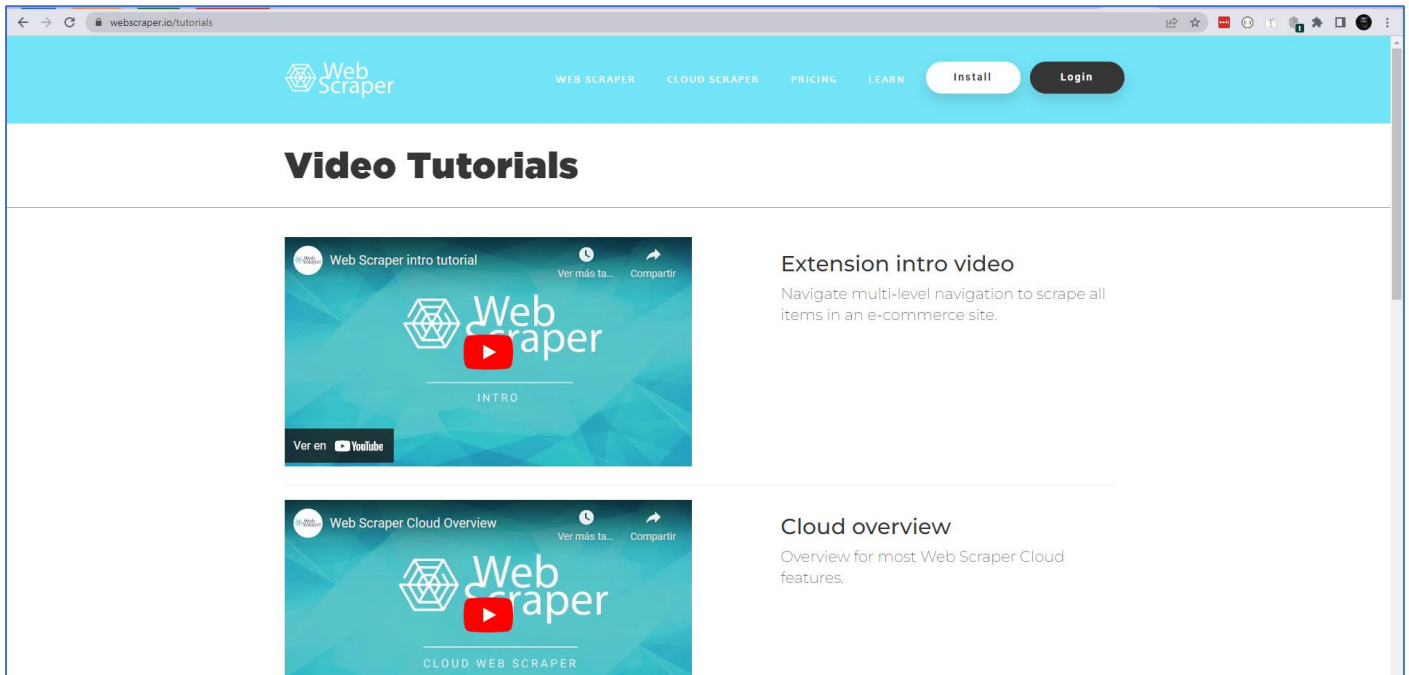


MONGO DB: Web Scraper with Pagination.

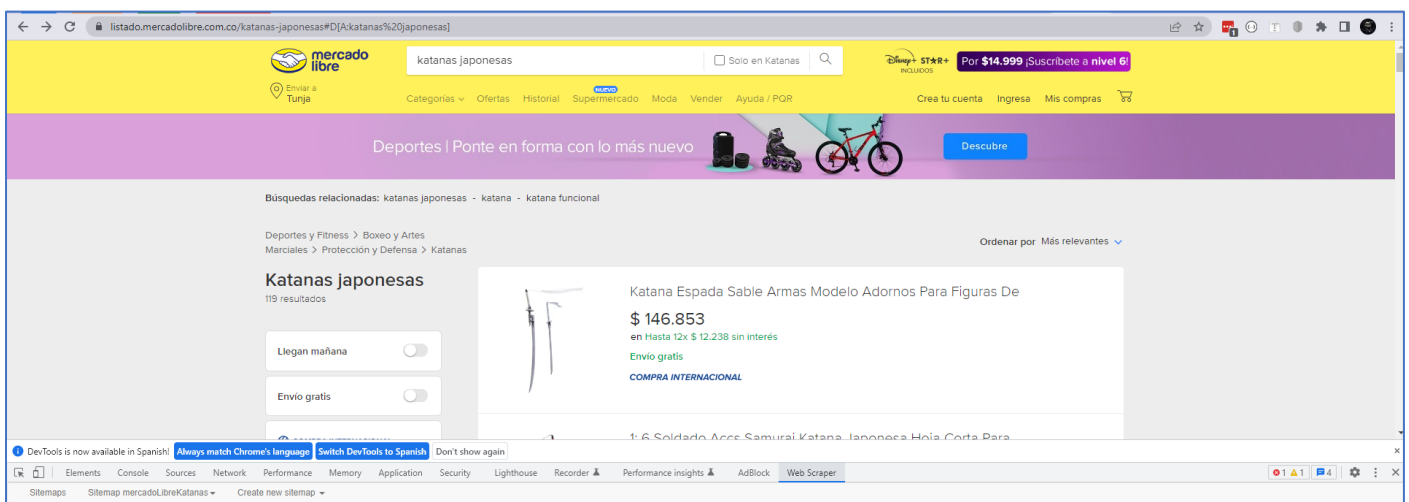
Ing. Luis Felipe Narvaez Gomez. E-mail: Luis.narvaez@usantoto.edu.co. Cod: 2312660. Facultad de Ingeniería de Sistemas

GUÍA BÁSICA DE WEB SCRAPER PARA PAGINACIÓN.

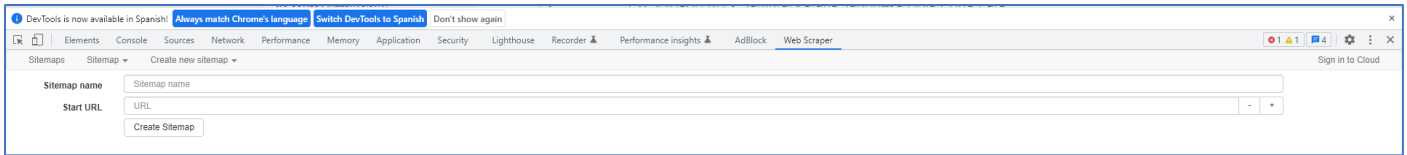
Para poder utilizar la herramienta de Web Scraper en tu navegador, es necesario instalarla, para ello se recomienda entrar al siguiente enlace <https://www.webscraper.io/tutorials> o ir al apartado de la tienda de extensiones de tu navegador en cuestión.



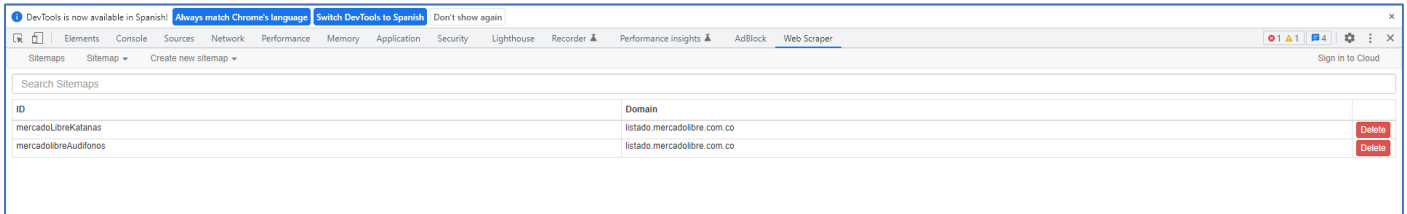
Una vez instalada la herramienta del navegador, nos dirigimos a la página web de nuestro interés, en nuestro caso mercado libre, en ella damos una búsqueda de algún producto al cual queramos hacerle Scraper. Dentro de la Búsqueda damos el comando [CTRL]+[SHIFT]+[I] para acceder al código de la pagina 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.

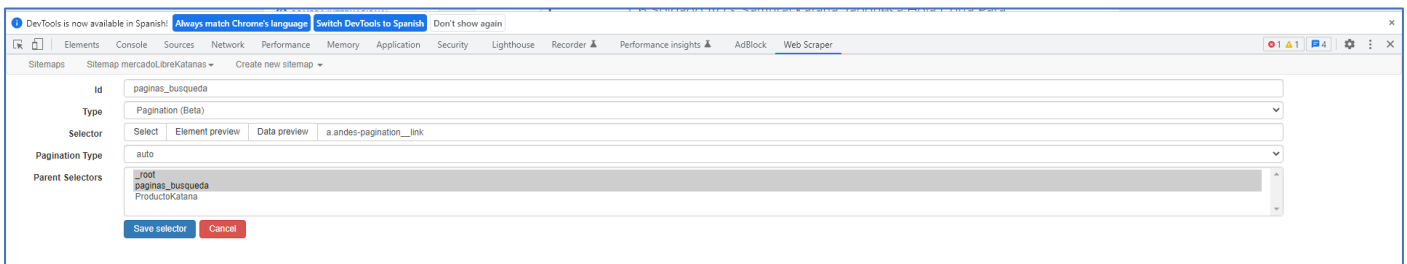


Una vez creado nuestro SiteMap nuevo podremos encontrarlo en la lista de sitemaps que tengamos creados. Podemos acceder a la lista de Sitemap seleccionando en la barra de herramientas “Sitemaps”.



Seleccionamos el SiteMap que acabamos de crear, en este caso para el producto de Katanas japonesas y dentro de él 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.

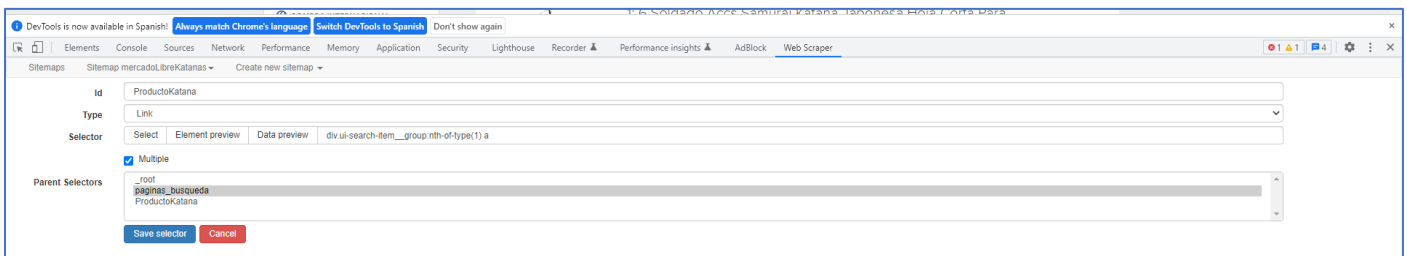
Debemos tener en cuenta que el tipo de este elemento es el que se muestra la imagen, la parte de selección corresponderá a la consecución de números de las páginas de nuestra búsqueda en el E-Commerce o bien la selección del Botón “Siguiente” o “Next”.



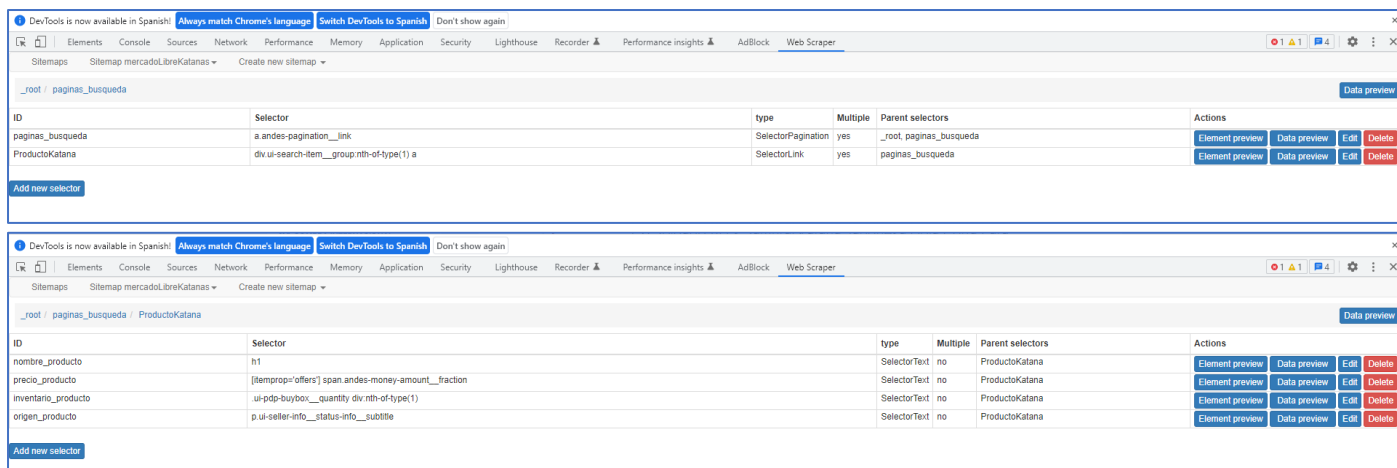
Al crear este primer selector, entraremos en él, analizando la barra de dirección superior que nos asegurara la ubicación de nuestros elementos de selección. Dentro de esta primera selección daremos la tarjeta o “Card” de nuestro producto individual.

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.

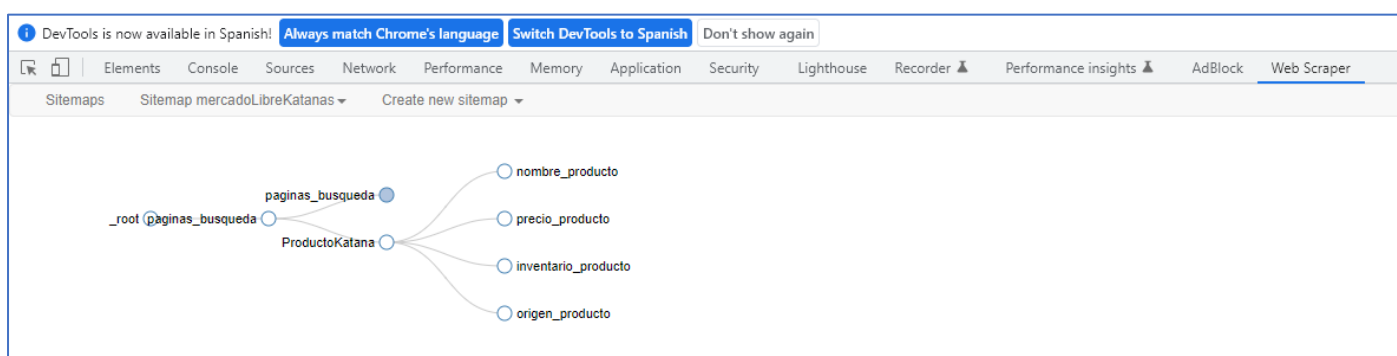
Algo que debemos tener en cuenta es que debemos seleccionar la casilla de “Multiple” esto debido a que no solo vamos a entrar en un Link o en un solo producto, si no en cada uno de los que se presenten en la página.



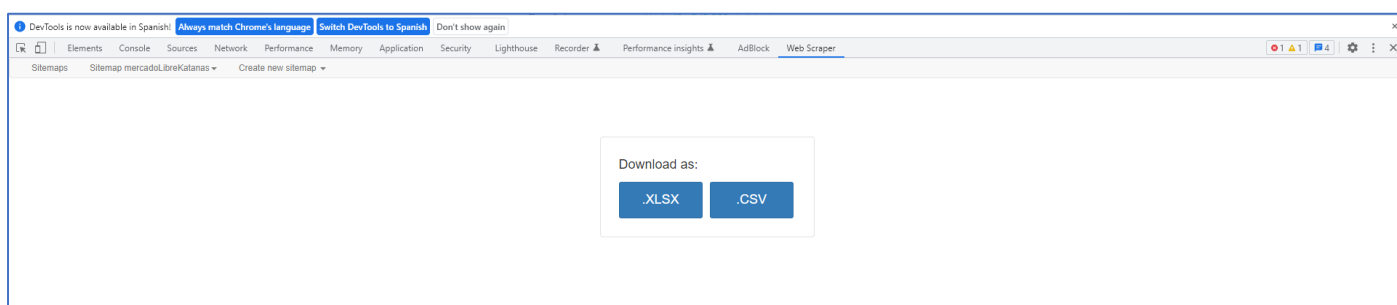
Dentro de este nuevo elemento crearemos tantos elementos queramos saber de un producto y sea común entre todos los productos, en este ejercicio estos elementos serán de tipo TEXT, pero no necesariamente tienen que ser siempre de este tipo. En este caso cuando ingresamos a un producto individual, solo 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 “,”.


```
mercadolibre_katanas_limpio.csv: Bloc de notas
Archivo Edición Formato Ver Ayuda
nombre_producto,precio_producto,inventory_producto,origen_producto
Soporte De Espada De Dos Niveles,182900,Cantidad:1 unidad(5 disponibles),Engativa, Bogota D.C.
Kit De Limpieza Espadas, Katana Samurai, Kenjutsu, Ninjutsu,97800,Cantidad:1 unidad(2 disponibles),Bucaramanga, Santander
Lyesword Forjado A Mano Damasco Doblado Acero Arcilla Temp,3112791,Cantidad:1 unidad(12 disponibles),Bucaramanga, Santander
Katana Espada Sin Funda Para Decoracion Y Coleccion Cosplay,229900,null,Fontibon, Bogota D.C.
Katana Japonesa De Exhibicion,350000,Cantidad:1 unidad(5 disponibles),Usaquen, Bogota D.C.
Pallillos Samurai Katana Japoneses Tres Heros Ieyasu Tokugawa,201398,Cantidad:1 unidad(12 disponibles),Bosa, Bogota D.C.
Exhibidor En Madera Para Katanas Espadas,sables Ejercito Daga,51900,null,Fontibon, Bogota D.C.
Para La Espada Danceonlinehasebe Prototipos Escala 1/8 Abs Y,1209859,Cantidad:1 unidad(12 disponibles),Floridablanca, Santander
Saki Katana - Tijeras Japonesas 6 Unidades Color Negro,1607900,Cantidad:1 unidad(10 disponibles),San Cristobal Sur, Bogota D.C.
Katana Shusui De Zoro De One Piece Cosplay,250000,null,Antonio Narino, Bogota D.C.
Lego Dual Scabbard Con 2 Espadas Katana De Plata,72990,Cantidad:1 unidad(12 disponibles),AguaChica, Cesar
Correa Accesorio Para Porta Katanas Para Cosplay Sofa Anime,21900,null,Fontibon, Bogota D.C.
Katana De Dragon De 40 Pulgadas, Afilada En Acero Damasco,1799000,null,Sabaneta, Antioquia
Katana Dragon De Acero Damasco, Rojo Sangre , Afilada,1790000,null,Sabaneta, Antioquia
Espada Ninja Ninjato Katana Solo Decoracion Y Lujo Cerbatana,337900,null,Fontibon, Bogota D.C.
Katana Musha Gold Warrior,1000000,null,Cali, Valle Del Cauca
Katana Dragon De Acero Damasco, Rojo Sangre , Afilada,1790000,null,Sabaneta, Antioquia
1: 6 Soldado Accs Samurai Katana Japonesa Hoja Corta Para,146853,Cantidad:1 unidad(11 disponibles),China
Pallillos Samurai Katana Japoneses Tres Heros Ieyasu Tokugawa,228795,Cantidad:1 unidad(12 disponibles),Chapinero, Bogota D.C.
Espada Katana Negra En Acero Para Arte Marcial Kendo Aikido,306900,Cantidad:1 unidad(2 disponibles),Fontibon, Bogota D.C.
Saki Katana - Tijeras Japonesas 6 Unidades Color Negro,1607900,Cantidad:1 unidad(10 disponibles),San Cristobal Sur, Bogota D.C.
Samurai Tactical Wakizashi - Mochila Tactica,269900,Cantidad:1 unidad(10 disponibles),AguaChica, Cesar
Makoto - Bolsa De Seda Para Espadas Japonesas De Katana, Wak,109400,Cantidad:1 unidad(10 disponibles),AguaChica, Cesar
Katana Japonesa De Exhibicion,330000,Cantidad:1 unidad(5 disponibles),Usaquen, Bogota D.C.
Soporte Para Espada, Soporte Para Katana De 1 Pieza, Soporte,107029,Cantidad:1 unidad(100 disponibles),China
Nuevo Y De Alta Calidad Samurai Katana Metal Palanca De,308957,Cantidad:1 unidad(3 disponibles),China
Katana Espada 1060 Hoja Acero Carbono Samurai Battle Ready,1361510,Cantidad:1 unidad(10 disponibles),Chapinero, Bogota D.C.
Baston Katana Oculta,defensa Personal,judo,karate,navaja,209900,Cantidad:1 unidad(2 disponibles),Fontibon, Bogota D.C.
Sh Kaeru Rana Wakizashi Multiples,7852990,null,Medellin, Antioquia
Tsuba Decorativa Samurai X, Katana, Sakabatou, Espada, Sword,82900,Cantidad:1 unidad(2 disponibles),Bucaramanga, Santander
Espada De Madera Lacado Para Soporte Samurai Katana,133900,Cantidad:1 unidad(10 disponibles),AguaChica, Cesar
Catillas De Palanca De Palanca De 20 Centimetros Negro Rojo,320997,Cantidad:1 unidad(2 disponibles),China
Correa Accesorio Para Porta Katanas Para Cosplay Sofa Anime,21900,null,Fontibon, Bogota D.C.
Katana Espada Sable Coleccion Anime Fontibon Daga No Stand,306900,Cantidad:1 unidad(2 disponibles),Fontibon, Bogota D.C.
Espada Katana Samurai Blanca Dragon (3 Piezas) Con Estante,800000,Cantidad:1 unidad(3 disponibles),Sabaneta, Antioquia
Espada Katana Sable Daga Deadpool Samuraix No Stand Cosplay,306900,Cantidad:1 unidad(3 disponibles),Fontibon, Bogota D.C.
```

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.

← → ↻ csvjson.com/csv2json

Home / csv2json

CSV or TSV > JSON

To get started, upload or paste your data from Excel (saved as CSV or TSV).

Upload a CSV file

Seleccionar archivo Ninguno archivo selec.

Or paste your CSV here

```
nombreproducto,precioproducto,unidadesproducto,ubicacionproducto
Tractomula Camion De Coleccion A Escala 1/50 Mula
Contenedor,109900,null,Bosa, Bogota D.C.
Tractomula Kenworth T700 De Coleccion A Escala 1 /68
Camion,59900,null,Bosa, Bogota D.C.
Tractomula Camion De Coleccion A Escala
International Mula,149900,null,Bosa, Bogota D.C.
Kenworth T 700 Tractomula Escala 1.68 32 Cm
Blanca,53700,(3 disponibles),Engativa, Bogota D.C.
Peterbilt Redbull, 1/32, Newray, 58 Cms De
Largo.,280000,null,Engativa, Bogota D.C.
Tractomula Camion Kadiwei Contenedores De Coleccion
A Escala,179900,null,Bosa, Bogota D.C.
Tractomula Camion De Coleccion A Escala Mula
Contenedor,109900,null,Bosa, Bogota D.C.
Tractomula Niner De Friction 2 Carros Juguetes
Ninos,25000,(49 disponibles),Santa Fe, Bogota D.C.
```

Separator

Comma

☒ Parse numbers ☒ Parse JSON ☐ Transpose

Output: ☒ Array ☐ Hash ☐ Minify

JSON

```
[
{
  "nombreproducto": "Tractomula Camion De Coleccion A Escala 1/50 Mula
Contenedor",
  "precioproducto": 109900,
  "unidadesproducto": null,
  "ubicacionproducto": "Bosa"
},
{
  "nombreproducto": "Tractomula Kenworth T700 De Coleccion A Escala 1 /68
Camion",
  "precioproducto": 59900,
  "unidadesproducto": null,
  "ubicacionproducto": "Bosa"
},
{
  "nombreproducto": "Tractomula Camion De Coleccion A Escala International
Mula",
```

> Convert Clear

Download Save Copy to clipboard

```
D:\TempDesktop\MercadoLibreAudifonos\Katanas\mercado_libre.json - Sublime Text (UNREGISTERED)
File Edit Selection Find View Goto Tools Project Preferences Help

mercado_libre.json
1 {
2   {
3     "nombre_producto": "Soporte De Espada De Dos Niveles",
4     "precio_producto": 182900,
5     "inventario_producto": "Cantidad:1 unidad(5 disponibles)",
6     "origen_producto": "Engativa"
7   },
8   {
9     "nombre_producto": "Kit De Limpieza Espadas",
10    "precio_producto": "Katana Samurai",
11    "inventario_producto": "Kenjutsu",
12    "origen_producto": "Ninjutsu"
13  },
14  {
15    "nombre_producto": "Lyuesword Forjado A Mano Damasco Doblado Acero Arcilla Templ",
16    "precio_producto": 3112791,
17    "inventario_producto": "Cantidad:1 unidad(12 disponibles)",
18    "origen_producto": "Bucaramanga"
19  },
20  {
21    "nombre_producto": "Katana Espada Sin Funda Para Decoracion Y Coleccion Cosplay",
22    "precio_producto": 229900,
23    "inventario_producto": null,
24    "origen_producto": "Fontibon"
25  },
26  {
27    "nombre_producto": "Katana Japonesa De Exhibicion",
28    "precio_producto": 350000,
29    "inventario_producto": "Cantidad:1 unidad(5 disponibles)",
30    "origen_producto": "Usaquen"
31  },
32  {
33    "nombre_producto": "Palillos Samurai Katana Japoneses Tres Heros Ieyasu Tokugawa",
34    "precio_producto": 201398,
35    "inventario_producto": "Cantidad:1 unidad(12 disponibles)",
36    "origen_producto": "Bosa"
37  },
38  {
39    "nombre_producto": "Exhibidor En Madera Para Katanas Espadas",
40    "precio_producto": "sables Ejercito Daga",
41    "inventario_producto": 51900,
42    "origen_producto": null
43  },
44  {
45    "nombre_producto": "Para La Espada Danceonlinehasebe Prototipos Escala 1/8 Abs Y",
46    "precio_producto": 1209859,
47    "inventario_producto": "Cantidad:1 unidad(12 disponibles)",
48    "origen_producto": "Floridablanca"
49  },
50  {
51    "nombre_producto": "Saki Katana - Tijeras Japonesas 6 Unidades Color Negro",
52    "precio_producto": 1607900
53  }
54 }
```

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

```
Welcome x db.katanas.insert( [ { ... } ]
MercadoLibreAudifonos localhost:27017 test
db.katanas.insert(
{
  {
    "nombre_producto": "Soporte De Espada De Dos Niveles",
    "precio_producto": 182900,
    "inventario_producto": "Cantidad:1 unidad(5 disponibles)",
    "origen_producto": "Engativa"
  },
  {
    "nombre_producto": "Kit De Limpieza Espadas",
    "precio_producto": "Katana Samurai",
    "inventario_producto": "Kenjutsu",
    "origen_producto": "Ninjutsu"
  },
  {
    "nombre_producto": "Lyuesword Forjado A Mano Damasco Doblado Acero Arcilla Templ",
    "precio_producto": 3112791,
    "inventario_producto": "Cantidad:1 unidad(12 disponibles)",
    "origen_producto": "Bucaramanga"
  },
  {
    "nombre_producto": "Katana Espada Sin Funda Para Decoracion Y Coleccion Cosplay",
    "precio_producto": 229900,
    "inventario_producto": null,
    "origen_producto": "Fontibon"
  },
  {
    "nombre_producto": "Katana Japonesa De Exhibicion",
    "precio_producto": 350000,
    "inventario_producto": "Cantidad:1 unidad(5 disponibles)",
    "origen_producto": "Usaquen"
  },
  {
    "nombre_producto": "Palillos Samurai Katana Japoneses Tres Heros Ieyasu Tokugawa",
    "precio_producto": 201398,
    "inventario_producto": "Cantidad:1 unidad(12 disponibles)",
    "origen_producto": "Bosa"
  },
  {
    "nombre_producto": "Exhibidor En Madera Para Katanas Espadas",
    "precio_producto": "sables Ejercito Daga",
    "inventario_producto": 51900,
    "origen_producto": null
  },
  {
    "nombre_producto": "Para La Espada Danceonlinehasebe Prototipos Escala 1/8 Abs Y",
    "precio_producto": 1209859,
    "inventario_producto": "Cantidad:1 unidad(12 disponibles)",
    "origen_producto": "Floridablanca"
  },
  {
    "nombre_producto": "Saki Katana - Tijeras Japonesas 6 Unidades Color Negro",
    "precio_producto": 1607900
  }
}
)

0.27 sec.
Inserted 1 record(s) in 267ms
```


WinRobo-IT - Download here.

Welcome x db.katanas.find() x

MercadoLibreAudfonos localhost:27017 test

db.katanas.find()

katanas 0.101sec.

```
/* 1 */
{
  "_id" : ObjectId("630d09f0274d81afb356bf58"),
  "nombre_producto" : "Soporte De Espada De Dos Niveles",
  "precio_producto" : 182900.0,
  "inventario_producto" : "Cantidad:1 unidad(6 disponibles)",
  "origen_producto" : "Engetiva"
}

/* 2 */
{
  "_id" : ObjectId("630d09f0274d81afb356bf59"),
  "nombre_producto" : "Kit De Limpieza Espadaas",
  "precio_producto" : "Katana Samurai",
  "inventario_producto" : "Kenjutsu",
  "origen_producto" : "Ninjutsu"
}

/* 3 */
{
  "_id" : ObjectId("630d09f0274d81afb356bf5a"),
  "nombre_producto" : "Espesword Forjado A Mano Damasco Doblado Acero Arcilla Templ",
  "precio_producto" : 3112791.0,
  "inventario_producto" : "Cantidad:1 unidad(12 disponibles)",
  "origen_producto" : "Bucaramanga"
}

/* 4 */
{
  "_id" : ObjectId("630d09f0274d81afb356bf5b"),
  "nombre_producto" : "Katana Espada Sin Funda Para Decoracion Y Coleccion Cosplay",
  "precio_producto" : 229900.0,
  "inventario_producto" : null,
  "origen_producto" : "Fontibon"
}

/* 5 */
{
  "_id" : ObjectId("630d09f0274d81afb356bf5c"),
  "nombre_producto" : "Katana Japonesa De Exhibicion",
  "precio_producto" : 350000.0,
  "inventario_producto" : "Cantidad:1 unidad(6 disponibles)",
  "origen_producto" : "Usaquen"
}

/* 6 */
{
  "_id" : ObjectId("630d09f0274d81afb356bf5d"),
```

136

0.001sec.