Análisis de datos de Twitter

|  |
| --- |
| Miguel Angel Pozo Jacome, Bryan David Guerrero Velasco, Jefferson Llumiquinga |
| Escuela de formación de tecnólogos |
| Escuela Politécnica Nacional |
| Ladrón de Guevara E11·253 Quito · Ecuador  [Miguel.pozo@epn.edu.ec](mailto:Miguel.pozo@epn.edu.ec), |

*Resumen*- La obtención de los datos se darán mediante las API y claves proporcionadas por Twitter con los que se sacaran conclusiones sobre tendencias o el número de usuarios que hablan sobre ciertos tópicos.

*Palabras Clave*- Twitter, análisis, claves.

**Introducción**

El análisis se realizará obteniendo los datos de Twitter directamente, pero para ello se debe de aplicar al proceso de aprobación de Twitter Developer.

El proceso se basa principalmente en responder un numero reducido de enunciados, donde lo primordial es averiguar de que forma o para que se van a usar los datos a partir de las API y claves.

Además, todos los datos serán guardados en una base de datos la cual usaremos como reflejo para utilizar los datos guardados en un HTML.

**herramientas**

A lo largo de nuestro analisis las herramientas usadas son las API de Twitter, Visual Studio Code y CouchDB.

Estos programas proporcionan un interfaz visual y amigable con el usuario para crear un entorno de uso mas sencillo.

Las API de twitter se obtendran tras aprobar el pequeño procesos que existe y las cuales serviran para acceder a los datos necesarios de Twitter.

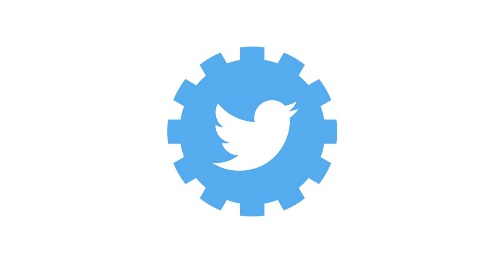


Ilustración 1 "Logo Twitter Developer"

La siguiente herramienta a usar es Visual Studio Code que nos servira como editor de codigo y nos proporcionara live server para observar los datos obtenidos.

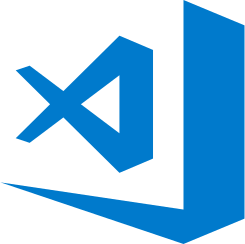


Ilustración 2 "Logo Visual Studio Code"

Y por ultimo, se usara CouchDB que es una base de datos de codigo abierto con un facilidad de uso y accesibilidad ya que se puede ejecutar en la web.

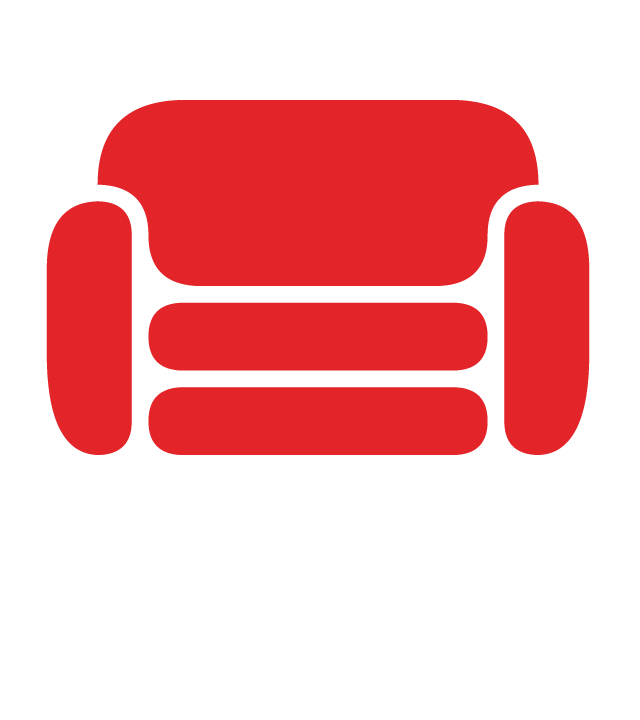


Ilustración 3 "Logo CouchDB"

# Eclipse Solar 2019

Los datos que se analizaran en este primer caso esta determinados por el hashtag ***#EclipseSolar2019***.

El primer paso para obtener los datos es obtener las API y tokens respectivos que se nos proporcionan en Twitter Developer despues de crear nuestra primera app.

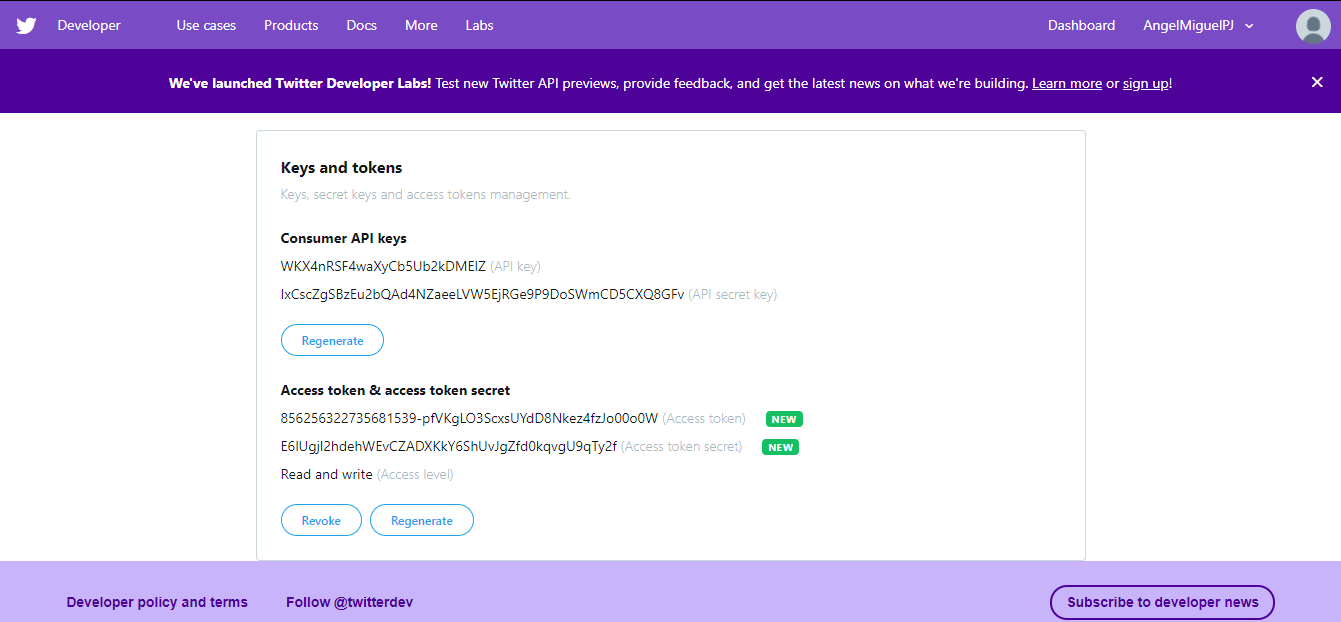


Ilustración 4 "API y Tokens Twitter"

A continuación, se dara paso a modificar el archivo **cosechar\_tweets.py** proporcionado en clases, el cual modificaremos cambiando las API y Tokens. Además, debemos modificar el nombre de la base de datos a crear en CouchDB y por ultimo poner el nombre del hashtag a buscar.

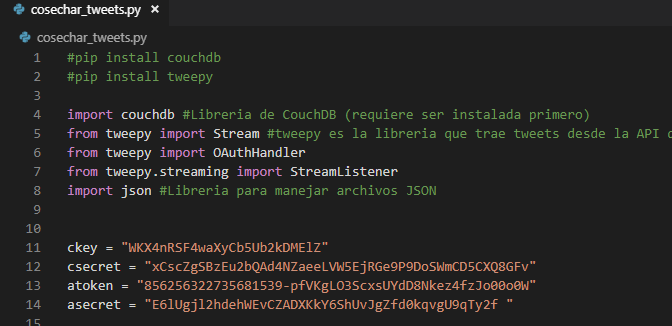


Ilustración 5 "Cambio de API y Tokens"

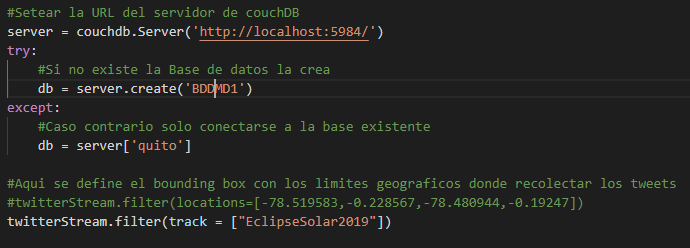


Ilustración 6 "Base de datos y Hashtag"

Después, ejecutaremos el archivo desde la terminal de nuestra computadora para comenzar el proceso de recolección de datos, pero antes debemos importar las librerías de CouchDB y de Twitter como lo haremos de la siguiente manera.

1. Librería de CouchDB.

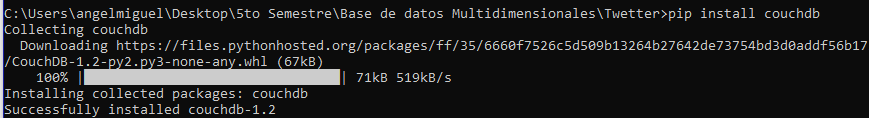


Ilustración 7 "Librerias CouchDB"

1. Actualización **pip** para python.

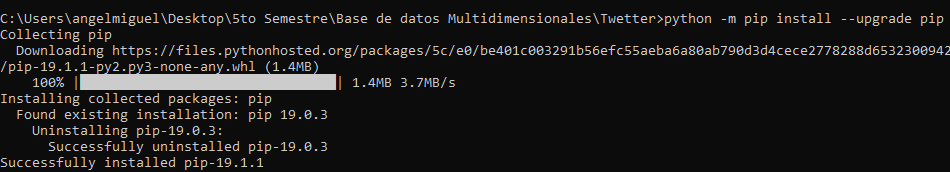


Ilustración 8 "Actualización pip"

1. Librerías Twitter.

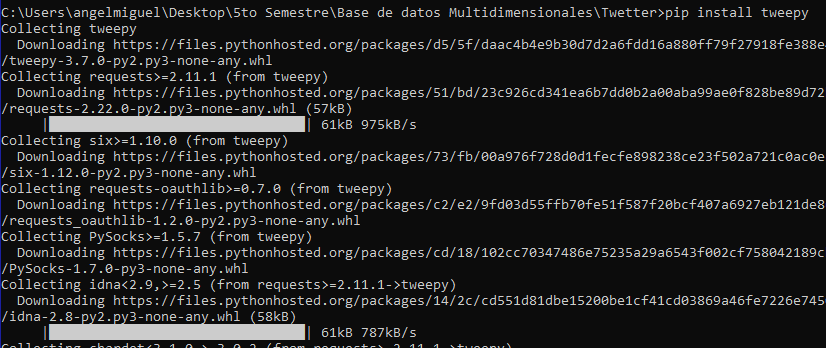


Ilustración 9 "Librerías Twitter"

Luego de la instalación de las librerías procederemos a abrir el archivo y empezara la recolección de datos.

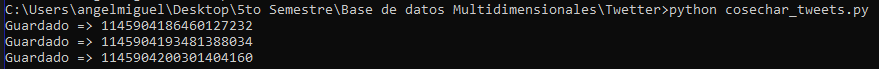


Ilustración 10 "Recolección de datos"

Ya empezado el proceso de recolección empezaremos a ver como se guardan en la base de datos y como el número de registros aumenta.



Ilustración 11 "Base de datos de la recolección"

Continuando con el proceso de recolección, daremos paso al la pagina princiapal donde tendremos las diferentes opciones para el reflejo de los datos según los temas o tendencias escogidas mediante el archivo **index.html**

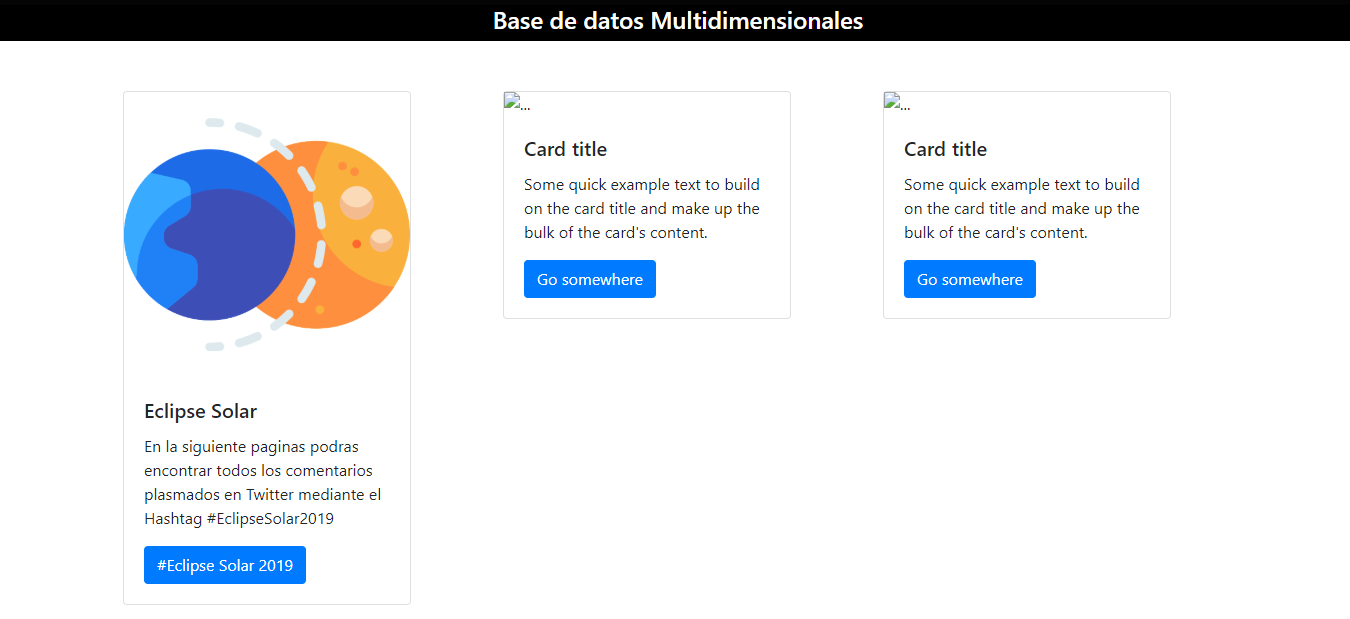


Ilustración 12 "Index - Ventana Principal"

Luego daremos Click en el boton ***“#Eclipse Solar 2019”***, y acederemos a la página donde visualizaremos todos los datos obtenidos hasta ahora, los cuales continuaran actualizandose.

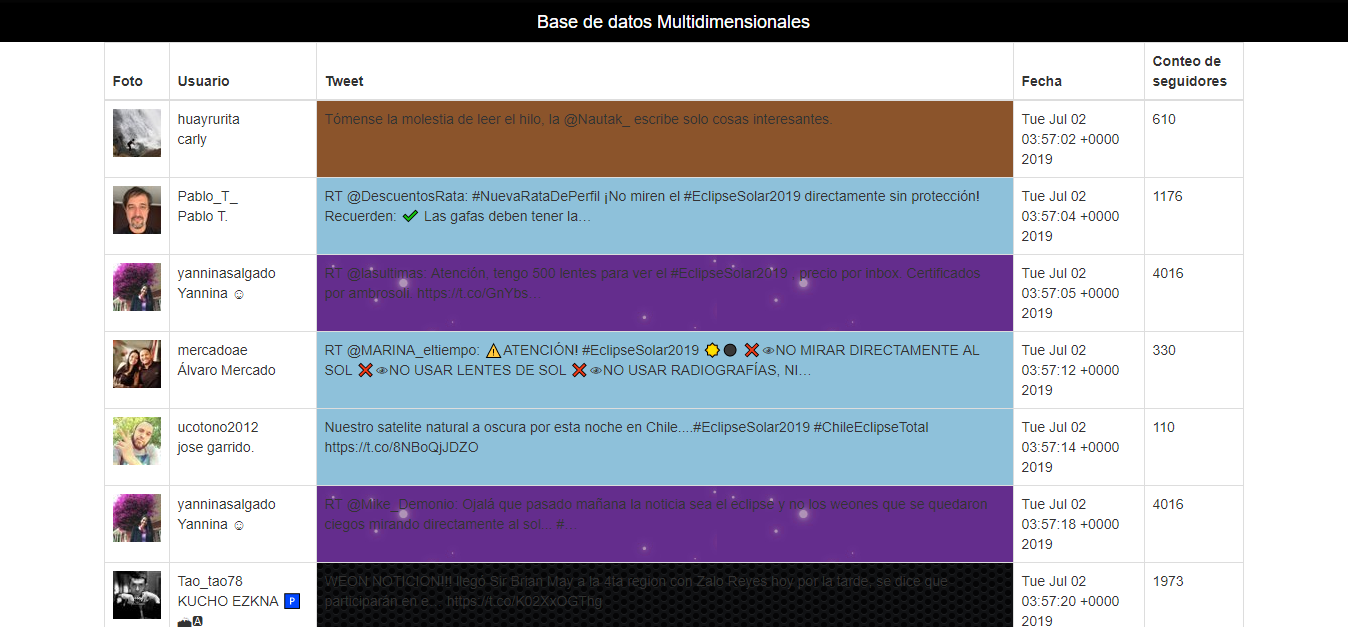


Ilustración 13 "Datos #EclipseSolar2019”

# Copa America

Como estamos en época de futbol aprovechamos la cantidad de datos que se pueden ir generando por la Copa América que se está realizando en Brasil.

Obtuvimos una cierta cantidad de api´s que nos ayudan a buscar datos más fácilmente.

Apis ocupadas en esta búsqueda:

ckey = "m4Fq2Pr4yHn1YLLg6nmPYxXYz"

csecret = "0WC0ZlD9sT4aMiBY5xDrRjFIQqT3KbU8oSaNEsFEkKPHZCASe4"

atoken = "999027411613356032-NvGF9YveYjVjQq4sf61x5IbFDe0KBej"

asecret = "ZHTEn2rxBoKLIkFa57Ksm3Hs4jYtwimhgYcsq8TtAXVXv"

en cada computador que se realiza este tipo de búsquedas debemos revisar si se encuentran instaladas esta librería en el terminal de Windows

#pip install couchdb

#pip install tweepy

Busqueda de datos con la frase #CopaAmerica2019

**C:\Users\jefer\OneDrive\Escritorio>python cosechar\_tweets.py**

**Guardado => 1145936209480474625**

Dentro de couchDB observamos que los datos se van guardando, el encuentro de datos varia según se haya ocupado la frase o palabra dentro de twitter



Ilustración 14 "Datos #CopaAmericar2019”

# Huawei

Seguimos con la recolecciòn de datos, en este caso nos basaremos en el #Huawei cambiando el archivo cosechar-3.py

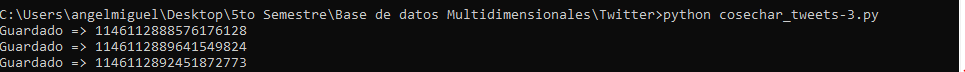


Ilustración 5 Recolecciòn

Ya empezado el proceso de recolección empezaremos a ver como se guardan en la base de datos y como el número de registros aumenta.



Ilustración 16 Base de Datos

Se añadirá a la página de html la sección el la cual presentaremos la secciòn que contendra todos los tweets de Huawei.



Ilustración 17 Sección Huawei

Luego daremos Click en el boton ***“#Visualizar #Huawei”***, y acederemos a la página donde visualizaremos todos los datos obtenidos hasta ahora, los cuales continuaran actualizandose.

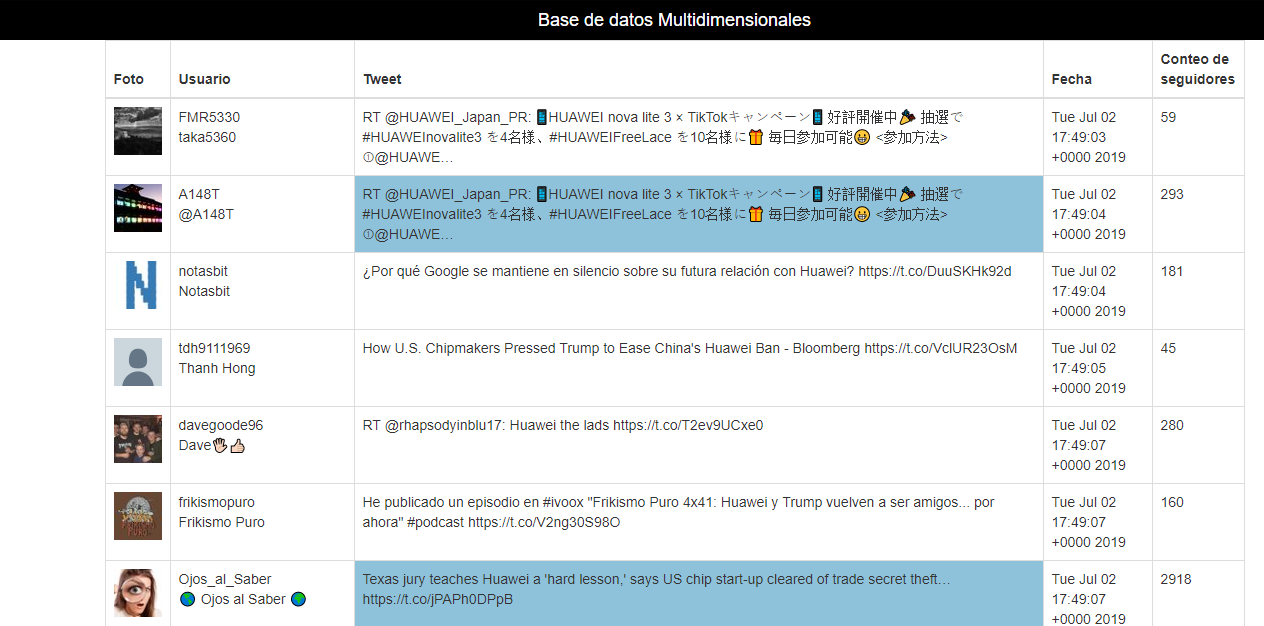


Ilustración 18"Datos #Huawei”

# Ubicación

Datos encontrados en la recolección de datos por ubicación

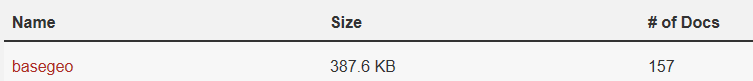


Ilustración 19"Datos en couchDB”

Ubicación de la búsqueda en la provincia del Guayas



Ilustración 20"Geolocalización”

Localización

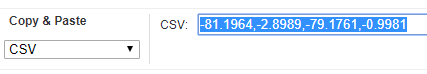


Ilustración 21