

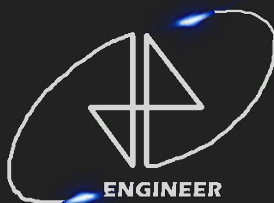
JPE_Twitter

Aplicación Splunk para Minería de Twitter

El presente documento fue pensado, diseñado y redactado única y exclusivamente para la comprensión general del aplicativo JPE_Twitter, por lo tanto, no puede ser editado y distribuido sin la previa autorización escrita de su autor.

Mayo de 2019

Juan Alejandro Pérez Chandía
Ingeniero Informático
jalejandro.ingeniero@gmail.com



JPEngineer ©

Introducción

JPE_Twitter es una aplicación Splunk desarrollada en Python, cuya finalidad es la minería en Twitter e integración completa con la herramienta de análisis de logs anteriormente mencionada. Con JPE_Twitter podrás extraer de Twitter información relevante basada en filtros de palabras, frases, oraciones, hashtag, etc. y obtener datos que indican lo que está pasando en un negocio o servicio determinado en dicha red social, construyendo así, una base de desarrollo para el análisis de sentimiento en Twitter.

Generalidades del Desarrollo

El aplicativo esta desarrollado en python 2.7, con la finalidad de compartir la instalación base de python de la plataforma Splunk (6.x o superior) y cuenta con cinco módulos esenciales:

- **config** : El aplicativo cuenta con un archivo de configuración llamado *twitter.conf*, donde el modulo config lee y escribe los parámetros necesarios para la autenticación del listener con la API de Twitter, configuración base de logs, encriptación, entre otros.
- **security** : Para la seguridad de datos sencibles (Token autenticación), este módulo cuenta con lo necesario para cifrar y descifrar en base a la clave que el usuario determine en la configuración de los token en Splunk.
- **splunk_read** : El objetivo de este módulo es tomar todos los parámetros ingresados desde la plataforma Splunk y llevarlos al módulo correspondiente (config, security, service, entre otros).
- **service** : Encargado de iniciar, detener y reiniciar el servicio, además de obtener su estado.
- **twitter** : Corresponde al listener propiamente tal, quién retornará todo lo filtrado en Twitter y posteriormente ser indexado en Splunk.

Además, JPE_Twitter cuenta con sus propios logs, para hacer debug del proceso en curso mediante splunk o visualizando directamente los archivos logs:

- **config.log** : Muestra el proceso ejecutado por el módulo *config*, incluyendo sus problemas (si existe alguno).
- **service.log** : Muestra el estado del servicio y su PID correspondiente, además incluye los problemas que dicho servicio pudiese presentar.
- **twitter.log** : Muestra el estado de la autenticación con la API de Twitter y los mensajes filtrados desde dicha red social.
- **splunk_read.log** : Muestra el proceso ejecutado por el módulo *splunk_read*, incluyendo sus problemas (si existe alguno).

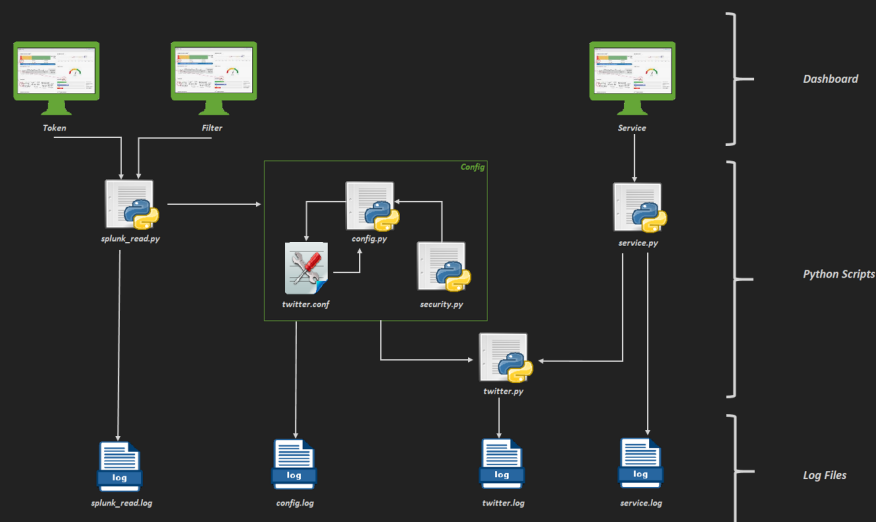


Fig. 1: JPE_Twitter General Diagram

Importante: Todas las librerías necesarias para el correcto funcionamiento del aplicativo, se encuentran incluidas en el paquete de instalación

Dashboards

Para la administración del proceso de escucha de Twitter existen tres dashboard importantes, que permiten configurar los token de autenticación, configurar los filtros correspondientes e iniciar, detener, reiniciar o verificar el estado del servicio:

- Token and Security
- Filter
- Service

Además, cuenta con tres dashboard básicos que permiten visualizar información general de Twitter, en base a los filtros aplicados, tales como: usuario con más seguidores, hashtags más usados y usuarios mencionados:

- Twitter Users
- Twitter Hashtag
- Twitter Comments

Token and Security

Este dashboard permite configurar los token de autenticación de la API de Twitter, los cuales son: *consumer key*, *consumer secret*, *access token* y *access secret*. Estos token quedan almacenados en el archivo de configuración *twitter.conf* y en lo que a seguridad respecta, los token de autenticación pueden ser encriptados para evitar que terceros accedan a ellos, solo es necesario escribir una clave de encriptación en el mismo panel, luego de indicar que deseas encriptar los token.

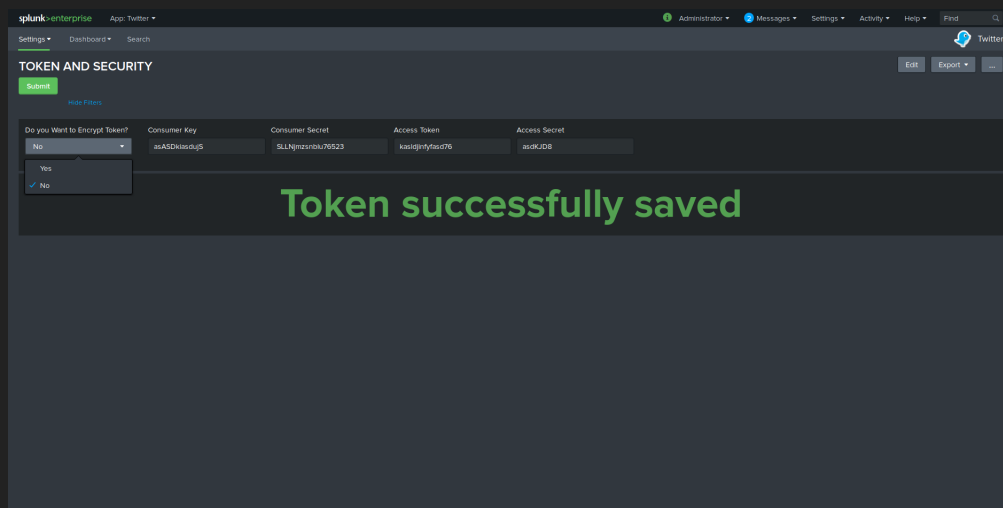


Fig. 2: Setting: Token and Security Dashboard

Es importante considerar que: para obtener los token, se debe crear un cuenta de desarrollador en Twitter (<https://developer.twitter.com>)

Filter

Este dashboard permite configurar los filtros que se aplicarán en el proceso de escucha. Las opciones de configuración incluyen la acción de agregar nuevo filtro al que se encuentra actualmente configurado o configurar nuevos filtros reemplazando a los anteriores, mediante la elección de *add* o *new* respectivamente.

Los filtros ingresados pueden ser palabras, frases, hashtag, entre otros y deben ser separados por una coma.

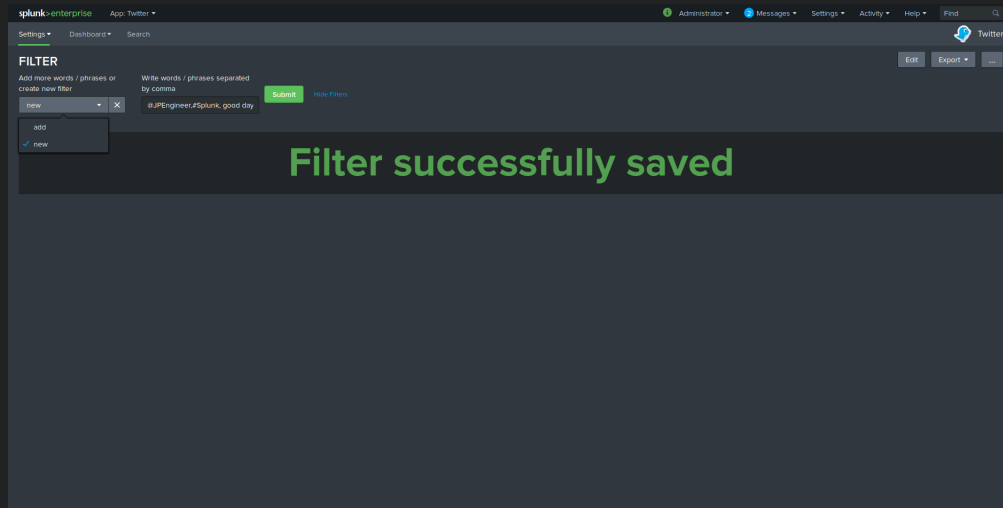


Fig. 3: Setting: filter Dashboard

Service

Este dashboard permite ejecutar acciones sobre el servicio de Twitter. La opción Start permite iniciar el servicio y comenzar a indexar data obtenida desde Twitter y como requisito previo, se requiere tener configurado los Token de autenticación y los filtros. La opción Stop es la encargada de detener el servicio. La opción Restart permite reiniciar el servicio si es requerido por desarrollos propios u otras razones. La opción Status indica si el servicio se encuentra activo o detenido.

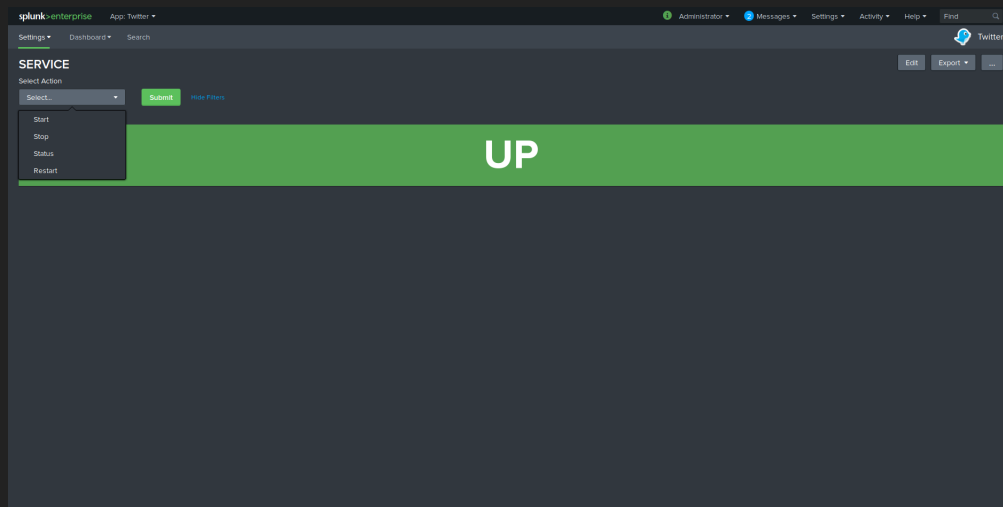


Fig. 4: Setting: Service Dashboard

Twitter Users

Este dashboard presenta información relacionada a usuarios de la red social, tales como: cantidad de seguidores, nombre, alias y ubicación

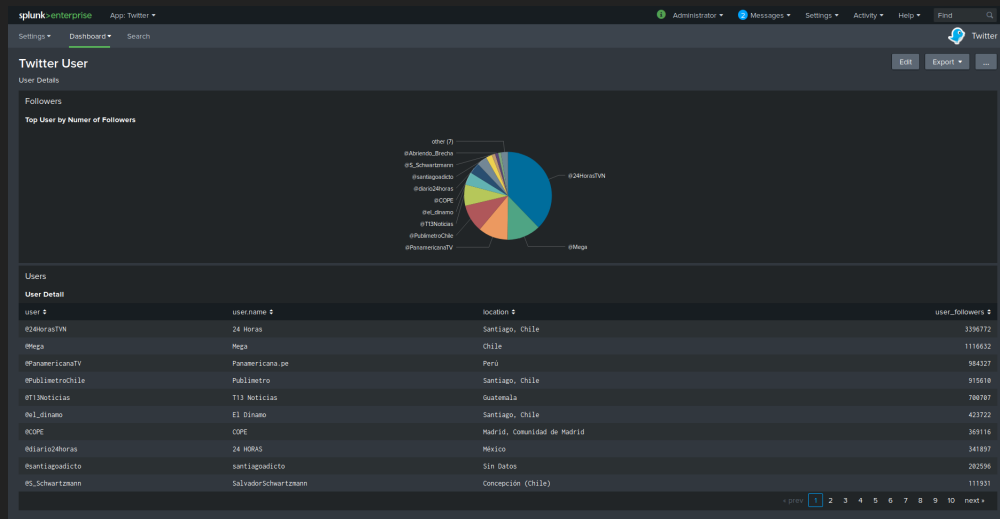


Fig. 5: Twitter User Dashboard