

Il Porco Scraper

Instructivo del programa

v. 4.6



1. ¿Qué es Il Porco Scraper?

Il Porco Scraper es un desarrollo de Software cuyo objetivo es realizar automáticamente la búsqueda de los precios de los artículos que están en la base de datos de Il Porco, con el motivo de cambiarlos en el sistema.

El contexto de inflación sostenida que atraviesa el país obliga a Il Porco a cambiar los precios de los productos que compra en sus sistema semanalmente, sin embargo buscar cada uno de estos precios se convierte en una tarea tediosa, repetitiva y que consume mucho tiempo. Por ende, se propone como solución un sistema informático que realice estas búsquedas de precios en sus respectivas páginas de manera completamente automática.

En su versión actual, el sistema cuenta con la capacidad de búsqueda dentro de las páginas de los siguientes proveedores: Maxiconsumo, Oscar David, Andina y La Serenísima.

Utilizando esta herramienta, el personal administrativo de Il Porco se ahorra la búsqueda de más de 700 artículos que pertenecen a estos proveedores.

2. ¿Cuáles son las componentes que hacen al Scraper?

Atención: La siguiente información es relevante para el funcionamiento de la aplicación a nivel de Software, pero no lo es desde el punto de vista de personal administrativo o usuario de la aplicación.

El Scraper está compuesto por 7 componentes que interactúan entre sí y tienen distintas funcionalidades:

- La GUI o Interfaz gráfica, que a la vez llama a ejecución al Operador, al Administrador de Credenciales y al Administrador de Tabla Intermedia.
- El Operador, que es el encargado del algoritmo general y la administración de los procesos de búsqueda de los datos en las páginas llamando a componentes más especializados, como el Lector y los Valuadores
- El Lector, cuyo objetivo es leer, interpretar y editar los archivos de Excel donde se toma la información y posteriormente guardar la información de forma organizada en el archivo de Excel resultante.
- Los Valuadores (uno por distribuidor), encargados de recibir los datos interpretados por el lector y buscar los artículos en sus respectivas páginas con sus respectivos algoritmos especializados.
- El Sesionador, que es el encargado de abrir las sesiones en las páginas de los proveedores para que la información que el programa encuentre sea la misma que se encuentra habiendo iniciado sesión en cada página.
- El Administrador de Credenciales, cuya función es cargar los datos que sirven para abrir las sesiones correspondientes a cada proveedor y editar esa información, encriptando las contraseñas a su vez.
- El Administrador de Tabla Intermedia, que sirve para editar la Tabla Intermedia, una parte fundamental del sistema que asocia el código de Il Porco con el código interno del artículo en cada página según su proveedor, para poder encontrarlo posteriormente.

3. Cómo realizar una búsqueda de precios con Il Porco Scraper

Se necesitan dos elementos principales para poder hacer la búsqueda de precios. Estos elementos son: Los ID de sesión de Maxiconsumo, Andina, Oscar David y La Serenísima, y un archivo EXCEL que contenga todos los elementos cuyo precio se desea buscar en una disposición igual a la que provee el sistema de Il Porco.

Pasos a seguir

1. Crear un archivo EXCEL que contenga todos los artículos cuyo precio debe ser buscado, y su formato debe ser el mismo que el del sistema de precios de Il Porco. Basta con copiar en el sistema y pegar en un Excel.
2. Guardar el archivo con el nombre que se desee en la ruta de la aplicación, en la carpeta llamada “archivos”. Otra alternativa es utilizar la funcionalidad del programa de buscar el archivo por una ruta alterna, desde la pestaña de “Opciones”.
3. Ejecutar el programa y correr el sesionador para buscar automáticamente los ID de Sesión de los proveedores que se pueda (Maxiconsumo, Andina y La Serenísima).
4. Buscar manualmente el ID de Sesión de Oscar David y pegarlo en el lugar correspondiente.
5. Colocar en el campo indicado el nombre del archivo de Excel que creamos previamente (No es necesario en caso de elegir una ruta alterna) y correr el programa.
6. Esperar a que la búsqueda de precios termine. Un nuevo archivo EXCEL actualizado será creado en el directorio de archivos, y se abrirá automáticamente.

Creación del archivo EXCEL

Primero, en el sistema de Il Porco hay que filtrar los artículos que se quieren buscar por los parámetros que se desee (en este caso, Maxiconsumo como proveedor).

SKU	Cód.	Artículo	Marca	Rubro	Sucursal	Proveedores	Costo anterior desde	Costo anterior S/IVA	Costo anterior C/IVA	Costo nuevo fecha	Costo nuevo S/IVA	Costo nuevo C/IVA	Variación	Act.
070330703629	3222	MAQUINA AFEITAR BIC SENSITIVE	BIC	CUIDADO PERS.	ALVEAR	MAXICONSUMO	03/01/2024	257,12	311,12	23/01/2024	257,12	311,12	0,00%	
0703307031806	3224	MAQUINA DE AFEITAR BIC SOLEIL X1 UN	BIC	CUIDADO PERS.	ALVEAR	MAXICONSUMO	21/12/2023	527,09	637,78	23/01/2024	527,09	637,78	0,00%	
070330909229	3550	ENCENDEDOR BIC MAXI RECARGABLE X1 UN	BIC	OTROS	ALVEAR	MAXICONSUMO	08/12/2023	725,36	877,69	23/01/2024	725,36	877,69	0,00%	
3014260846725	2104	CEPILLO DE DIENTES ORAL-B PRO X1 UN	ORAL-B	CUIDADO PERS.	ALVEAR	MAXICONSUMO	22/11/2023	272,64	329,89	23/01/2024	272,64	329,89	0,00%	
4005080257584	3458	ANTITRANSPIRANTE NIVEA DRY COMFORT ROLL ON X51 GR	NIVEA	CUIDADO PERS.	ALVEAR	MAXICONSUMO	02/01/2024	847,11	1.025,00	23/01/2024	847,11	1.025,00	0,00%	
4005080910328	3217	ANTITRANSPIRANTE NIVEA POWDER COMFORT X150 ML	NIVEA	CUIDADO PERS.	ALVEAR	MAXICONSUMO	02/01/2024	1.533,88	1.855,99	23/01/2024	1.533,88	1.855,99	0,00%	
40050900071279	2994	ESPUÑA DE AFEITAR NIVEA HIDRATANTE X200 ML	NIVEA	CUIDADO PERS.	ALVEAR	MAXICONSUMO	02/01/2024	1.614,88	1.954,38	23/01/2024	1.614,88	1.954,38	0,00%	
40050900122186	3220	ANTITRANSPIRANTE NIVEA PROTECTOCARE X150 ML	NIVEA	CUIDADO PERS.	ALVEAR	MAXICONSUMO	02/01/2024	1.533,88	1.855,99	23/01/2024	1.533,88	1.855,99	0,00%	
40050900122186	3220	ANTITRANSPIRANTE NIVEA PROTECTOCARE X150 ML	NIVEA	CUIDADO PERS.	SAN RAFAEL	MAXICONSUMO SA	02/01/2024	1.533,88	1.855,99	23/01/2024	1.533,88	1.855,99	0,00%	
40050900143969	2995	ESPUÑA DE AFEITAR NIVEA SENSITIVE X200 ML	NIVEA	CUIDADO PERS.	ALVEAR	MAXICONSUMO	02/01/2024	1.615,19	1.954,38	23/01/2024	1.615,19	1.954,38	0,00%	
40050900389593	3540	CREMA CORPORAL NIVEA CEREZA 400ML	NIVEA	CUIDADO PERS.	ALVEAR	MAXICONSUMO	28/12/2023	3.222,72	3.899,49	23/01/2024	3.222,72	3.899,49	0,00%	
40050900301323	3216	ANTITRANSPIRANTE NIVEA AGUA DE COCO X150 ML	NIVEA	CUIDADO PERS.	ALVEAR	MAXICONSUMO	02/01/2024	1.533,88	1.855,99	23/01/2024	1.533,88	1.855,99	0,00%	
40050900301354	3218	ANTITRANSPIRANTE NIVEA BLACK&WHITE X150 ML	NIVEA	CUIDADO PERS.	ALVEAR	MAXICONSUMO	02/01/2024	1.533,88	1.855,99	23/01/2024	1.533,88	1.855,99	0,00%	
40050900301354	3218	ANTITRANSPIRANTE NIVEA BLACK&WHITE X150 ML	NIVEA	CUIDADO PERS.	SAN RAFAEL	MAXICONSUMO SA	02/01/2024	1.533,88	1.855,99	23/01/2024	1.533,88	1.855,99	0,00%	
40050900395110	3471	CREMA CORPORAL NIVEA KARITE PIEL SECA X125 ML	NIVEA	CUIDADO PERS.	ALVEAR	MAXICONSUMO	02/01/2024	1.188,25	1.437,78	23/01/2024	1.188,25	1.437,78	0,00%	
6902004095218	2894	ENCENDEDOR CANDELA X1 UN	CANDELA	OTROS	ALVEAR	MAXICONSUMO	27/12/2023	247,85	299,90	23/01/2024	247,85	299,90	0,00%	
6910021007206	1648	CEPILLO DE DIENTES COLGATE PREMIER CLEAN X1 UN	COLGATE	CUIDADO PERS.	SAN RAFAEL	MAXICONSUMO SA	28/12/2023	388,35	469,90	23/01/2024	388,35	469,90	0,00%	
6910021007206	1648	CEPILLO DE DIENTES COLGATE PREMIER CLEAN X1 UN	COLGATE	CUIDADO PERS.	ALVEAR	MAXICONSUMO	28/12/2023	388,35	469,90	23/01/2024	388,35	469,90	0,00%	
7500405138505	3223	MAQUINA DE AFEITAR GILLETTE VENUS SKIN&GLOR X1U	GILLETTE	CUIDADO PERS.	ALVEAR	MAXICONSUMO	16/01/2024	1.322,23	1.599,90	23/01/2024	1.322,23	1.599,90	0,00%	
7500405138505	3223	MAQUINA DE AFEITAR GILLETTE VENUS SKIN&GLOR X1U	GILLETTE	CUIDADO PERS.	SAN RAFAEL	MAXICONSUMO SA	16/01/2024	1.322,23	1.599,90	23/01/2024	1.322,23	1.599,90	0,00%	
7500405138769	2950	MAQUINA DE AFEITAR GILLETTE PRESTOBARBA 2 PLUS	GILLETTE	CUIDADO PERS.	ALVEAR	MAXICONSUMO	23/01/2024	826,33	999,86	23/01/2024	826,33	999,86	0,00%	
7500405219848	9591	ESPUÑA DE AFEITAR GILLETTE PIEL SENSIBLE 175GRS	GILLETTE	CUIDADO PERS.	ALVEAR	MAXICONSUMO	16/01/2024	2.479,26	2.999,90	23/01/2024	2.479,26	2.999,90	0,00%	
7501056344092	3102	SHAMPOO CLEAR CONTROL CAIDA X400ML	CLEAR	CUIDADO PERS.	ALVEAR	MAXICONSUMO	21/12/2023	2.066,03	2.499,90	23/01/2024	2.066,03	2.499,90	0,00%	
7501056346157	3214	CREMA CORPORAL DOVE NOU BODY CARE ESSENCIAL X400ML	DOVE	CUIDADO PERS.	ALVEAR	MAXICONSUMO	28/12/2023	1.462,39	1.769,49	23/01/2024	1.462,39	1.769,49	0,00%	

Luego, los datos se copian del sistema y se pegan desde la celda número 1, fila 1 columna A, hacia adelante. Deberá quedar de la siguiente forma:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	70330703629	3222	MAQUINA AFEITAR BIC SENSITIVE	BIC	CUIDADO PERSONA	ALVEAR	MAXICONSUMO	3/1/2024	257,12	311,12	#####	257,12	311,12	0,00%									
2	70330731806	3224	MAQUINA DE AFEITAR BIC SOLEIL X1 UN	BIC	CUIDADO PERSONA	ALVEAR	MAXICONSUMO	#####	527,09	637,78	#####	527,09	637,78	0,00%									
3	70330909229	3550	ENCENDEDOR BIC MAXI RECARGABLE X1 UN	BIC	OTROS	ALVEAR	MAXICONSUMO	#####	725,36	877,69	#####	725,36	877,69	0,00%									
4	3,01426E+12	2104	CEPILLO DE DIENTES ORAL-B PRO X1 UN	ORAL-B	CUIDADO PERSONA	ALVEAR	MAXICONSUMO	#####	272,64	329,89	#####	272,64	329,89	0,00%									
5	4,00581E+12	3458	ANTITRANSPIRANTE NIVEA DRY COMFORT ROLL ON X51 GR	NIVEA	CUIDADO PERSONA	ALVEAR	MAXICONSUMO	2/1/2024	847,11	1.025,00	#####	847,11	1.025,00	0,00%									

Ahora podemos guardar el archivo, con el nombre que se desee. Es importante recordar este nombre porque será otro de los datos a ingresar en el sistema.

Una vez guardado el archivo se puede mover a la carpeta de “Archivos” ubicada en el directorio del programa (por defecto, en Archivos de Programa/Il Porco

Scraper/archivos) o abrir desde el programa como una ruta alternativa desde la pestaña de “Opciones”.

Apertura de sesiones utilizando el sesionador

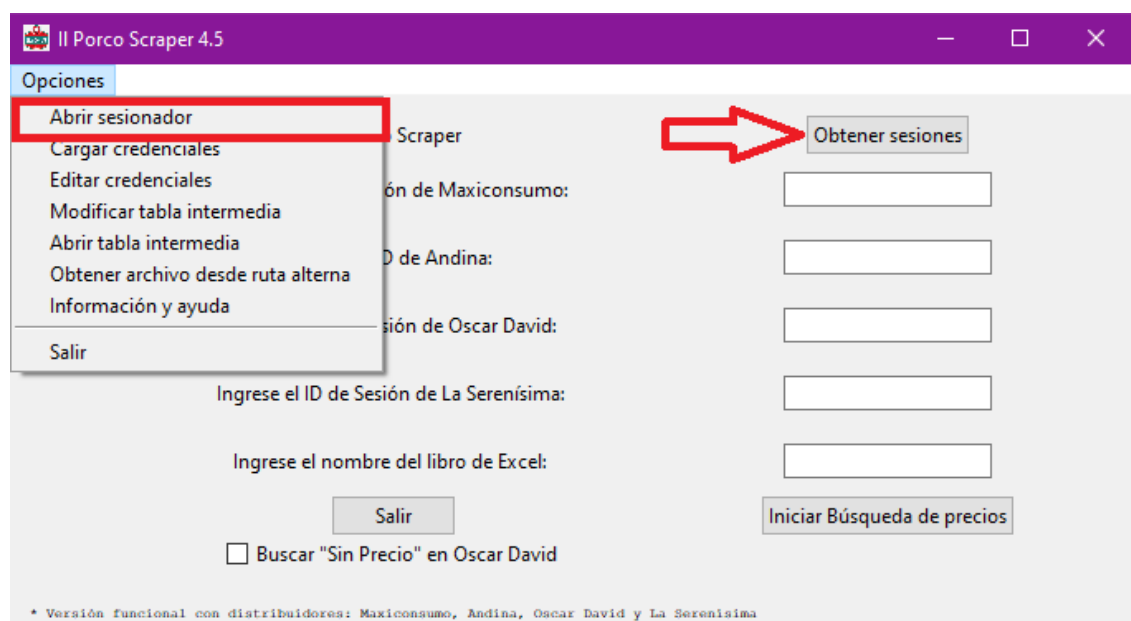
Para que el programa encuentre los precios de las páginas tiene que iniciar sesión con las cuentas empresariales en dichos sitios web.

Generalmente, cuando un usuario envía las credenciales (es decir, el nombre de usuario o correo electrónico y la contraseña) a una página y se autentica correctamente, la misma devuelve una clave de sesión que sirve como llave para acceder a todas las funcionalidades que otorga una sesión iniciada. Este tipo de clave se conoce como **Session ID**, y para que el programa pueda acceder a la información restringida para el usuario común de las páginas de proveedores, necesita una de estas llaves por cada proveedor.

Nótese que los Session ID tienen cierta vida útil que suele ser de unas pocas horas, por lo que cada vez que se vaya a buscar precios lo recomendable es darle al programa una nueva Session ID por cada proveedor.

La herramienta de “Sesionador” del Il Porco Scraper le permite al usuario abrir las sesiones y recolectar los Session ID automáticamente en el caso de 3 de los 4 proveedores que es capaz de buscar. Más abajo veremos exactamente cómo se busca un Session ID de forma manual.

Para utilizar el sesionador del Il Porco Scraper, damos click en el botón “Obtener Sesiones” que está sobre los campos a completar con los Session ID, o en su defecto vamos al menú de opciones y elegimos “Abrir Sesionador”.



Esto abrirá una nueva ventana con las opciones de sesión.

Sesionador de Il Porco Scraper

Credenciales de Maxiconsumo

Correo Electrónico:

Contraseña:

Credenciales de La Serenísima

Correo Electrónico:

Contraseña:

* La sesión de Oscar David debe ser abierta de forma manual

Abrir sesiones

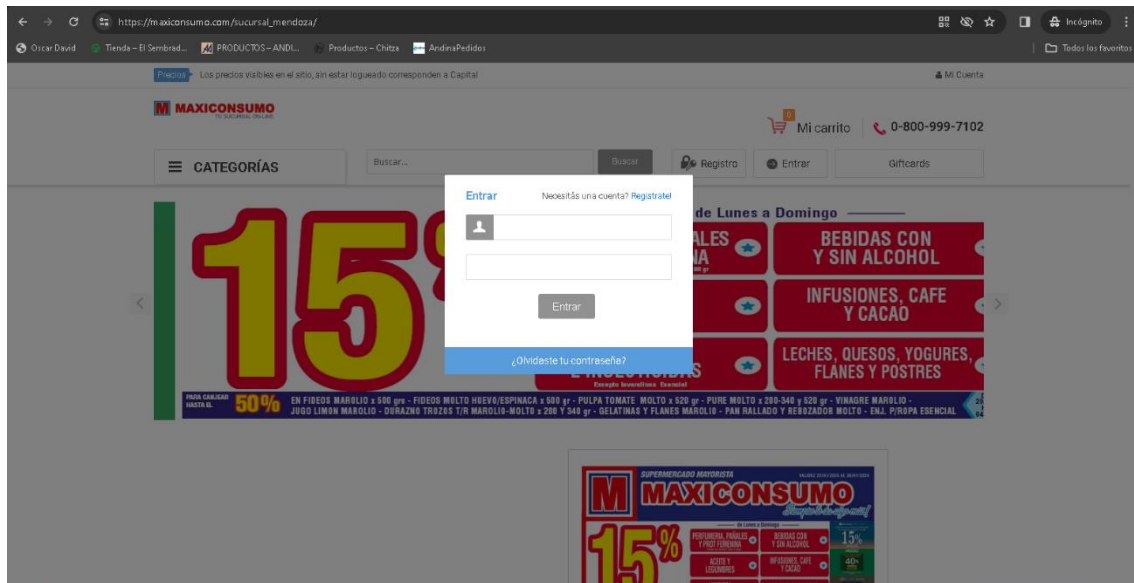
Ahora podemos cargar manualmente las credenciales de Maxiconsumo y La Serenísima en los campos correspondientes, o también podemos cargar estas credenciales desde la pestaña de Opciones, Cargar Credenciales (Si no hay credenciales cargadas de antemano, se pueden editar desde la misma pestaña de Opciones, Editar Credenciales). Una vez hecho esto, los cuatro campos del Sesionador deben estar llenos.

Es importante que al darle click al botón de “Abrir Sesiones” no se toque nada, porque se ejecutará un ambiente automatizado controlado por un bot y cualquier click puede entorpecer la sucesión de eventos que están programados para iniciar las sesiones.

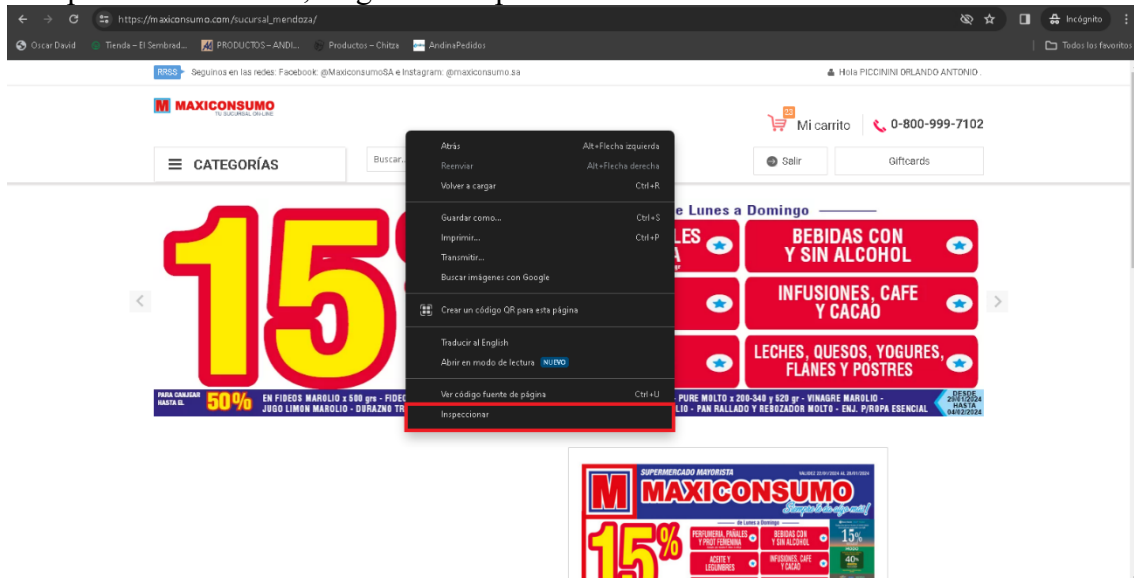
El ambiente automatizado realizará los inicios de sesión correspondientes con las credenciales proveídas y luego se cerrará automáticamente (Si no se cierra de forma automática, el navegador automatizado se puede cerrar una vez aparezcan los Session ID en sus campos correspondientes) al igual que la ventana del sesionador.

Obtención manual de ID de Sesión

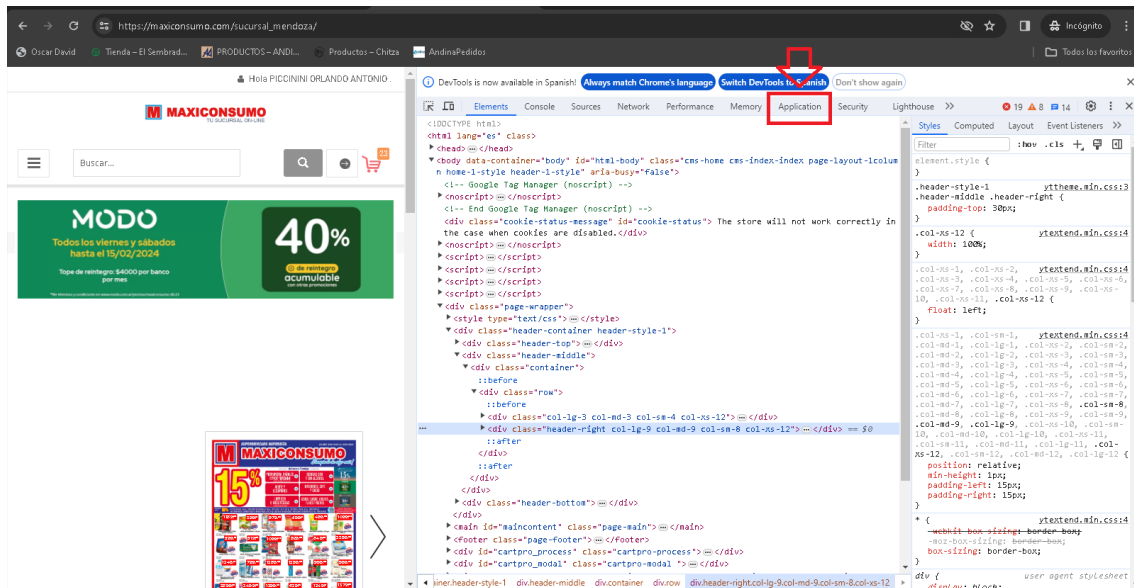
Para conseguir el ID de Sesión que el programa requiere, primero hay que ingresar a la página del proveedor. Una vez adentro, hay que iniciar sesión como uno haría normalmente utilizando las credenciales de la empresa.



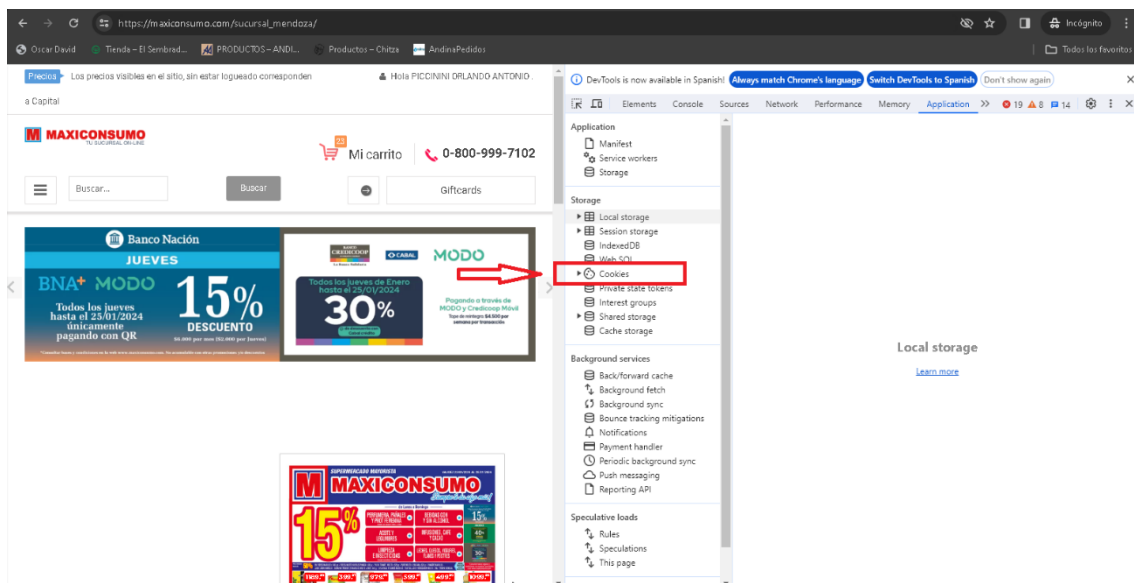
Una vez iniciada la sesión, hacemos click derecho en cualquier lugar de la página. Entre las opciones mostradas, elegimos “Inspeccionar”.



Al darle click a esta opción, se nos abrirá una ventana que contiene el código fuente de la página y otro tipo de información que no nos interesa. Lo que nos interesa son las opciones que se encuentran encima de dicho código fuente, en particular la pestaña que dice “Application” o “Aplicación”.



Dándole click a esta opción, se nos desvelará una serie de categorías de información. Damos click en donde dice “Cookies” en la sección de “Storage”.



Hay que darle click a la sección de “Cookies” y se desplegarán dos opciones. Seleccionamos la primera:

The screenshot shows a web browser with the URL https://maxiconsumo.com/sucursal_mendoza/. The page features a banner for 'HORARIO DE VERANO' (Summer Hours) and a '15%' discount offer. The Chrome DevTools Application tab is open, displaying a list of cookies. The 'PHPSESSID' cookie is highlighted, indicating it is the session ID being tracked.

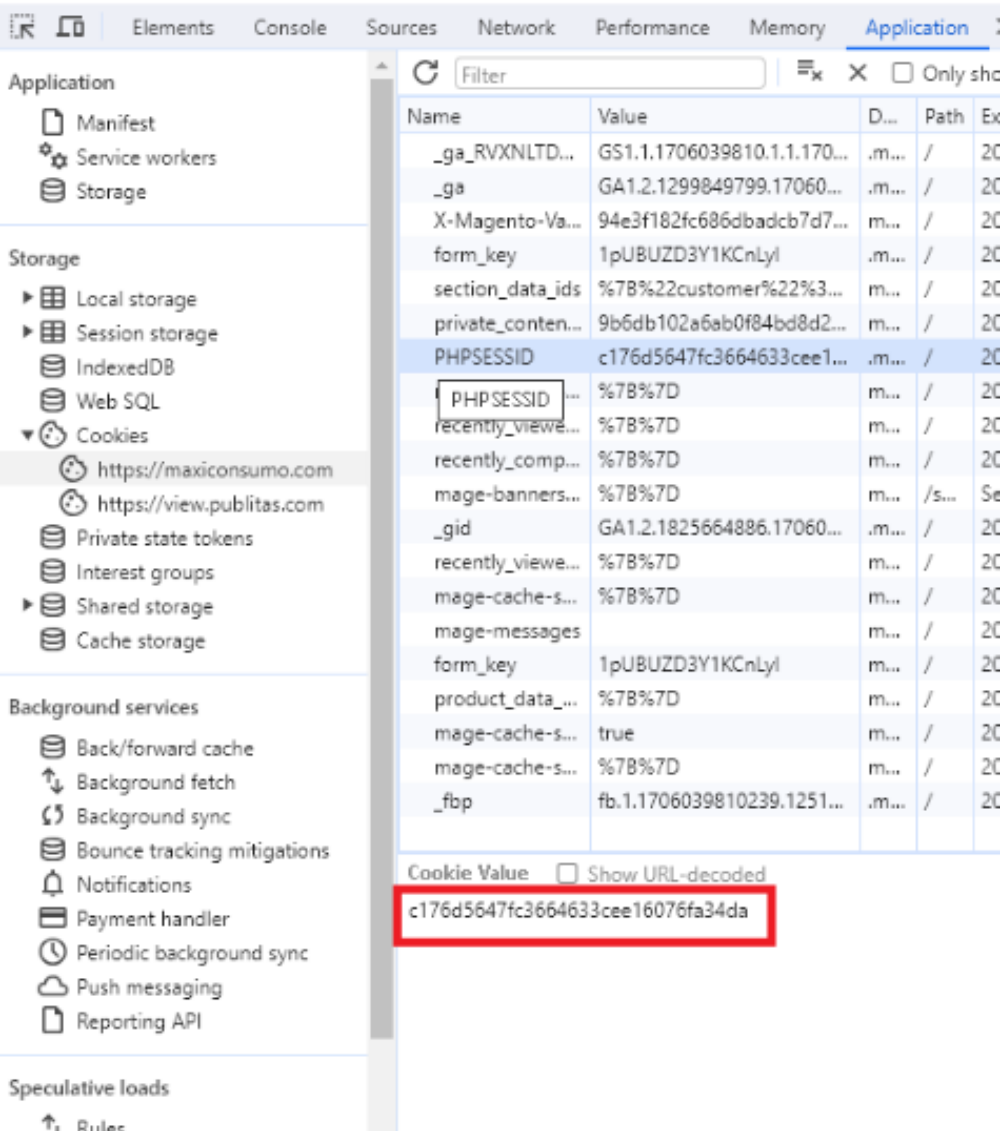
Name	Value	D...	Path	Ex...	Size	Ht...	Se...	Sa...	Pa...	P...
_ga_RVXNLT...	GS1.1.1706039810.1.1.170...	.m...	/	20...	52					M...
_ga	GA1.2.1299849799.17060...	.m...	/	20...	30					M...
X-Magento-Va...	94e3f182fc686dbadcb7d7...	.m...	/	20...	54	✓	✓	N...		M...
form_key	1pUBUZZD3Y1KCnLyl	.m...	/	20...	24			✓	N...	M...
section_data_ids	%7B%22customer%22%3...	.m...	/	20...	602			✓	Lax	M...
private_conten...	9b6dcb102a6ab0f84bd8d2...	.m...	/	20...	55			✓	N...	M...
PHPSESSID	c176d5647fc3664633cee1...	.m...	/	20...	41	✓	✓	N...		M...
recently_comp...	%7B%7D	.m...	/	20...	31			✓	Lax	M...
recently_view...	%7B%7D	.m...	/	20...	38			✓	Lax	M...
recently_comp...	%7B%7D	.m...	/	20...	40			✓	Lax	M...
mage-banners...	%7B%7D	.m...	/s...	Se...	32			✓	Lax	M...
_gid	GA1.2.1825664886.17060...	.m...	/	20...	31					M...
recently_view...	%7B%7D	.m...	/	20...	29			✓	Lax	M...
mage-cache-s...	%7B%7D	.m...	/	20...	24			✓	Lax	M...
mage-messages	%7B%7D	.m...	/	20...	13				St...	M...
form_key	1pUBUZZD3Y1KCnLyl	.m...	/	20...	24			✓	Lax	M...
product_data_...	%7B%7D	.m...	/	20...	26			✓	Lax	M...
mage-cache-s...	true	.m...	/	20...	21			✓	Lax	M...
mage-cache-s...	%7B%7D	.m...	/	20...	45			✓	Lax	M...
_fbp	fb.1.1706039810239.1251...	.m...	/	20...	33				Lax	M...

Aquí podremos ver las “cookies” de nuestro navegador. Una de estas “cookies” es el ID De Sesión que estamos buscando. Al darle click a la primera opción del paso anterior, nos mostrará todas las “cookies” de esta página a la derecha. Debemos dar click en la que nos interesa a nosotros, que es “PHPSESSID”.

The screenshot shows the Chrome DevTools Application tab with the 'Cookies' section selected. The 'PHPSESSID' cookie is highlighted, and the text 'Select a cookie to preview its value' is displayed below the list.

Name	Value	D...	Path	Ex...	Size	Ht...	Se...	Sa...	Pa...	P...
_ga_RVXNLT...	GS1.1.1706039810.1.1.170...	.m...	/	20...	52					M...
_ga	GA1.2.1299849799.17060...	.m...	/	20...	30					M...
X-Magento-Va...	94e3f182fc686dbadcb7d7...	.m...	/	20...	54	✓	✓	N...		M...
form_key	1pUBUZZD3Y1KCnLyl	.m...	/	20...	24			✓	N...	M...
section_data_ids	%7B%22customer%22%3...	.m...	/	20...	602			✓	Lax	M...
private_conten...	9b6dcb102a6ab0f84bd8d2...	.m...	/	20...	55			✓	N...	M...
PHPSESSID	c176d5647fc3664633cee1...	.m...	/	20...	41	✓	✓	N...		M...
recently_comp...	%7B%7D	.m...	/	20...	31			✓	Lax	M...
recently_view...	%7B%7D	.m...	/	20...	38			✓	Lax	M...
recently_comp...	%7B%7D	.m...	/	20...	40			✓	Lax	M...
mage-banners...	%7B%7D	.m...	/s...	Se...	32			✓	Lax	M...
_gid	GA1.2.1825664886.17060...	.m...	/	20...	31					M...
recently_view...	%7B%7D	.m...	/	20...	29			✓	Lax	M...
mage-cache-s...	%7B%7D	.m...	/	20...	24			✓	Lax	M...
mage-messages	%7B%7D	.m...	/	20...	13				St...	M...
form_key	1pUBUZZD3Y1KCnLyl	.m...	/	20...	24			✓	Lax	M...
product_data_...	%7B%7D	.m...	/	20...	26			✓	Lax	M...
mage-cache-s...	true	.m...	/	20...	21			✓	Lax	M...
mage-cache-s...	%7B%7D	.m...	/	20...	45			✓	Lax	M...
_fbp	fb.1.1706039810239.1251...	.m...	/	20...	33				Lax	M...

Si le damos click, en la parte de abajo quedará especificado claramente el ID de Sesión. Es cuestión de copiarlo para más tarde pegarlo en el programa.



Nótese que este mismo método sirve para cualquiera de los proveedores, solo que La Serenísima maneja su Session ID con el nombre de “session_id” en vez de “PHPSESSID”.

Búsqueda de precios

Una vez realizados todos los preparativos (ya se tiene el EXCEL con los datos de los artículos a buscar filtrados por los proveedores que se desean buscar y cada proveedor tiene su Session ID cargado) se procede a ejecutar el programa para que realice la búsqueda de precios.

El programa listo para funcionar quedará de la siguiente forma:

II Porco Scraper 4.5

Opciones

II Porco Scraper

Obtener sesiones

Ingrese el ID de Sesión de Maxiconsumo: 25c3ea7a5e9bed01b4c...

Ingrese el ID de Andina: e88d134e4f63da739e0e

Ingrese el ID de Sesión de Oscar David: 0bf657409232a8be97fe

Ingrese el ID de Sesión de La Serenísima: d6f3c0add90e41a84921

Ingrese el nombre del libro de Excel: maxi-od-3

Salir

☐ Buscar "Sin Precio" en Oscar David

Iniciar Búsqueda de precios

* Versión funcional con distribuidores: Maxiconsumo, Andina, Oscar David y La Serenísima

O si se optó por elegir un archivo desde una ruta alterna, de la siguiente forma:

II Porco Scraper 4.5

Opciones

II Porco Scraper

Obtener sesiones

Ingrese el ID de Sesión de Maxiconsumo: 25c3ea7a5e9bed01b4c...

Ingrese el ID de Andina: e88d134e4f63da739e0e

Ingrese el ID de Sesión de Oscar David: 0bf657409232a8be97fe

Ingrese el ID de Sesión de La Serenísima: d6f3c0add90e41a84921

Ingrese el nombre del libro de Excel:

Salir

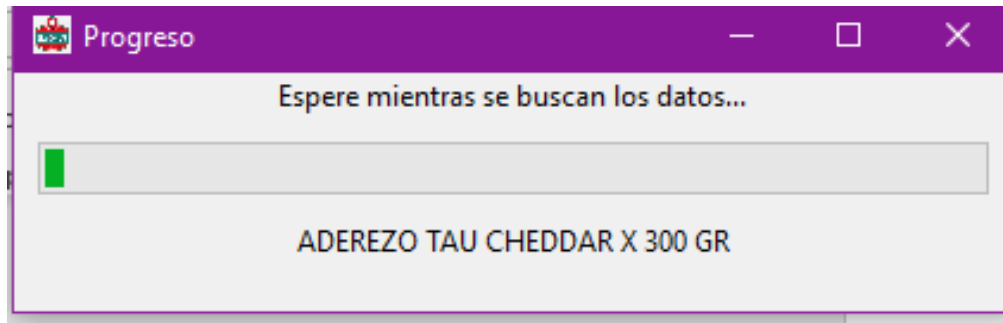
☐ Buscar "Sin Precio" en Oscar David

Iniciar Búsqueda de precios

* Versión funcional con distribuidores: Maxiconsumo, Andina, Oscar David y La Serenísima

Nótese que existe una opción de buscar los precios no encontrados en Maxiconsumo, en Oscar David. Si esta opción está prendida, todos los precios que no sean encontrados en la página de Maxiconsumo para los artículos serán automáticamente buscados en la página de Oscar David, y si se encuentra en este último proveedor un precio para el producto, será añadido con una advertencia de que el precio conseguido pertenece a la página de Oscar David.

Una vez todo esté preparado para la búsqueda, se corre el Operador del programa dando click en “Iniciar Búsqueda de Precios”. Si todo sale bien, se mostrará una nueva ventana que lleva el progreso de la búsqueda de los artículos, así como el artículo que se está buscando actualmente.



Una vez la barra termine de llenarse, se abrirá un archivo de EXCEL automáticamente con los datos actualizados