

Scraper de Biobio Chile

Proyecto 22 por cechiang, carloslugook e iggypopri

Pasos a seguir

Se navegará la página web <https://www.biobiochile.cl/> con beautiful soup, pues tenemos un código preexistente que utiliza esta tecnología y funciona.

Se le pedirá al scraper que descargue las noticias con más vistas para un periodo de tiempo ojalá mayor a 1 mes por lo menos, y tanto como se quiera.

De esas noticias se extraerá toda la información que nos pueda entregar el html y además se intentará extraer palabras claves de los títulos, y si es que no es mucho trabajo, del cuerpo de la noticia, para identificar temáticas.

Pasos a seguir

Se analizará la evolución en el tiempo de las temáticas más populares de Biobio Chile, que para efectos de este trabajo diremos que son aquellas con más visitas

Eventualmente se podría replicar esta secuencia de pasos con otro medio de comunicación y comparar, para sacar más conclusiones.

Borrador del scraper en pseudocodigo

for i in len(cantidad de noticias):

 scrap these values of these variables: visits, date, title, hashtags, news type

 arrange variables and values as data frame

 download as xlsx or csv

(desarrollar código más específico para el paso que parte con scrap)

Borrador del csv o xlsx final

Cantidad de filas: Esto dependerá de la cantidad de noticias que seamos capaces de scrapear

Cantidad de columnas: igual a la cantidad de variables, por definir