

# Webscraper

Cesar Israel Vargas León  
Ramírez Herrera Marco Antonio  
Garduño Reyes Santiago

Diciembre 2023

## WebScrapping

El webscraper se apoya de la librería BeautifulSoup para poder acceder al archivo html de las páginas correspondientes a las que se hará web scrapper. Comenzamos importando las librerías de las que haremos uso, csv para generar el archivo csv con los datos, request para para acceder a la url del sitio, collections para tener los datos almacenados en tuplas y datetime para obtener la fecha actual. Creamos el archivo donde guardaremos los datos llamado “Articulos” Guardamos la url del sitio en una variable url, y los objetos que crearemos en la variable Artículo junto con los atributos que se le darán. Creamos una lista vacía a la que le agregaremos los artículos y sus atributos, pasamos el html a texto con el método text del objeto requests y crearemos el objeto BeautifulSoup para acceder a los elementos del html, primero con este mismo objeto buscaremos las coincidencias dependiendo del sitio en el que nos encontremos para lograr acceder a los productos, haremos uso del método find\_all para encontrar dichas coincidencias. Cabe mencionar que dependiendo del sitio, los elementos a los que haremos referencia se encontraran de una u otra forma, en este caso los productos del html se encuentran listados en una clase, a la que accederemos encontrando todos los objetos de la misma clase. Ya que tenemos los objetos metidos en la lista que genera find\_all accederemos al precio, nombre, etc; iterando en el objeto guardado en la variable articulos\_mercadolibre(esto para el primer sitio), finalmente guardamos los datos obtenidos en un objeto de tipo articulo que se pasara a “leer” para finalmente escribir sus atributos en el csv. Este procedimiento es muy parecido para el resto de productos y sitios, cambiando el modo de acceder a los datos que queremos, dependiendo del html. Cabe destacar que si la pagina no esta abierta para que se le realice webscrapping, este método no servirá.