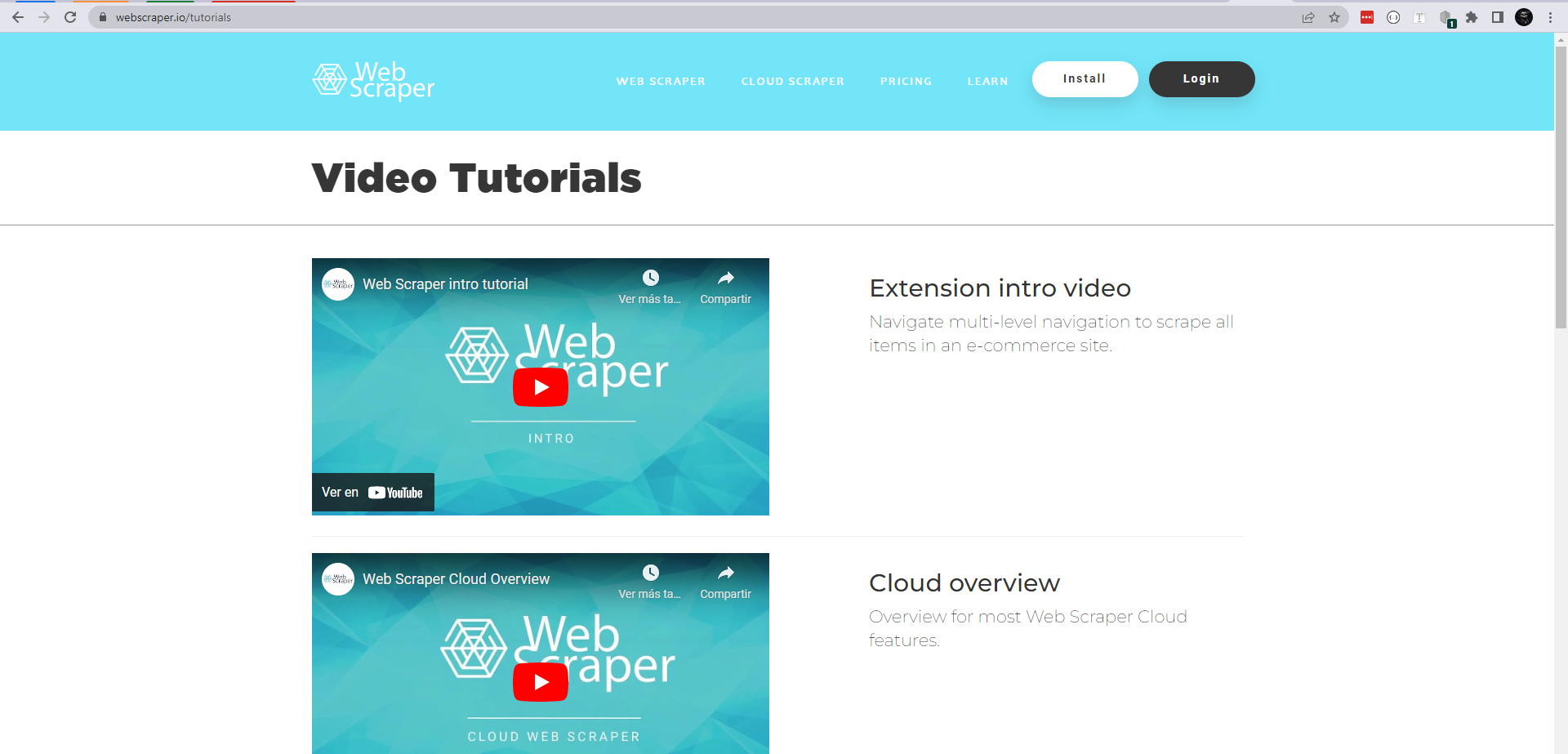
**MONGO DB: Web Scraper with Pagination.**

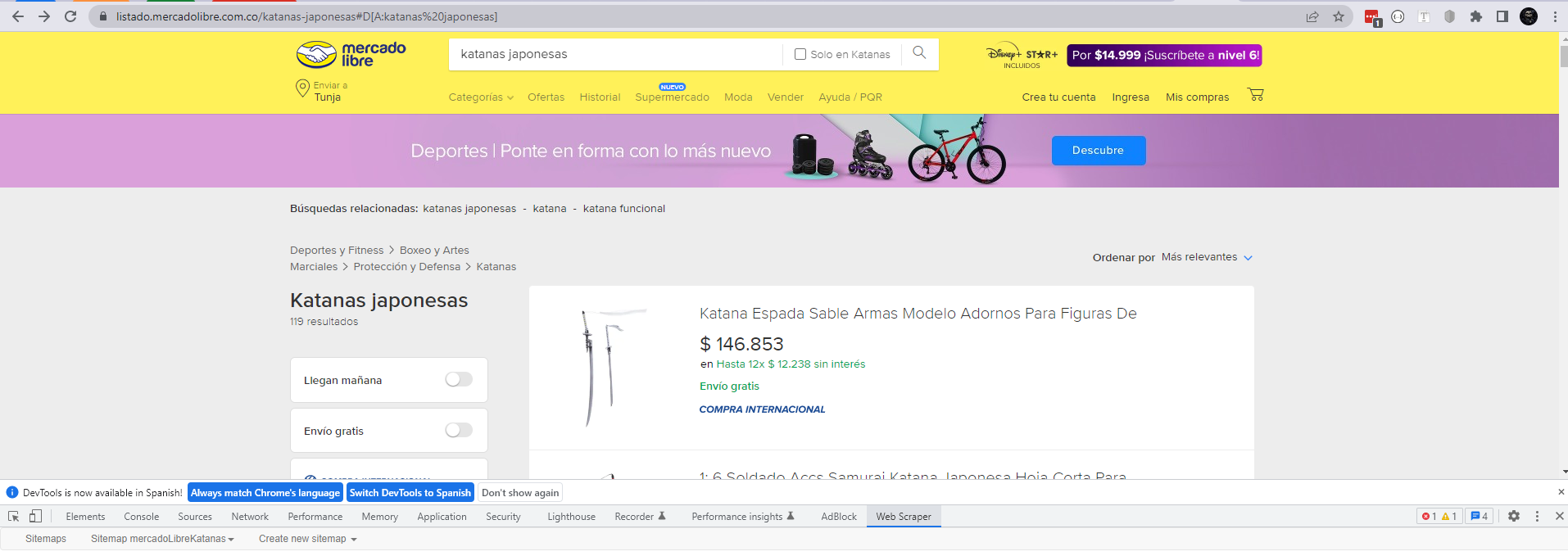
Ing. Luis Felipe Narvaez Gomez. E-mail: [Luis.narvaez@usantoto.edu.co](mailto:Luis.narvaez@usantoto.edu.co). Cod: 2312660. Facultad de Ingeniería de Sistemas

**GUÍA BÁSICA DE WEB SCRAPER PARA PAGINACIÓN.**

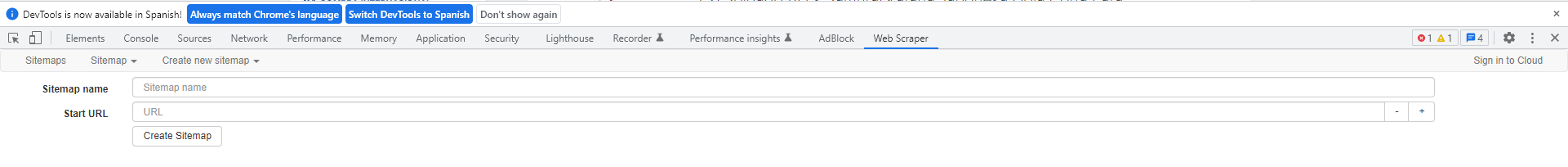
Para poder utilizar la herramienta de Web Scraper en tu navegador, es necesario instalarla, para ello se recomienda entrar al siguiente enlace <https://www.webscraper.io/tutorials> o ir al apartado de la tienda de extensiones de tu navegador en cuestión.



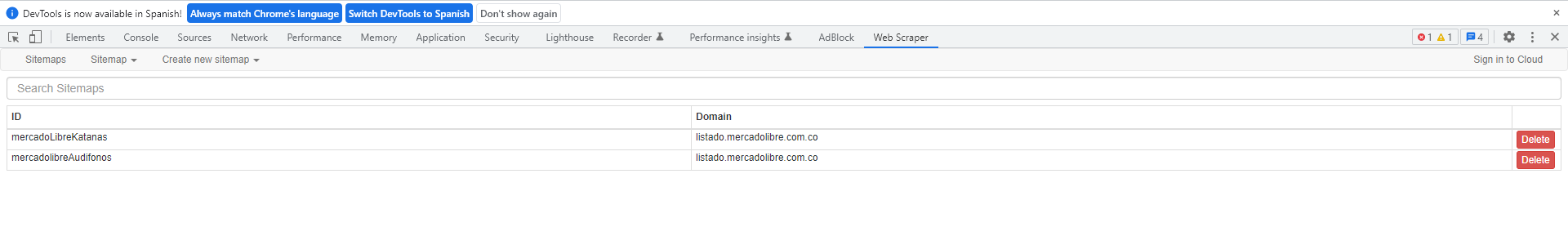
Una vez instalada la herramienta del navegador, nos dirigimos a la página web de nuestro interés, en nuestro caso mercado libre, en ella damos una búsqueda de algún producto al cual queramos hacerle Scraper. Dentro de la Búsqueda damos el comando [CTRL]+[SHIFT]+[I] para acceder al código de la pagina en el situamos la ventana emergente de forma horizontal para hacer más cómodo su uso y en la barra de menús, ubicamos WebScraper.



Ahora vamos a crear un nuevo SITEMAP, para ello nos referimos al menú desplegable “Create new sitemap”, ahí seleccionamos “CreateSitemap”. Allí encontraremos un espacio donde podremos denominar el SiteMap y otro espacio justo debajo donde podemos insertar la URL de nuestra búsqueda, la misma la hallaremos en la barra superior de búsqueda de nuestro navegador.

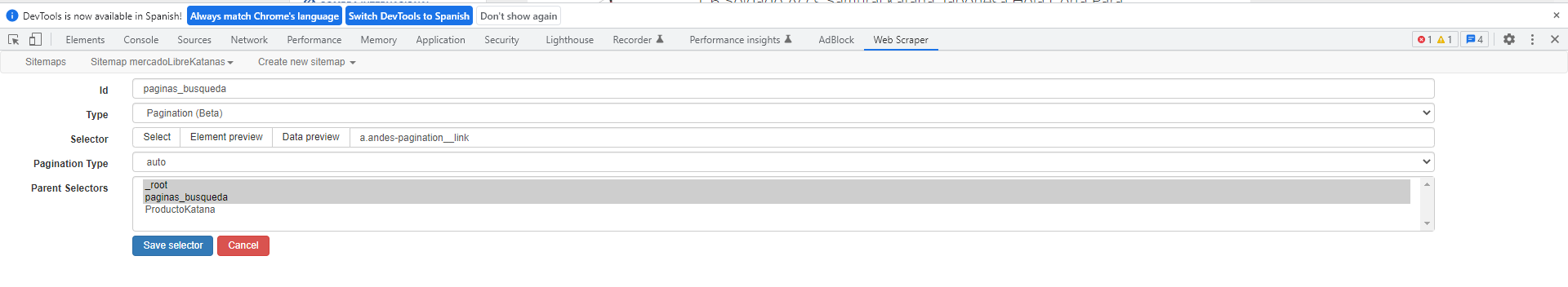


Una vez creado nuestro SiteMap nuevo podremos encontrarlo en la lista de sitemaps que tengamos creados. Podemos acceder a la lista de Sitemap seleccionando en la barra de herramientas “Sitemaps”.



Seleccionamos el SiteMap que acabamos de crear, en este caso para el producto de Katanas japonesas y dentro de el crearemos un nuevo “selector”, para eso daremos clic izquierdo sobre el botón azul “add new selector”. Este primer selector será el encargado de dar a paginación de nuestra búsqueda.

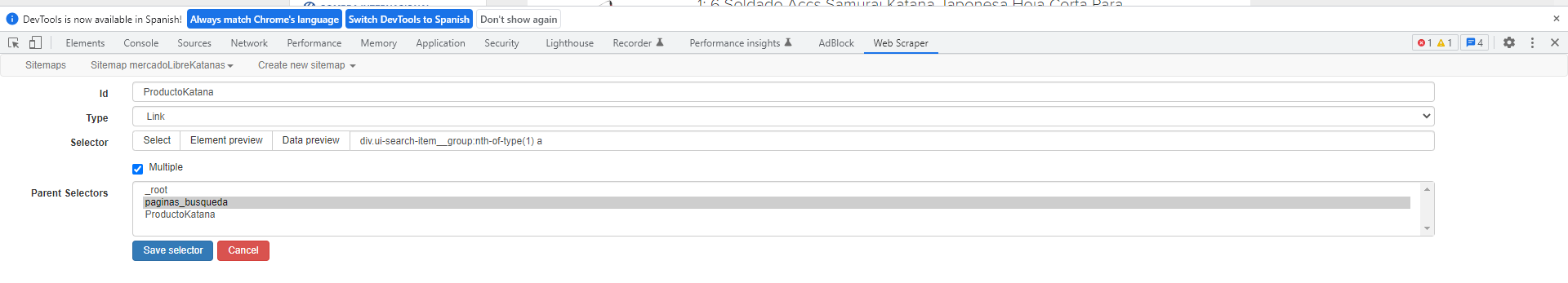
Debemos tener en cuenta que el tipo de este elemento es el que se muestra la imagen, la parte de selección corresponderá a la consecución de números de las páginas de nuestra búsqueda en el E-Commerce o bien la selección del Botón “Siguiente” o “Next”.



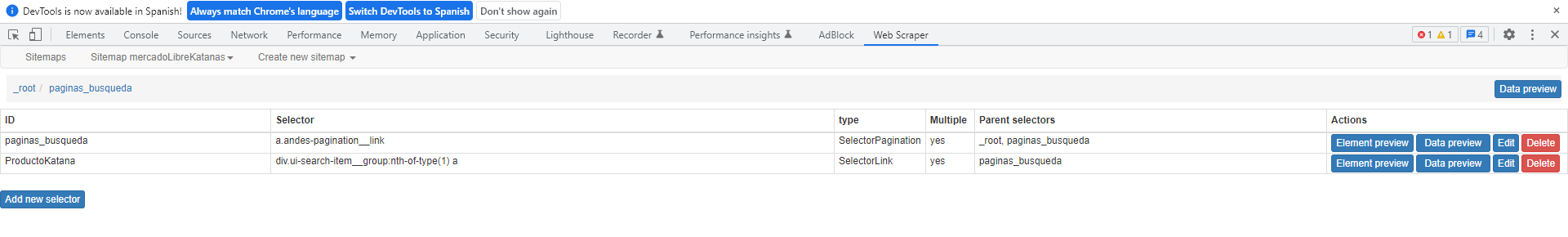
Al crear este primer selector, entraremos en él, analizando la barra de dirección superior que nos asegurara la ubicación de nuestros elementos de selección. Dentro de esta primera selección daremos la tarjeta o “Card” de nuestro producto individual.

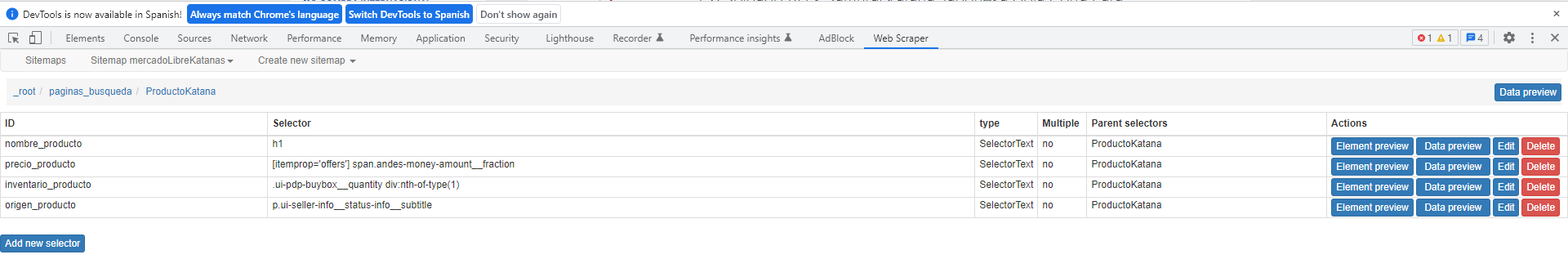
Este será de tipo Link y en el seleccionaremos con la herramienta “select” algún atributo de la tarjeta que individualiza el producto y permite ingresar a verlo en mas detalle, en este caso se selecciona el nombre del producto. Ojo, aquí seleccionamos el nombre del producto no porque queramos el nombre en este punto, si no porque queremos el link de ingreso para ver y producto individualmente.

Algo que debemos tener en cuenta es que debemos seleccionar la casilla de “Multiple” esto debido a que no solo vamos a entrar en un Link o en un solo producto, si no en cada uno de los que se presenten en la página.

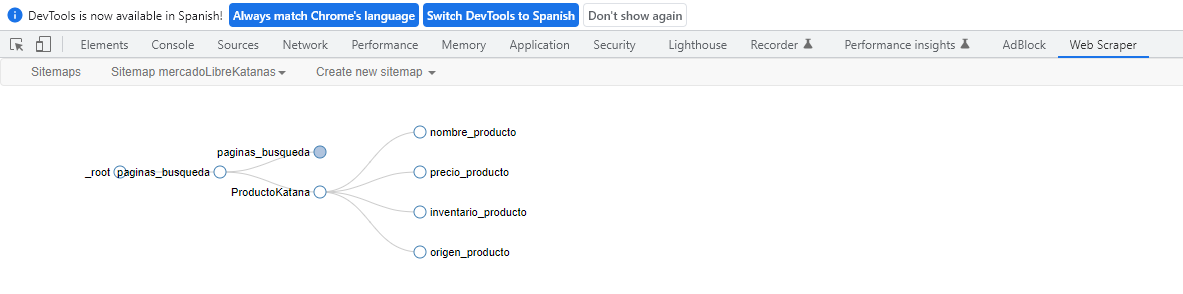


Dentro de este nuevo elemento crearemos tantos elementos queramos saber de un producto y sea común entre todos los productos, en este ejercicio estos elementos serán de tipo TEXT, pero no necesariamente tienen que ser siempre de este tipo. En este caso cuando ingresamos a un producto individual, solo existirá un atributo único por cada búsqueda, por lo que no es necesario seleccionar “Multiple”.

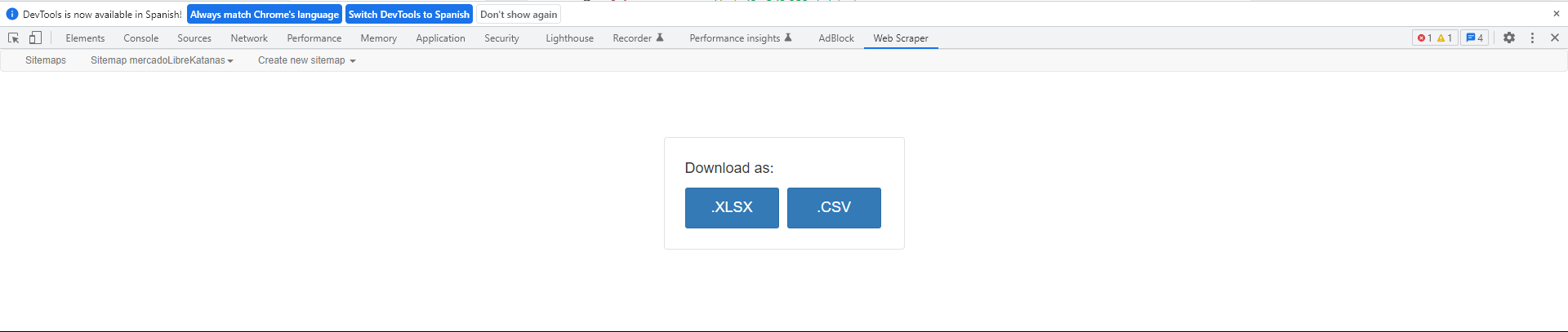




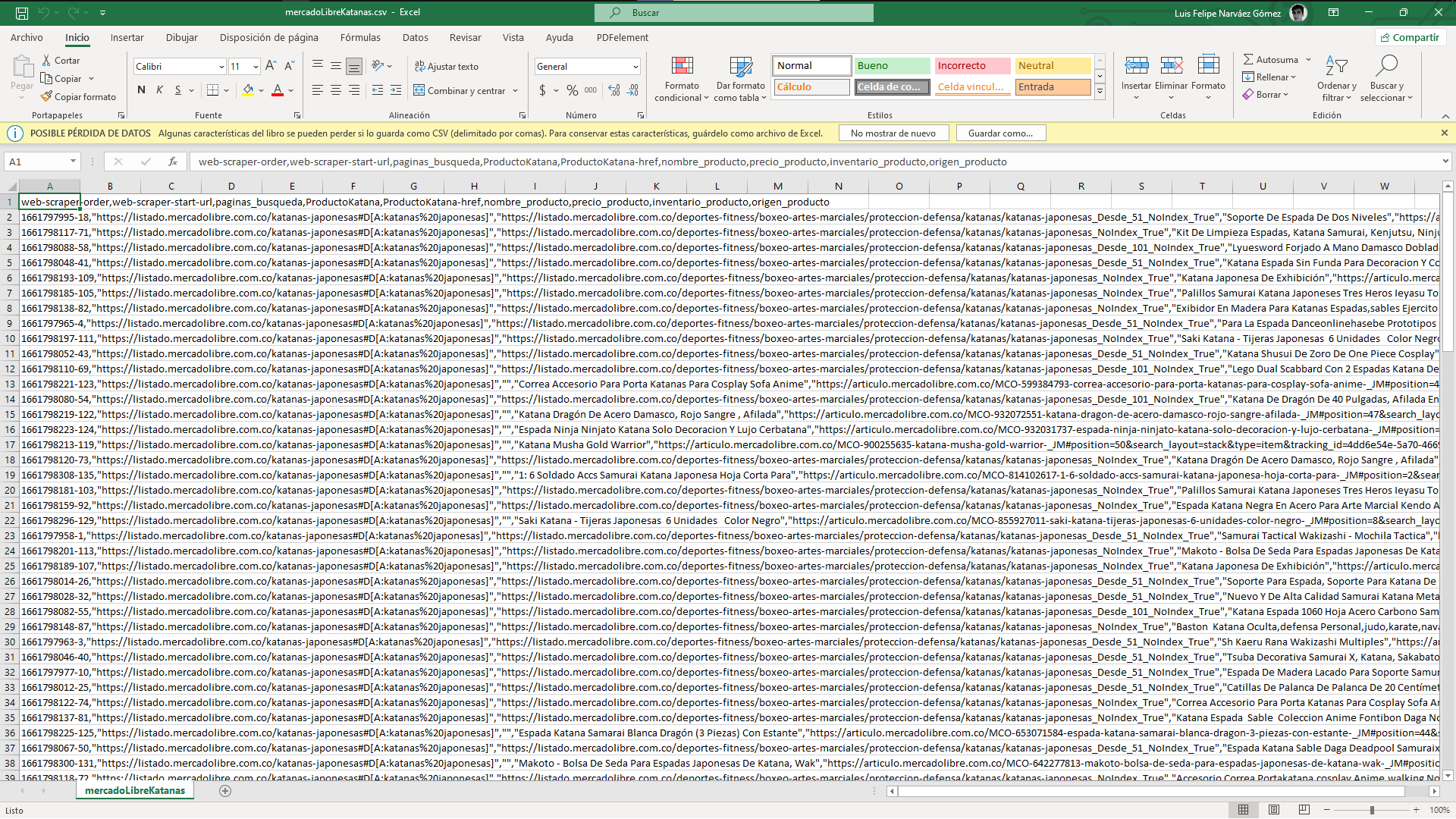
Una vez tenemos los atributos verificamos nuestro Grafo de nuestro Sitemap que se vera similar al siguiente. Para poderlo ver, podeos seleccionar en la barra de herramientas “SiteMap nombre\_sitemap”, ahí seleccionamos “selector graph”.

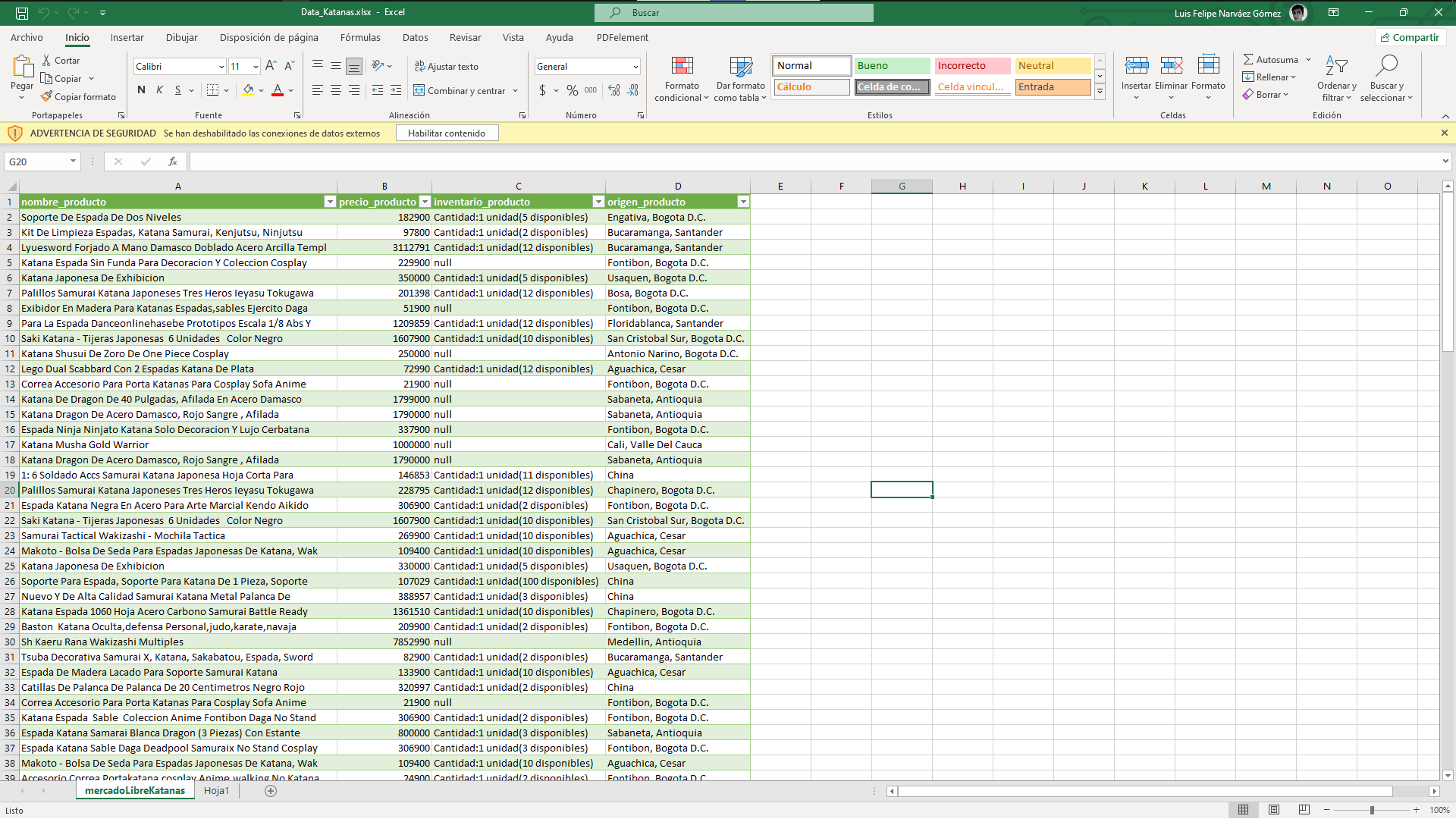


Como vemos en la parte de paginación, generamos un bucle infinito que corresponderá a la búsqueda que realizará el Scraper para adquirir la información. Teniendo todo listo podemos dirigirnos a la barra de herramientas “Sitemap nombre\_sitemap” y seleccionamos “Scrape”. Una vez seleccionado se ejecutará una ventana emergente que empezara a indagar en orden del árbol del scraper que creamos, todos los atributos de la información serán capturados. Una vez termine podremos dar clic izquierdo en el botón “Refresh” y dirigirnos nuevamente a la barra de herramientas y selecciona la opción “Export Data”.



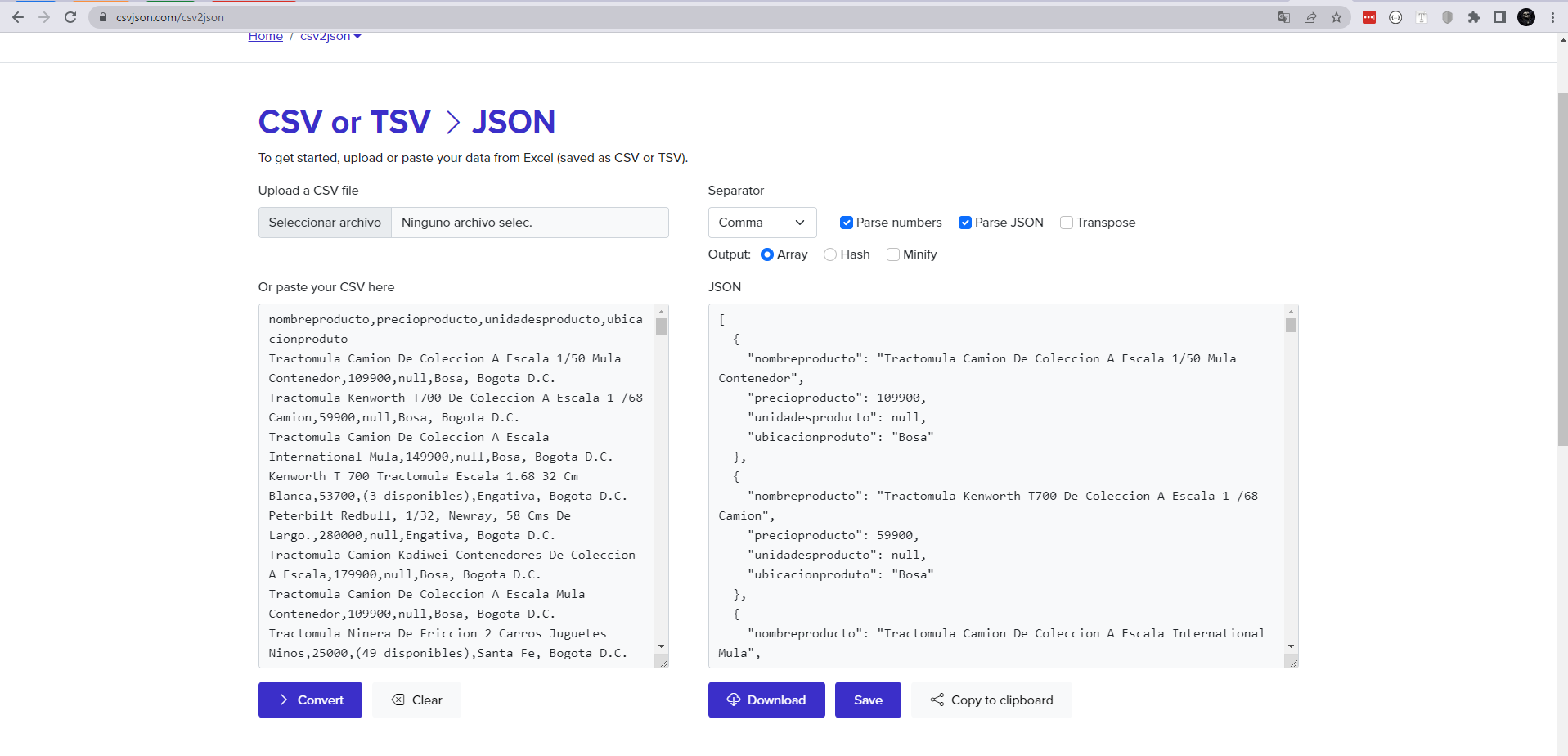
Ahí podemos obtener nuestros datos en extensión. XLSX y en extencion .CSV, en mi caso quiero los datos en esta última. Ahora queda limpiar los datos, esto lo haremos en Excel, quitando columnas de información que no queramos, eliminando caracteres extraños, cambiando las vocales con tilde a las mismas sin tilde, quitando la Ñ por la N o otra letra extraña. Luego abrimos el CSV con un editor de texto básico como el Bloc de Notas y reemplazamos el separador de los datos “;” por la “,”.

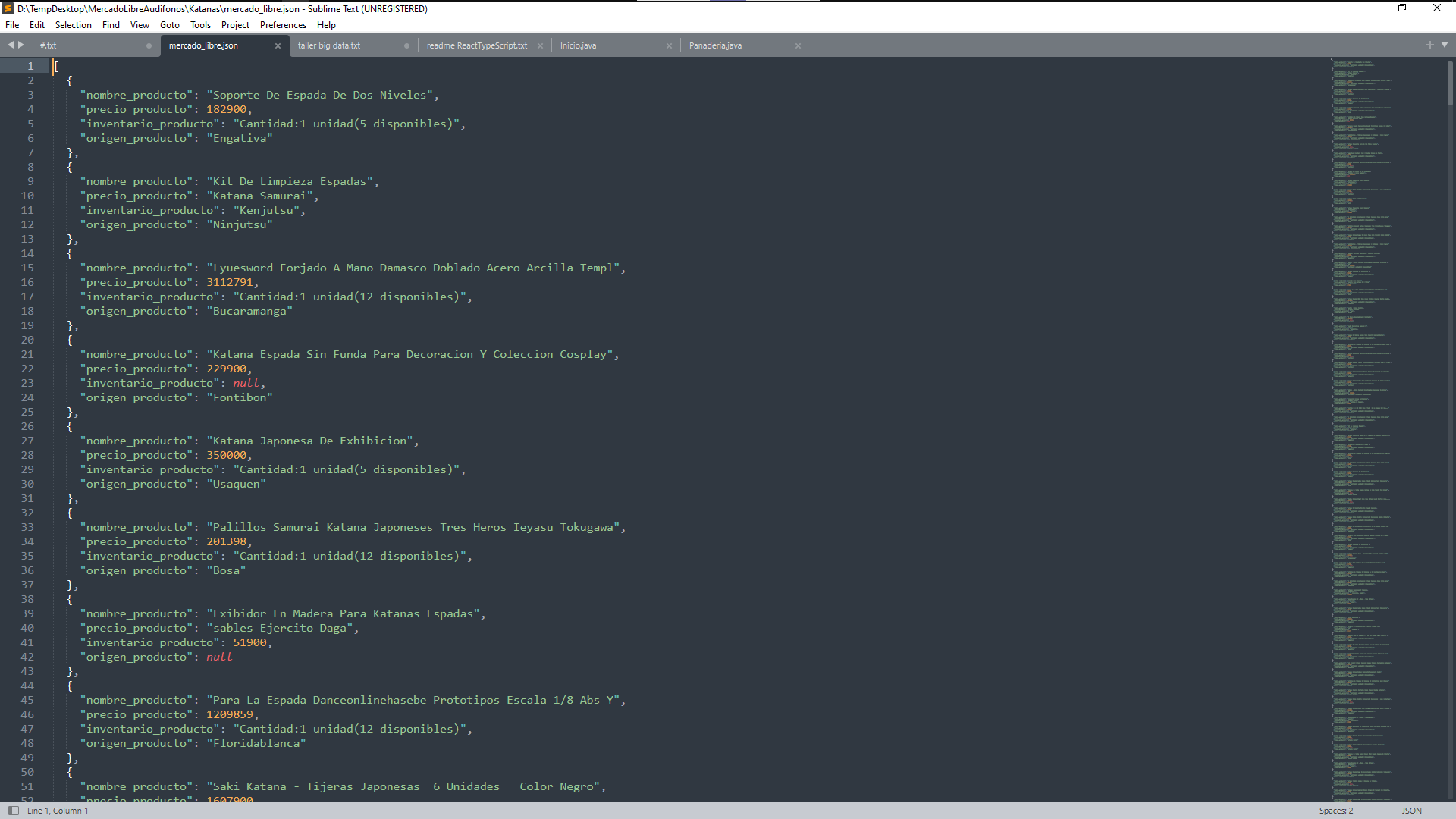






De ahí nos dirigimos al siguiente link <https://csvjson.com/csv2json> donde podemos transformar el CSV a JSON, archivo de tipop diccionario de información en NombreValor que podemos utilizar para crear la colección en la base de datos.





Ya dentro de Mongo DB solo queda insertar esta nueva información para tener nuestra colección de a base de datos.



