

@Es\_Noticia

Mario Sepúlveda Cornejo  
Iñaki Urrutia Sánchez

June 6, 2018

# 1 Introducción

En esta práctica se nos pedía realizar una aplicación de temática libre. Como requisitos indispensables teníamos que:

1. Obtener información de una fuente de datos.
2. Almacenar datos intermedios en la nube.
3. Ofrecer una salida final de datos (estadísticos, gráficas,... ) que será o bien visualizados usando un interfaz web, o almacenados en otro fichero Dropbox/Google drive.
4. Realizar al menos una tarea asíncrona

Teniendo estos requisitos en cuenta decidimos realizar una aplicación que dada una palabra y un número de días, buscara en las cuentas de twitter de los 4 periódicos más leídos en España el número de veces que aparecía esa palabra en los titulares de noticias desde el día indicado hasta hoy, y a partir de los datos conseguidos realizara un tuit con una gráfica que los mostrara.

## 2 Tecnologías utilizadas

Las tecnologías utilizadas en esta práctica son las siguientes:

- Python: Lenguaje de programación utilizado para unificar todas las tecnologías.
- Twitter: Se utiliza para recoger las peticiones de los usuarios, recopilar datos y para mostrar los resultados tras procesar los datos.
- Dropbox: Es donde almacenamos los datos recogidos en Twitter.
- Celery: Nos permite realizar tareas asíncronas.

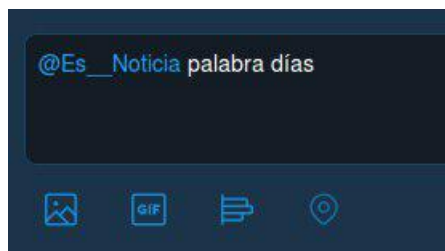
## 3 Funcionamiento

### 3.1 Twitter

Para empezar creamos una cuenta en Twitter a partir de la cual pudieramos recibir y responder las peticiones de los usuarios. Luego creamos una API de Twitter a partir de la cuenta creada desde <https://apps.twitter.com/>



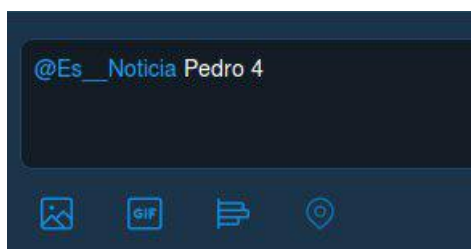
La cuenta espera a recibir nuevas peticiones de búsqueda con el siguiente formato:



Donde:

- Palabra es la palabra que se quiere buscar entre los titulares de las noticias
- Días es el número de días que se quiere buscar. En el caso en el que no se incluya el número de días su valor será 1 por defecto.

Ejemplo:



En este ejemplo se buscaría el número de apariciones de la palabra Pedro en los últimos 4 días. Una vez recibida la petición se buscarían coincidencias de la palabra en las siguientes cuentas de Twitter:

- @elpais\_espana
- @elmundoes
- @LaVanguardia
- @20m

Una vez procesados los datos en Python, una gráfica con los resultados es subida a Twitter como respuesta a la petición del usuario:



Como podemos observar, si no indicamos el número de días, se buscarán sólo en las noticias del día actual:



## 3.2 Dropbox

Para guardar los datos extraídos en la nube creamos una cuenta de Dropbox y su respectiva API desde <https://www.dropbox.com/developers/apps>

En Dropbox guardamos los ficheros en formato .csv con los tweets extraídos de cada cuenta de Twitter mencionada anteriormente. Cada cuenta tiene su propio fichero .csv. Además guardamos las menciones que ya han sido respondidas en un fichero llamado Menciones.csv para poder detectar cuando hay nuevas menciones.

Por seguridad, además guardamos copias de respaldo de todos los ficheros.

### 3.3 Celery

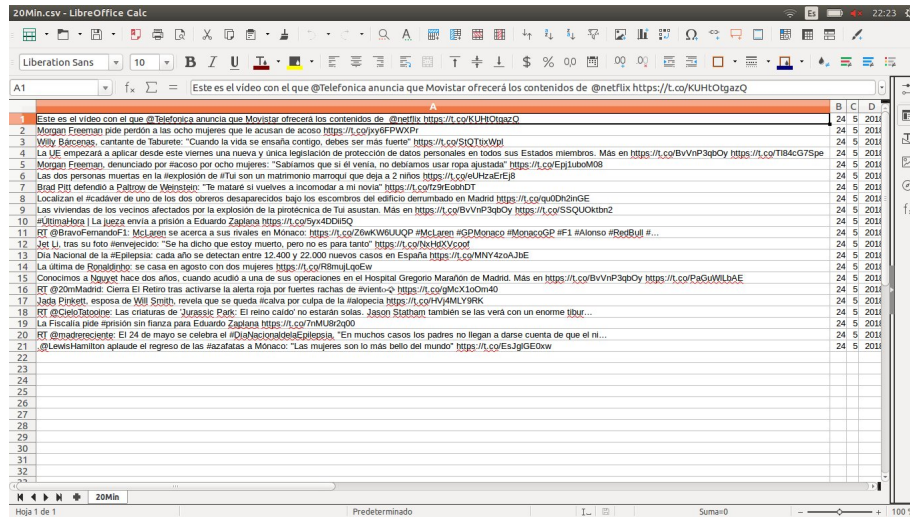
Al tener que realizar al menos una tarea asíncrona decidimos utilizar esta tecnología ya que era sencilla y cómoda de utilizar.

Aplicamos la petición asíncrona a la hora de contar el número de veces que aparece una palabra en los tuits extraídos. De esta manera conseguimos mejorar el rendimiento ya que ejecutábamos cuatro peticiones de forma paralela.



### 3.4 CSV

Usamos ficheros .csv como base de datos para guardar/cargar los tuits que previamente descargamos de Twitter. El formato que siguen los .csv es el mostrado en la siguiente imagen:



20Min.csv - LibreOffice Calc		
Liberation Sans 10 B I U T A % 00		
A1	Este es el video con el que @Telefonica anuncia que Movistar ofrecerá los contenidos de @netflix https://t.co/KUHTotgazQ	
1	Este es el video con el que @Telefonica anuncia que Movistar ofrecerá los contenidos de @netflix https://t.co/KUHTotgazQ	24 5 2014
2	Morgan Freeman pide perdón a las ocho mujeres que le acusan de acoso https://t.co/y6FPWXP	24 5 2014
3	Willy Baccus, cantante de Taburete: "Cuando la vida se enseña contigo, debes ser más fuerte" https://t.co/SQ1toWj	24 5 2014
4	La UE empezará a aplicar desde este viernes una nueva y única legislación de protección de datos personales en todos sus Estados miembros. Más en https://t.co/BvVnVP3qOy https://t.co/TtB4cG7Spe	24 5 2014
5	Morgan Freeman, denunciado por acoso por ocho mujeres: "Sabíamos que si él venía, no debíamos usar ropa ajustada" https://t.co/Epl1ubm08	24 5 2014
6	Las dos personas muertas en la explosión de #Tui son un matrimonio marroquí que deja a 2 niños https://t.co/ehztaEJE8	24 5 2014
7	Brad Pitt defendió a #Estepen: "Te mataré si vuelves a incomodar a mi novia" https://t.co/t2WESaDT	24 5 2014
8	Localizan el cadáver de uno de los dos obreros desaparecidos bajo los escombros del edificio demolido en Madrid https://t.co/9u0Dh2wGE	24 5 2014
9	Las viviendas de los vecinos afectados por la explosión de la protécnica de Tui asustan. Más en https://t.co/BvVnVP3qOy https://t.co/SSQUkttbn2	24 5 2014
10	#UltimaHora   La jueza envía a prisión a Eduardo Zaplana https://t.co/5yx4Dm5Q	24 5 2014
11	RT @BavofemandoF1: McLaren se acerca a sus rivales en Mónaco: https://t.co/2WkKWUJQD #McLaren #GPMonaco #MonacoGP #F1 #Alonso #RedBull #...	24 5 2014
12	Jet Li, tras su foto #envejecido: "Se ha dicho que estoy muerto, pero no es para tanto" https://t.co/NxH0XVcod	24 5 2014
13	Día Nacional de la #Epilepsia: cada año se detectan entre 12.400 y 22.000 nuevos casos en España https://t.co/MNY4zoAJNE	24 5 2014
14	La última de #Rozalby: se casa en agosto con dos mujeres https://t.co/8mujapCw	24 5 2014
15	Conocimos a Nguet hace dos años, cuando acudí a una de sus operaciones en el Hospital Gregorio Marañón de Madrid. Más en https://t.co/BvVnVP3qOy https://t.co/PaGvWLBaE	24 5 2014
16	RT @20nMadrid: Cierra El Retiro tras activarse la alerta roja por fuertes rachas de viento-> https://t.co/gMcK1oOm40	24 5 2014
17	Jada Pinkett, esposa de Will Smith, revela que se queda #calva por culpa de la alopecia https://t.co/HV4MLY9K	24 5 2014
18	RT @CrisioBatorre: Las ciruelas de Jureiss: País. El reino caído no estarán solos. Jasson Sibson también se las verá con un enorme #bur...	24 5 2014
19	La Fiscalía pide prisión sin fianza para Eduardo Zaplana https://t.co/7mMUB2q00	24 5 2014
20	RT @madrescientee: El 24 de mayo se celebra el #DíaNacionaldeEpilepsia. "En muchos casos los padres no llegan a darse cuenta de que el ni...	24 5 2014
21	@LewisHamilton aplaude el regreso de las #azulatas a Mónaco: "Las mujeres son lo más bello del mundo" https://t.co/ESJyGDEow	24 5 2014
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
Hoja 1 de 1 Predeterminado Suma=0 100 %		

En el primer campo, guardamos el tuit, y en los campos siguientes guardamos DÍA MES y AÑO para poder implementar la búsqueda según un período de tiempo concreto.

## 4 Ejecución

Los pasos para ejecutar el bot son los siguientes:

1. Nos ubicamos en el directorio donde están ubicados los ficheros .py
2. Abrimos tres terminales:
  - (a) En el primer terminal ejecutar el comando: `celery -A tareas worker`
  - (b) En el segundo terminal ejecutar el comando: `flower`
  - (c) En el tercer terminal ejecutar el comando: `python3 main.py`
3. Mencionamos a la cuenta de Twitter (@Es\_\_Noticia) con nuestra consulta.
4. El bot responde a nuestro tuit con la gráfica correspondiente.