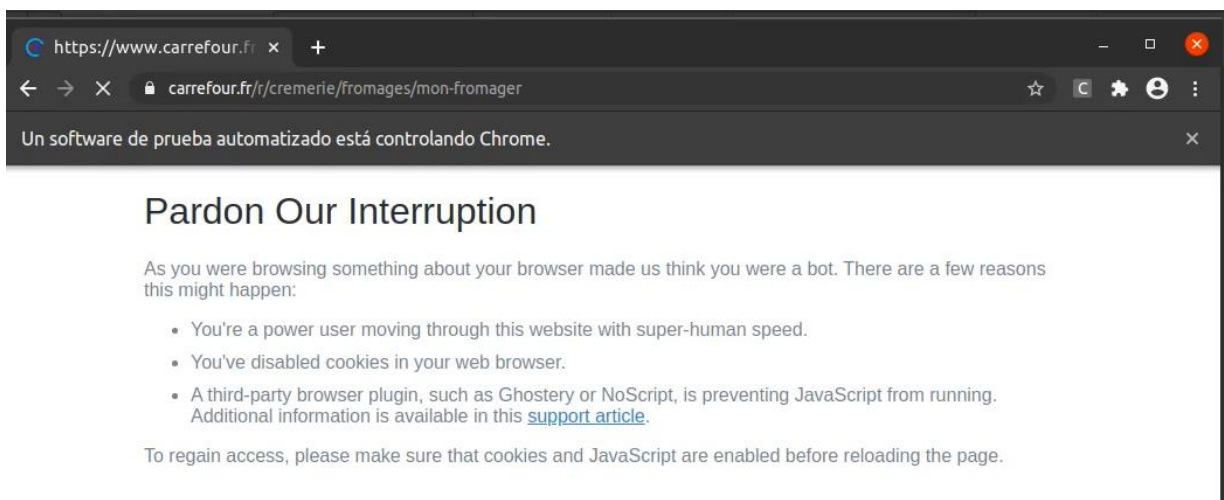
En curso	Terminadas
Ninguna.	Extraer los enlaces de productos procesados.
Ninguna.	Diseñar la base de datos.
Ninguna.	Almacenar las URL.
Obtener los enlaces HTML de acuerdo con las URL extraídas.	Ninguna.
Almacenar los enlaces HTML.	Ninguna.
Obtener los enlaces de producto de acuerdo con los enlaces HTML extraídos.	Ninguna.
Almacenar los enlaces de producto en la base de datos.	Ninguna.
Obtener los enlaces HTML de acuerdo con los enlaces de producto extraídos.	Ninguna.
Extracción de información nutrimental de acuerdo con los enlaces HTML de producto obtenidos.	Ninguna.

**TABLA 15. Product backlog actualizado.**

### 3.4.2.- Sprint 2

En el desarrollo de este sprint el primer problema que tuve fue el no obtener el enlace HTML a tiempo, ya que, el tiempo de espera del código `app_webDriverio` era poco, entonces no me daba tiempo de omitir el captcha por lo cual en el enlace HTML salía dañado.

Otro problema fue que me detectaba Carrefour, entonces ya no me permitía extraer el enlace HTML **FIGURA 39**.



**FIGURA 39. Bloqueo de Carrefour.**

En la **FIGURA 40** se muestra el resultado final de este sprint, el cual es la inserción de los enlaces HTML en la base de datos.

```
MariaDB [BD_Carrefour_FR]> SELECT * FROM ENLACES_PRODUCTOS;
+-----+-----+
| id_producto | url |
+-----+-----+
1 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace1.html |
2 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace2.html |
3 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace3.html |
4 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace4.html |
5 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace5.html |
6 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace6.html |
7 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace7.html |
8 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace8.html |
9 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace9.html |
10 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace10.html |
11 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace11.html |
12 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace12.html |
13 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace13.html |
14 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace14.html |
15 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace15.html |
16 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace16.html |
17 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace17.html |
18 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace18.html |
19 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace19.html |
20 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace20.html |
21 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace21.html |
22 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace22.html |
23 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace23.html |
24 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace24.html |
25 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace25.html |
26 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace26.html |
27 | file:///home/angelmitzin/Documentos/Estancia_ll/webdriver_Carrefour_FRANCIA/WebDriver/Carrefour/EnlacesCarrefourHTML/enlace27.html |
+-----+-----+
```

**FIGURA 40. Inserción de los enlaces HTML en la base de datos.**

En la **TABLA 16** se muestra el producto backlog actualizado de acuerdo con las tareas finalizadas del sprint 2.

En curso	Terminadas
Ninguna.	Extraer los enlaces de productos procesados.
Ninguna.	Diseñar la base de datos.
Ninguna.	Almacenar las URL.
Ninguna.	Obtener los enlaces HTML de acuerdo con las URL extraídas.

Ninguna.	Almacenar los enlaces HTML.
Obtener los enlaces de producto de acuerdo con los enlaces HTML extraídos.	Ninguna.
Almacenar los enlaces de producto en la base de datos.	Ninguna.
Obtener los enlaces HTML de acuerdo con los enlaces de producto extraídos.	Ninguna.
Extracción de información nutricional de acuerdo con los enlaces HTML de producto obtenidos.	Ninguna.

**TABLA 16. Product backlog actualizado.**

### 3.4.3.- Sprint 3

En el desarrollo de este sprint el problema que tuve fue que los enlaces de producto se insertaban duplicados **FIGURA 41**.

```

angelmitzn@angel: ~/Documentos/Estancia_ll/Webdriver_Carrefour_FRANCIA/WebDriver/Carrefour
4007 | https://www.carrefour.fr/p/sushi-saumon-l-sushi-daily-3760199050558
4008 | https://www.carrefour.fr/p/sushi-saumon-l-sushi-daily-3760199050558
4009 | https://www.carrefour.fr/p/sushi-maki-california-sushi-daily-3760199050343
4010 | https://www.carrefour.fr/p/sushi-maki-california-sushi-daily-3760199050343
4011 | https://www.carrefour.fr/p/salade-de-choux-sushi-market-3760228711061
4012 | https://www.carrefour.fr/p/salade-de-choux-sushi-market-3760228711061
4013 | https://www.carrefour.fr/p/salade-d-algues-wakame-comptoir-sushi-3038680020169
4014 | https://www.carrefour.fr/p/salade-d-algues-wakame-comptoir-sushi-3038680020169
4015 | https://www.carrefour.fr/p/california-thon-sushi-daily-3760199052873
4016 | https://www.carrefour.fr/p/california-thon-sushi-daily-3760199052873
4017 | https://www.carrefour.fr/p/menu-san-sushi-daily-3760199059360
4018 | https://www.carrefour.fr/p/menu-san-sushi-daily-3760199059360
4019 | https://www.carrefour.fr/p/salade-d-algues-wakame-sushi-daily-3760199051203
4020 | https://www.carrefour.fr/p/salade-d-algues-wakame-sushi-daily-3760199051203
4021 | https://www.carrefour.fr/p/sushi-saumon-sushi-daily-3760199050480
4022 | https://www.carrefour.fr/p/sushi-saumon-sushi-daily-3760199050480
4023 | https://www.carrefour.fr/p/sushi-crunch-poulet-roll-sushi-daily-3760199054181
4024 | https://www.carrefour.fr/p/sushi-crunch-poulet-roll-sushi-daily-3760199054181
4025 | https://www.carrefour.fr/p/boissons-grenade-the-vert-mangafo-5060040190180
4026 | https://www.carrefour.fr/p/boissons-grenade-the-vert-mangafo-5060040190180
4027 | https://www.carrefour.fr/p/assortiment-de-sushi-combo-one-sushi-daily-3760199052682
4028 | https://www.carrefour.fr/p/assortiment-de-sushi-combo-one-sushi-daily-3760199052682
4029 | https://www.carrefour.fr/p/maki-fromage-sushi-daily-3760199050305
4030 | https://www.carrefour.fr/p/maki-fromage-sushi-daily-3760199050305
4031 | https://www.carrefour.fr/p/maki-fun-mix-sushi-daily-3760278205619
4032 | https://www.carrefour.fr/p/maki-fun-mix-sushi-daily-3760278205619
4033 | https://www.carrefour.fr/p/maki-saumon-sushi-daily-3760199050282
4034 | https://www.carrefour.fr/p/maki-saumon-sushi-daily-3760199050282
4035 | https://www.carrefour.fr/p/ravioli-crevette-sushi-daily-3760199051210
4036 | https://www.carrefour.fr/p/ravioli-crevette-sushi-daily-3760199051210
4037 | https://www.carrefour.fr/p/ravioli-poulet-sushi-daily-3760199051227
4038 | https://www.carrefour.fr/p/ravioli-poulet-sushi-daily-3760199051227
4039 | https://www.carrefour.fr/p/roll-veggie-fromage-sushi-daily-3760278203998
4040 | https://www.carrefour.fr/p/roll-veggie-fromage-sushi-daily-3760278203998
-----
4040 rows in set (0.001 sec)

```

**FIGURA 41. Duplicidad de enlaces producto en la base de datos.**

Al solucionar este problema, el resultado final de este sprint 3 fue la inserción de enlaces por cada producto en la base de datos **FIGURA 42**.

```

3102 | https://www.carrefour.fr/p/sushi-saumon-sushi-daily-3760199050480
3103 | https://www.carrefour.fr/p/sushi-crunch-poulet-roll-sushi-daily-3760199054181
3104 | https://www.carrefour.fr/p/boissons-grenade-the-vert-manga-jo-5060040190180
3105 | https://www.carrefour.fr/p/assortiment-de-sushi-combo-one-sushi-daily-3760199052682
3106 | https://www.carrefour.fr/p/maki-fromage-sushi-daily-3760199050305
3107 | https://www.carrefour.fr/p/maki-fun-mix-sushi-daily-3760278205619
3108 | https://www.carrefour.fr/p/maki-saumon-sushi-daily-3760199050282
3109 | https://www.carrefour.fr/p/ravioli-crevette-sushi-daily-3760199051210
3110 | https://www.carrefour.fr/p/ravioli-poulet-sushi-daily-3760199051227
3111 | https://www.carrefour.fr/p/roll-veggie-fromage-sushi-daily-3760278203998
3112 | https://www.carrefour.fr/p/sushi-duo-mix-sushi-daily-3760199054280
3113 | https://www.carrefour.fr/p/sushi-crunch-saumon-roll-sushi-daily-3760199051425
3114 | https://www.carrefour.fr/p/tartare-de-saumon-sushi-daily-3760199050794
3115 | https://www.carrefour.fr/p/sushi-maki-love-sushi-market-3760228710736
3116 | https://www.carrefour.fr/p/algue-grillee-sushi-daily-3760199054600
3117 | https://www.carrefour.fr/p/cheese-call-roll-sushi-daily-3760199050039
+
117 rows in set (0.003 sec)

```

**FIGURA 42. Inserción de enlaces producto en la base de datos.**

En la **TABLA 17** se muestra el producto backlog actualizado de acuerdo con las tareas finalizadas del sprint 3.

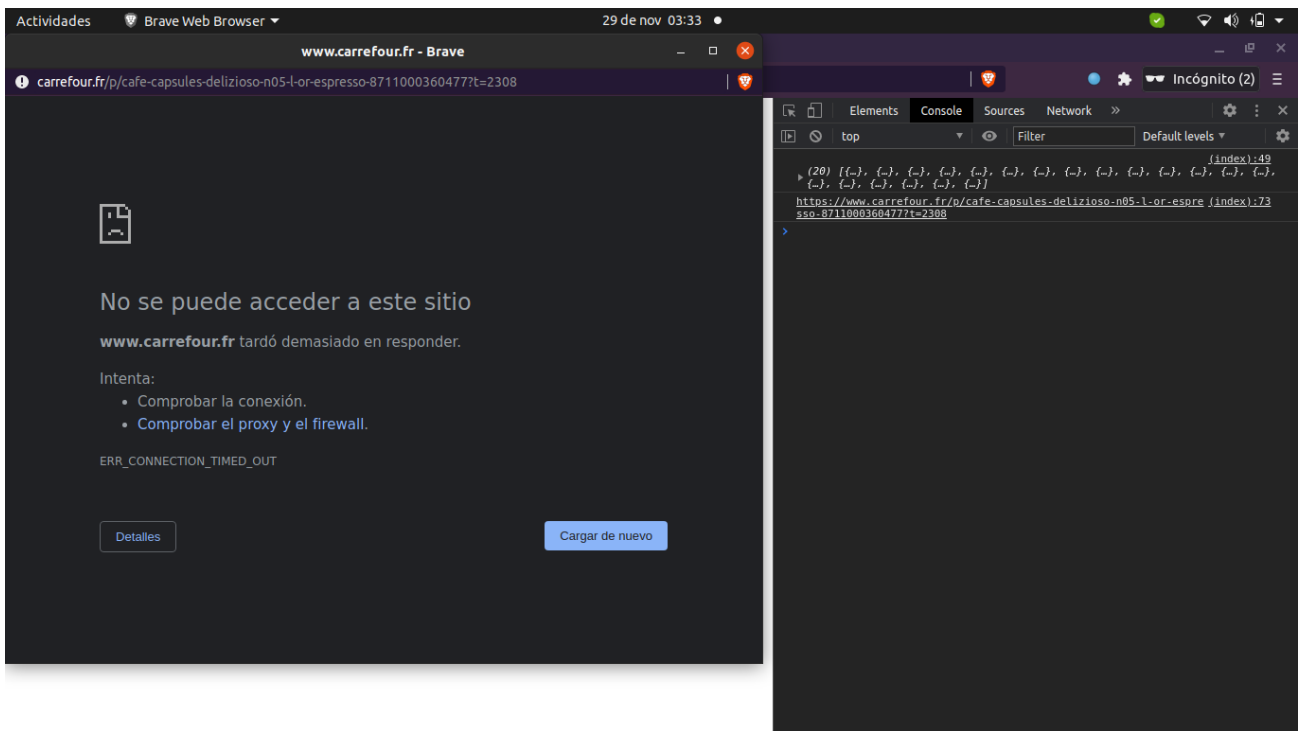
En curso	Terminadas
Ninguna.	Extraer los enlaces de productos procesados.
Ninguna.	Diseñar la base de datos.
Ninguna.	Almacenar las URL.
Ninguna.	Obtener los enlaces HTML de acuerdo con las URL extraídas.
Ninguna.	Almacenar los enlaces HTML.
Ninguna.	Obtener los enlaces de producto de acuerdo con los enlaces HTML extraídos.
Ninguna.	Almacenar los enlaces de producto en la base de datos.

Obtener los enlaces HTML de acuerdo con los enlaces de producto extraídos.	Ninguna.
Extracción de información nutricional de acuerdo con los enlaces HTML de producto obtenidos.	Ninguna.

**TABLA 17. Product backlog actualizado.**

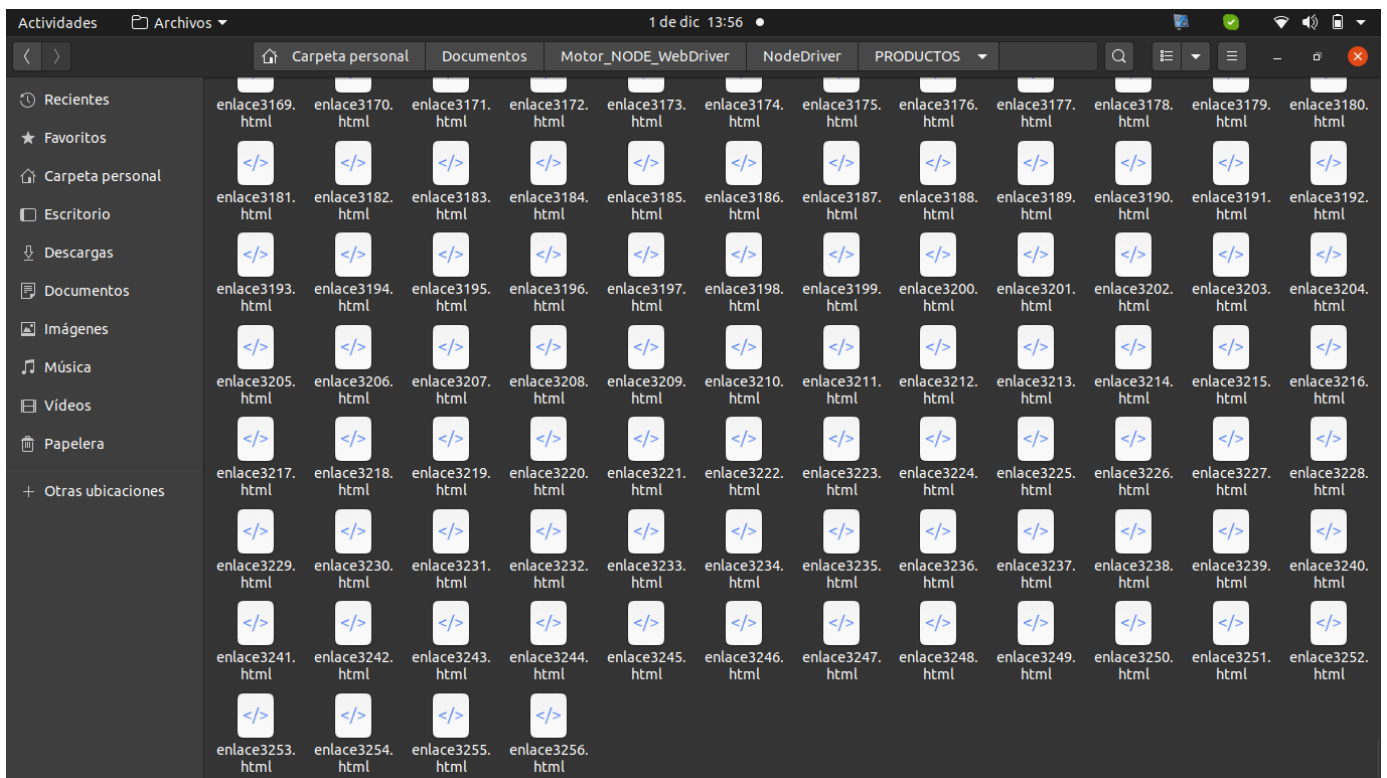
### 3.4.4.- Sprint 4

En el desarrollo de este sprint el problema que tuve fue que el servidor fallo, lo cual provocó que no cargaran los enlaces. **FIGURA 43.**



**FIGURA 43. Error al cargar el enlace.**

El producto final en este sprint fueron 3256 enlaces HTML de producto **FIGURA 44.**



**FIGURA 44. Enlaces HTML de producto.**

En la **TABLA 18** se muestra el producto backlog actualizado de acuerdo con las tareas finalizadas del sprint 4.

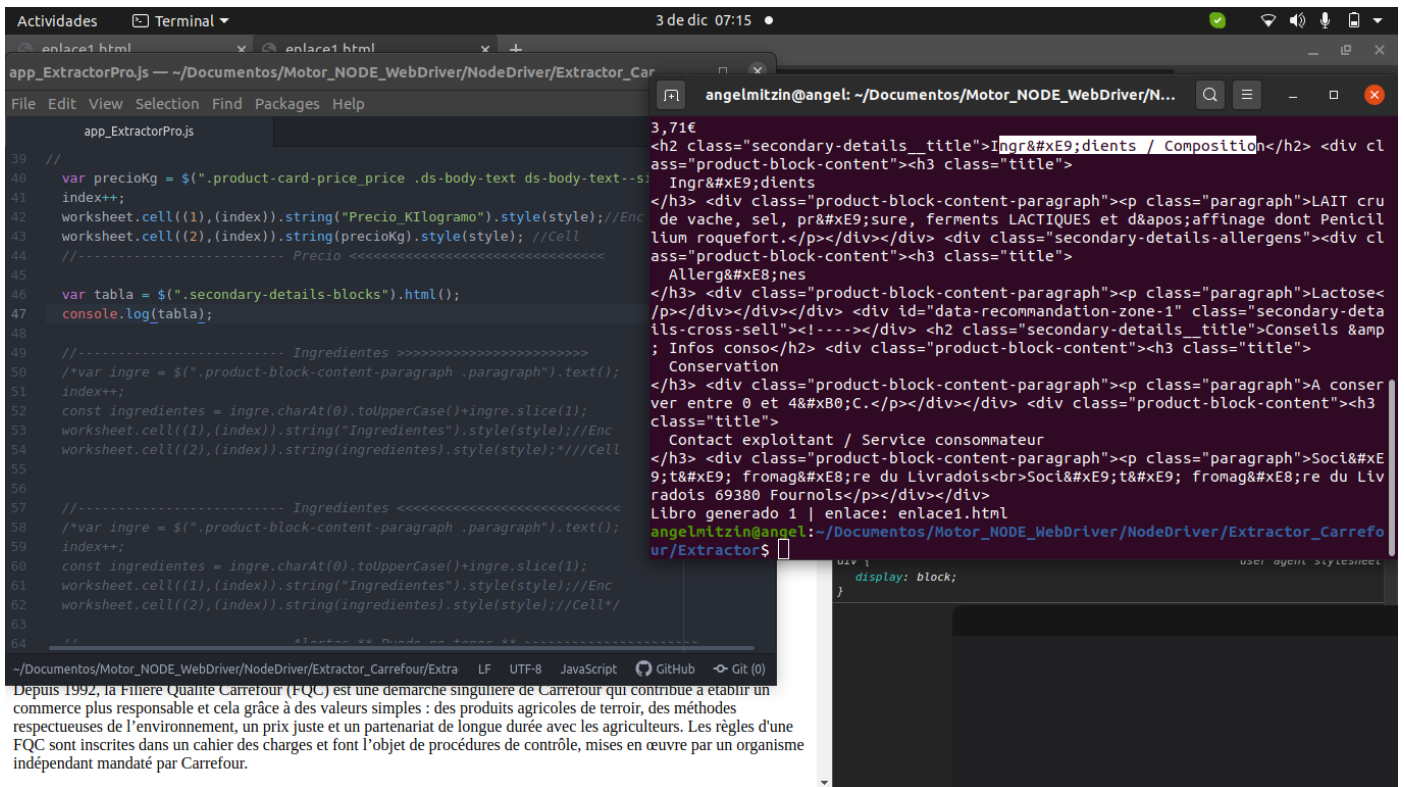
En curso	Terminadas
Ninguna.	Extraer los enlaces de productos procesados.
Ninguna.	Diseñar la base de datos.
Ninguna.	Almacenar las URL.
Ninguna.	Obtener los enlaces HTML de acuerdo con las URL extraídas.
Ninguna.	Almacenar los enlaces HTML.
Ninguna.	Obtener los enlaces de producto de acuerdo con los enlaces HTML extraídos.

Ninguna.	Almacenar los enlaces de producto en la base de datos.
Ninguna	Obtener los enlaces HTML de acuerdo con los enlaces de producto extraídos.
Extracción de información nutricional de acuerdo con los enlaces HTML de producto obtenidos.	Ninguna.

**TABLA 18. Product backlog actualizado.**

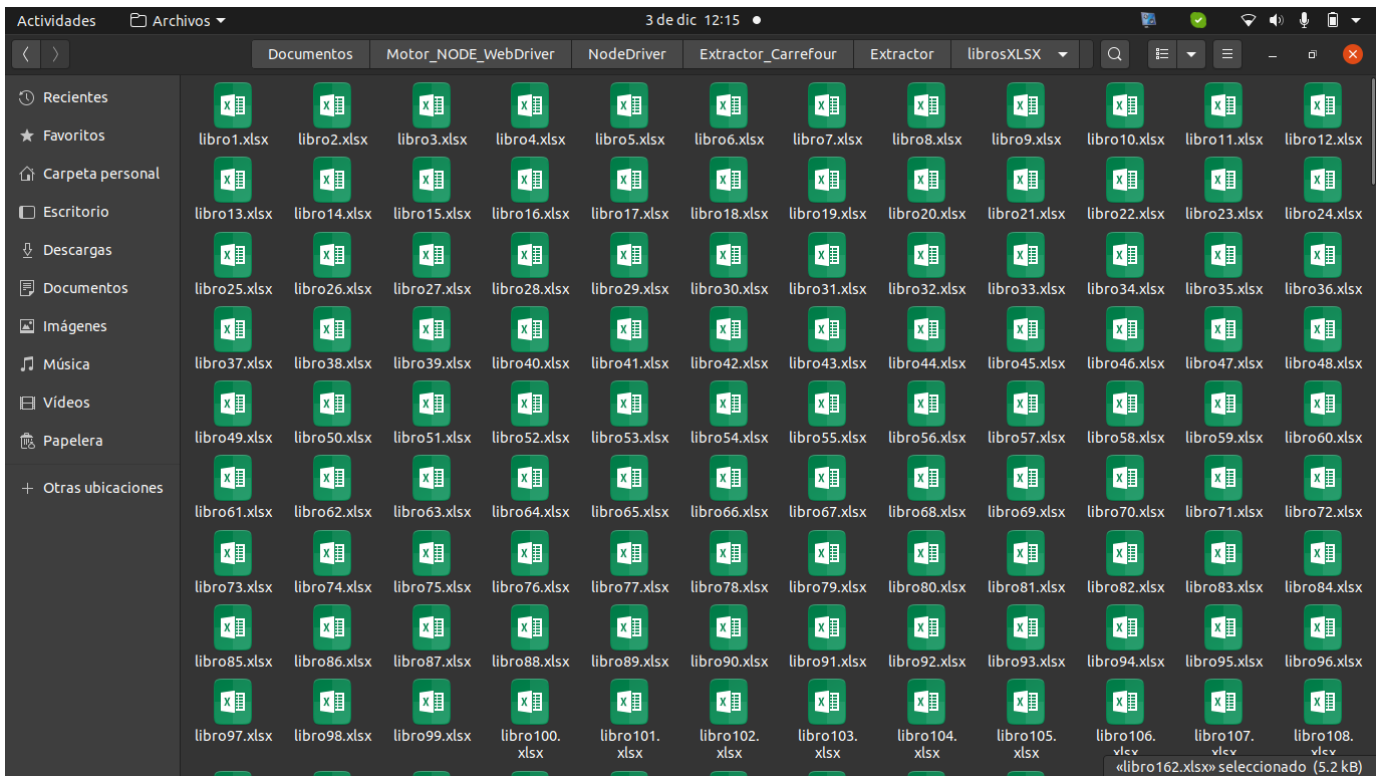
### 3.4.5.- Sprint 5

En el desarrollo de este sprint fue que no encontraba la etiqueta principal para poder extraer la información, por lo tanto, al imprimir me mandaba null o bien sólo me mostraba una parte de código HTML **FIGURA 45.**



**FIGURA 45. Impresión de código HTML.**

El producto final de este sprint fueron los archivos Excel de los enlaces HTML de producto **FIGURA 46.**



**FIGURA 46. Archivo Excel.**

En la **TABLA 19** se muestra el producto backlog actualizado de acuerdo con las tareas finalizadas del sprint 5.

En curso	Terminadas
Ninguna.	Extraer los enlaces de productos procesados.
Ninguna.	Diseñar la base de datos.
Ninguna.	Almacenar las URL.
Ninguna.	Obtener los enlaces HTML de acuerdo con las URL extraídas.
Ninguna.	Almacenar los enlaces HTML.
Ninguna.	Obtener los enlaces de producto de acuerdo con los enlaces HTML extraídos.
Ninguna.	Almacenar los enlaces de producto en la base de datos.



Ninguna	Obtener los enlaces HTML de acuerdo con los enlaces de producto extraídos.
Ninguna.	Extracción de información nutrimental de acuerdo con los enlaces HTML de producto obtenidos.

**TABLA 19. Product backlog actualizado.**

### 3.5.- Lanzamiento

CASO DE PRUEBA	
<b>Código:</b> 1	<b>Número de tarea:</b> 1
<b>Tarea:</b> Extraer las URL de productos procesados de Carrefour Francia.	
<b>Condiciones de ejecución:</b> Sólo extraer enlaces de productos procesados.	
<b>Entrada/Pasos ejecución:</b>	
<ol style="list-style-type: none"> <li>1.- Entrar a la página <a href="http://www.carrefour.fr">www.carrefour.fr</a>.</li> <li>2.- Identificar las categorías de productos procesados.</li> <li>3.- Entrar a las subcategorías de una categoría de producto procesado.</li> <li>4.- Copiar el enlace y pegarlo en un txt.</li> </ol>	
<b>Resultado esperado:</b> Enlace de productos procesados.	
<b>Evaluación de la prueba:</b> Exitosa.	

**TABLA 20. Caso de prueba 1.**

CASO DE PRUEBA	
<b>Código:</b> 2	<b>Número de tarea:</b> 3
<b>Tarea:</b> Almacenar las URL de Carrefour en la base de datos.	
<b>Condiciones de ejecución:</b> Tener completos los enlaces.	
<b>Entrada/Pasos ejecución:</b>	
<ol style="list-style-type: none"> <li>1.- Ejecutar el script <code>app_Insert_BD.js</code></li> <li>2.- Hacer una consulta en la base de dato.</li> </ol>	
<b>Resultado esperado:</b> Inserción de las URL en la base de datos	
<b>Evaluación de la prueba:</b> Exitosa.	

**TABLA 21. Caso de prueba 2.**

CASO DE PRUEBA	
<b>Código:</b> 3	<b>Número de tarea:</b> 4
<b>Tarea:</b> Obtener los enlaces HTML de acuerdo con las URL extraídas.	
<b>Condiciones de ejecución:</b> Abrir y cerrar Google Chrome, ejecutar el servidor selenium.	
<b>Entrada/Pasos ejecución:</b> 1.- Ejecutar el script app_Webdriverio.js 2.- Omitir el captcha de Carrefour. 3.- Verificar que el enlace se haya creado con éxito. 4.- Eliminar el enlace de la base de datos.	
<b>Resultado esperado:</b> Enlace HTML.	
<b>Evaluación de la prueba:</b> Exitosa.	

**TABLA 22. Caso de prueba 3.**

CASO DE PRUEBA	
<b>Código:</b> 4	<b>Número de tarea:</b> 6
<b>Tarea:</b> Obtener los enlaces de producto de acuerdo con los enlaces HTML extraídos.	
<b>Condiciones de ejecución:</b> Tener todos los enlaces HTML en orden numérico.	
<b>Entrada/Pasos ejecución:</b> 1.- Ejecutar el comando app_Cheerio.js 2.- Hacer una consulta para verificar que los enlaces HTML se hayan insertado correctamente y no haya duplicidad.	
<b>Resultado esperado:</b>	
<b>Evaluación de la prueba:</b> Exitosa.	

**TABLA 23. Caso de prueba 4.**

CASO DE PRUEBA	
<b>Código:</b> 5	<b>Número de tarea:</b> 7
<b>Tarea:</b> Obtener los enlaces HTML de acuerdo con los enlaces de producto extraídos.	
<b>Condiciones de ejecución:</b> Tener la VPN activa, navegar de incognito.	
<b>Entrada/Pasos ejecución:</b> 1.- Ejecutar el scrip app_Server.js 2.- Abrir el navegador brave y entrar al localhost:7000 3.- Abrir una ventana de inspeccionar código. 4.- Dar clic en el botón "Iniciar NodeDriver"	
<b>Resultado esperado:</b> Enlaces HTML de los productos	
<b>Evaluación de la prueba:</b> Exitosa.	

**TABLA 24. Caso de prueba 5.**

CASO DE PRUEBA	
<b>Código: 6</b>	<b>Número de tarea: 7</b>
<b>Tarea:</b> Obtener los enlaces HTML de acuerdo con los enlaces de producto extraídos.	
<b>Condiciones de ejecución:</b> Tener la VPN activa, navegar de incognito.	
<b>Entrada/Pasos ejecución:</b> 1.- Ejecutar el scrip app_Server.js 2.- Abrir el navegador brave y entrar al localhost:7000 3.- Abrir una ventana de inspeccionar código. 4.- Dar clic en el botón “Iniciar NodeDriver”	
<b>Resultado esperado:</b> Enlaces HTML de los productos	
<b>Evaluación de la prueba:</b> Exitosa.	

**TABLA 25. Caso de prueba 6.**

CASO DE PRUEBA	
<b>Código: 7</b>	<b>Número de tarea: 7</b>
<b>Tarea:</b> Extracción de información nutrimental de acuerdo con los enlaces HTML de producto obtenidos.	
<b>Condiciones de ejecución:</b> Ejecutar un enlace en el navegador, tener la consola abierta del navegador y de la computadora.	
<b>Entrada/Pasos ejecución:</b> 1.- En la consola identificar un id padre que describa el producto o bien un div. 2.- Hacer una impresión para identificar si se imprime la información requerida. 3.- Si se imprime la información se puede profundizar, esto depende de la etiqueta y la clase. Si no imprime, se debe volver a buscar. 4.- Después de haber identificado, se mete la etiqueta o el id o una clase, esto depende de donde se encuentre la información. 5.- Ejecutar el script app_Extractor.js con la librería para crear archivos Excel.	
<b>Resultado esperado:</b> Archivo Excel	
<b>Evaluación de la prueba:</b> Exitosa.	

**TABLA 26. Caso de prueba 7.**

# 4. Conclusiones y recomendaciones

---

## **Conclusión personal**

Para la elaboración de este proyecto trabaje con herramientas que no había utilizado como node.js, web driver e implementar una VPN. Pero aun lo que me motivo fue el ingeniero Ricardo Marcial, quien se dedicó a capacitarme, ya que, él fue el que me regalo de su tiempo, siempre estuvo pendiente de mi en todo y me enseñó muchas cosas con JavaScript. Además, la elaboración de esta técnica me impresionó mucho porque desconocida que se podía obtener así la información de manera más rápida, con esto, espero aprender nuevas técnicas de abstracción de información con otro tipo de herramientas.

## **Conclusión del proyecto.**

Le veo mucho potencial a esta técnica, ya que, uno como persona puede obtener mucha información y analizarla para determinar desde los gustos de personas, en cuestión de comida, productos, obtener información de los productos que hay en el mercado o bien, poder ofrecer este tipo de información a empresas para que ellos hagan sus análisis. Espero y muchas personas puedan aprender este tipo de técnicas porque es muy impresionante como se puede lograr esto con sólo aprender a programar.

# 5. Referencias bibliográficas

---

- [1] Bahit, E. (s. f.). Scrum and extreme programming. Recuperado 29 de octubre de 2020, de <http://umh2818.edu.umh.es/wp-content/uploads/sites/884/2016/02/Scrum-y-eXtrem-Programming-para-programadores.pdf>
- [2] URL: Que es MariaDB y mejoras sobre MySQL | Nerion. (s. f.). Recuperado 29 de octubre de 2020, de <https://www.nerion.es/soporte/que-es-mariadb-y-mejoras-sobre-mysql/>
- [3] URL: Selenium. (s. f.). Recuperado 5 de noviembre de 2020, de <https://magmax.org/blog/selenium-y-qa-automation/#:~:text=Selenium%20IDE,-Consiste%20en%20un&text=Ventajas%3A%20%2D%20Es%20muy%20f%C3%A1cil%20para,las%20pruebas%20as%C3%AD%20es%20lento.>
- [4] Stack Overflow Contributors. (s. f.). Aprendizaje Node.js (1.a ed., Vol. 1). Recuperado de <https://manual-informatica.com/download-file.html>
- [5] Ubuntu. (s. f.). Recuperado 5 de noviembre de 2020, de <http://www.uls.edu.sv/pdf/ubuntu.pdf>
- [6] Lopez, M. D. (s. f.). Introduccion-a-Nodejs.pdf. Recuperado 5 de noviembre de 2020, de [https://www.academia.edu/38559025/Introduccion\\_a\\_Nodejs\\_pdf](https://www.academia.edu/38559025/Introduccion_a_Nodejs_pdf)
- [7] Sánchez, J. (s. f.). Java. Recuperado 5 de noviembre de 2020, de <http://jorgesanchez.net/manuales/viejos/fpr/Java.pdf>
- [8] Kniberg, H. (s. f.). Scrum y XP desde las trincheras. Recuperado 5 de noviembre de 2020, de <http://www.proyectalis.com/wp-content/uploads/2008/02/scrum-y-xp-desde-las-trincheras.pdf>
- [9] URL: Ventajas de organizar las dependencias usando Node.js/npm. (2018, 13 febrero). Recuperado 12 de noviembre de 2020, de <https://es.stackoverflow.com/questions/4163/ventajas-de-organizar-las-dependencias-usando-node-js-npm>
- [10] URL: Araujo, D. S. (s. f.). Waterfall vs Agile. Recuperado 12 de noviembre de 2020, de <https://blog.softtek.com/es/waterfall-vs-agile>



Universidad Politécnica de Puebla  
Ingeniería en Informática

*Miguel Ángel Mitzin Izelo*  
*César Hernández Rosete*  
*Rebeca Rodríguez Huesca*

Este documento se distribuye para los términos de la  
Licencia 2.5 Creative Commons (CC-BC-NC-ND 2.5 MX)