Índice

1. [Lenguaje y librerías](#_pbvmcb4d0dp9)
2. [Funcionamiento](#_oty13husr7ec)
3. [Ejecución](#_k4h1sk1igr0f)

# Lenguaje y librerías

El programa se ha realizado en lenguaje Python, con el IDE Visual Studio Code.

Las librerías utilizadas son las siguientes:

from operator import indexOf

from urllib import request

import requests

import bs4

import re

import json

from bs4 import BeautifulSoup

from urllib.parse import urljoin

# Funcionamiento

El script consta de 7 métodos, cada uno con una funcionalidad distinta.

-Método **get\_descripciones(url):** Recibe la url de la publicación en cuestión para extraer la descripción del mismo. Con BeautifulSoup llego a la información desde la etiqueta de <h3>

-Método **get\_urls(soup):** Tras recibir el “soup” general del archivo en html, se busca los enlaces de las publicaciones recogidas en una etiqueta <a> mediante su atributo “href”. Luego se guardan en una lista, y este método lo devuelve para su posterior tratamiento.

-Método **get\_ids(url):** La url que recibe como parámetro, debe ser de la publicación correspondiente. Devuelve la referencia de dicha publicación.

-Método **get\_provincias(url):** Devuelve los lugares que aparecen en el texto detallado de cada publicación. Mediante una colección de los lugares de España, se compará con cada cadena recogida gracias a la librería bs4 con la etiqueta donde se encuentra el texto. Luego con un for, se recorre las cadenas del texto y se compará con la lista de lugares de España obteniendo el resultado que se busca. Este resultado se guarda en otra lista que posteriormente, este método, devuelve con “return”.

-Método **get\_firmas(url):** Con bs4, se recogen de las etiquetas “firma\_”, los correspondientes nombres de las publicaciones. Se guardán en una lista, y el método los devuelve para su posterior tratamiento.

-Método **get\_fechas(url):** Se buscan las fechas que aparecen en las publicaciones. Las cadenas del texto se comparan, mediante un for, con una serie de expresiones regulares guardadas en unas variables. Cada una de las fechas se guardará y se devolverá en una lista.

# Ejecución

Antes que nada, se deberá instalar las librerías previamente usadas para el script, en caso de no tenerlas preinstaladas.

El archivo de python deberá ser ejecutado en una IDE que admita la ejecución del mismo (Visual Studio Code, PyCharm, entre otros) o desde una ventana de la terminal del SO Windows.

En caso de ejecutarlo por terminal, se deberá ir a la dirección del archivo mediante el comando “cd”, o abrir la ventana PowerShell en la ubicación exacta en el explorador de archivos, y ejecutar el script.

Se escribe “archivo.py” en el terminal y este se ejecutará.

