



MAGÍSTER EN DATA SCIENCE

# Scraper de Biobio Chile

**Proyecto 22** por cechiang, Carloslugook e iggyppopri

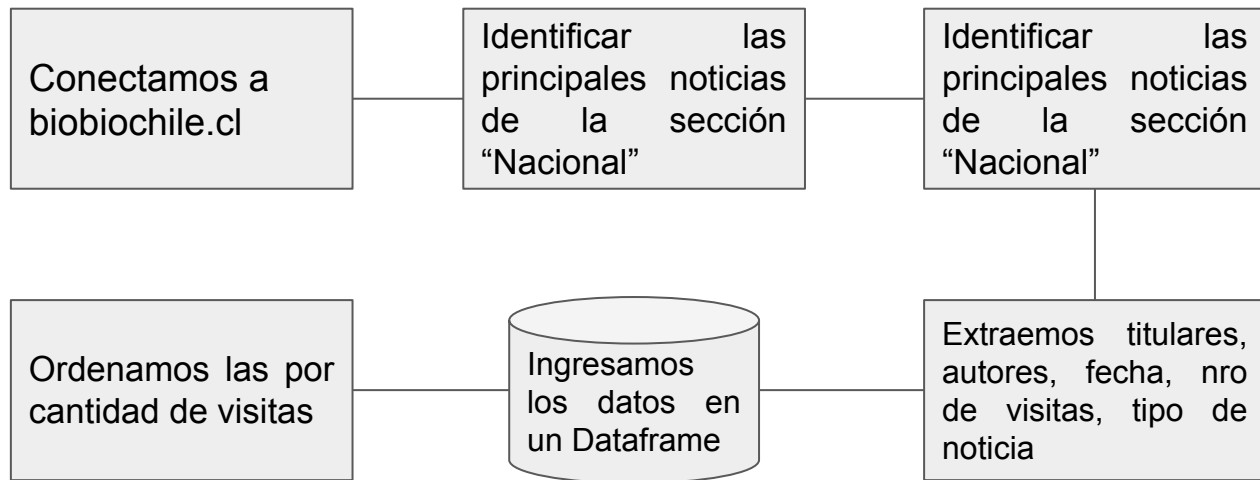
# Pasos a seguir

En el presente trabajo realizaremos web scraping de la página web <https://www.biobiochile.cl/> con beautiful soup.

Se programará un scraper con python para descargar información de las noticias nacionales con más vistas para un periodo de tiempo determinado, identificando el autor y la fecha de la noticias.



# Pasos a seguir



Todo el proceso de web scraping y manipulación de datos lo vamos a realizar con Python con la librería de beautiful soup

El sistema generado podrá servir de base para usuarios que en el futuro requieran realizar análisis de las noticias más visitadas de biobiochile.cl, así como también identificar quienes son los auter mas populares y los días con más visitas

# Pseudocódigo del scraper de BioBioChile

request biobiochile' url

click headline

extract headline link

for i in len(amount of headline links):

- extract these variables: amount of visits, date, headline, author, news type

- save the values in a list or array

- arrange variables and values as data frame

- sort rows by amount of visits

- download as xlsx or csv

create a function for extracting key words from headlines

# Csv final



**Cantidad de filas:** en la sección nacional hay 20 noticias cada día, si programamos el scraper para que funcione durante 5 días, entonces tendremos 100 filas.

**Cantidad de columnas:** Número de visitas, autor, titular, fecha, tipo de noticia