

Es Noticia

Bot para conteo de palabras en los principales periódicos de España.

Mario Sepúlveda Cornejo, Iñaki Urrutia Sánchez

1 Introducción

Este proyecto desarrolla las siguientes competencias:

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 esto en cuenta decidimos realizar una aplicación que dada una palabra y un número de días buscara en 4 cuentas de twitter sobre noticias el número de veces que aparecía esa palabra en los últimos número de días indicado, y a partir de los datos conseguidos creará un tuit con una gráfica que los mostrara.

2 Tecnologías

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:

- Palabra es la palabra que se quiere buscar.
- 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.

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.

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.

4 Ejecución

Los pasos para ejecutar el programa son los siguientes:

- Nos ubicamos en el directorio donde están ubicados los ficheros .py
- Abrimos tres terminales.
 - En el primer terminal ejecutar el comando: `celery -A tareas worker`
 - En el segundo terminal ejecutar el comando: `flower`
 - En el tercer terminal ejecutar el comando: `python3 main.py`
- Enviamos una petición a la cuenta de Twitter.
- Esperamos a que la cuenta nos responda con la gráfica.