



Software Solutions

Twitter Data Analysis

Manual de Usuario

Manejo de Datos no Estructurados

Dr. Francisco J. Moo Mena

Integrantes:

Daniel Fernando Baas Colonia

Rodrigo Castilla López

Luis Gerardo Castillo Pinkus

Rafael Rodríguez Guzmán

Introducción

Nintendo es una compañía que desarrolla videojuegos. Está posicionada a nivel mundial como una de las que más efectivo genera gracias a su productos, y su importancia en el mercado es indudable. Son responsables de grandes franquicias de videojuegos como Super Mario Bros, The Legend of Zelda o Pokémon. Cuentan con una base de usuarios enorme alrededor del mundo y cada juego que lanzan es sumamente esperado.

Una de las franquicias más importantes para Nintendo es Super Smash Bros. Este consiste en un juego de lucha, donde cuatro jugadores se enfrentan en una arena usando personajes distintivos de Nintendo, tales como Mario, Link o Pikachu. Dado que esta franquicia genera mucha expectativa entre sus usuarios, esta esperan con ansias cada vez que hay una nueva entrega del juego. Cada jugador selecciona su personaje favorito de su juego favorito y lo usa para enfrentar a otros jugadores. La mecánica de que puedas ver a tu personaje favorito es una de las cosas que los usuarios más disfrutan. Sin embargo, no siempre se incluyen todos los personajes que los jugadores desean, ya sea por cuestiones de tiempo, dinero (licencias) o esfuerzo. Por lo tanto, la mayoría de los jugadores de Super Smash Bros siempre están al tanto de qué personaje nuevo de incluirá en cada entrega.

Nintendo realiza eventos digitales cada cierto tiempo llamados "Nintendo Direct". Este consiste en un stream de video donde algún representante de la compañía habla sobre los próximos juegos que saldrán, anuncios nuevos, actualizaciones, etc. Uno de los últimos eventos de este año fue un Nintendo Direct dedicado completamente a mostrar información sobre el próximo juego de Super Smash Bros, que tiene fecha de salida para el próximo 7 de Diciembre.



1.11.2018

Los eventos digitales de Nintendo siempre generan mucho revuelo puesto las personas siempre esperan ver cosas nuevas e interesantes, y en el caso del Nintendo Direct de Super Smash Bros, las personas esperan ansiosamente por ver qué personajes nuevos incluirá la última entrega del juego.

Así, redes sociales como Twitter se llenan de información cada vez que un evento digital ocurre. Por lo tanto, para este proyecto se decidió recopilar y analizar tweets generados por usuarios durante el último Nintendo Direct de Super Smash Bros. Twitter ofrece una oportunidad de obtener información interesante sobre qué opinan los usuarios sobre el evento mientras ocurre, así como cuáles fueron sus expectativas al iniciar, y cómo quedaron sus opiniones al terminar. Es posible ver que durante momentos importantes del stream, los usuarios reaccionaron, y eso se refleja en el flujo de tweets.

Para almacenar tal cantidad de información, y tener una fácil forma de manipularla, es conveniente usar MongoDB. MongoDB es una base de datos no relacional, por lo cual, cada documento que se almacene es compatible con la forma anidada de elementos que contiene cada tweet.

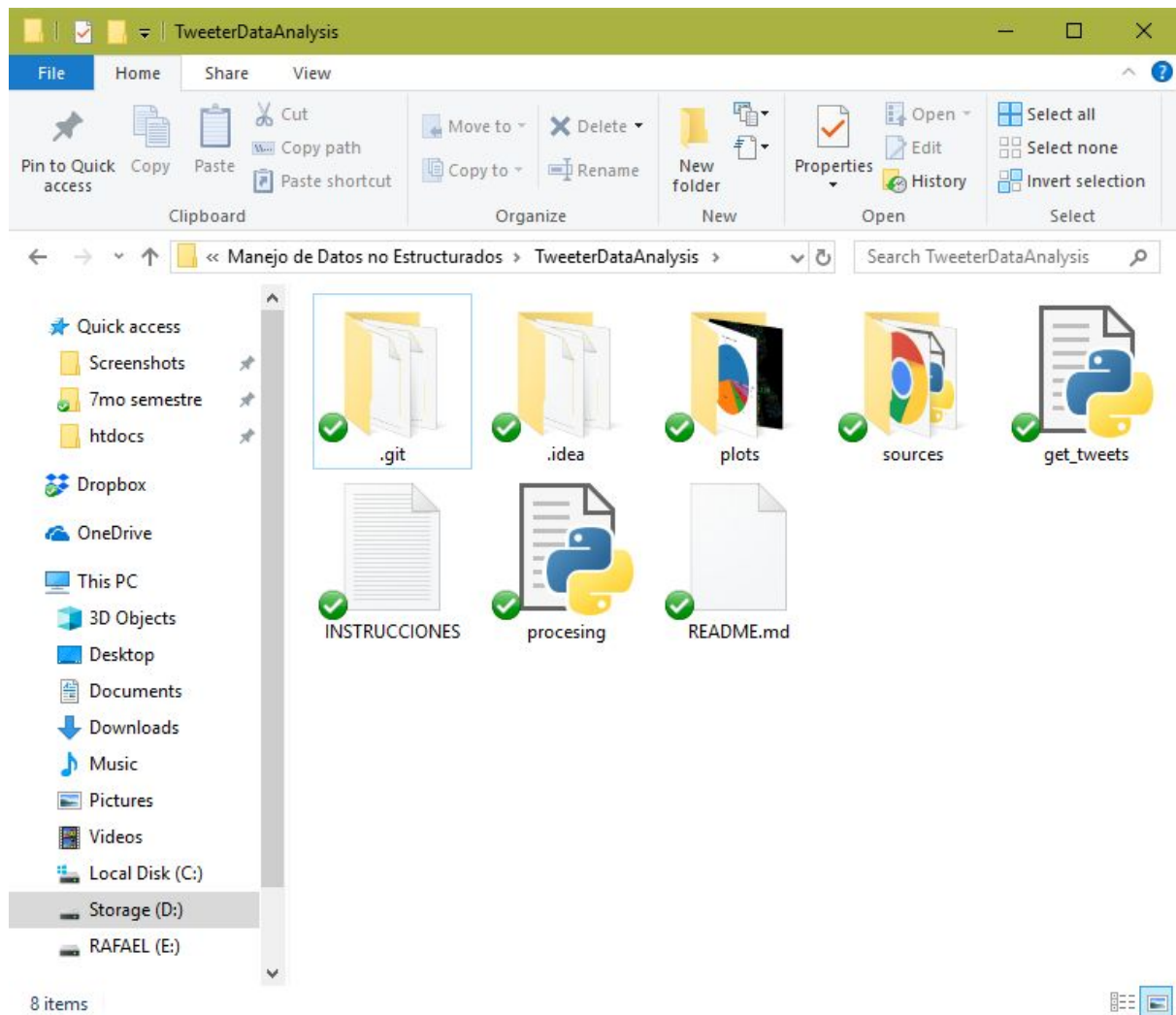
Para procesar toda la información obtenida, el lenguaje de programación Python resulta fácil, amigable y útil. Cuenta con una enorme cantidad de bibliotecas para obtener y procesar datos, así como generar gráficas que muestren los resultados obtenidos.

Así, se planea obtener tweets de Twitter, almacenarlos en la base de datos en línea de MongoDB, y procesarlos usando Python. Se desea ver qué fue lo que los usuarios mencionaron más, qué países estuvieron más activos, qué idiomas fueron los más usados y en qué aplicaciones de software se realizaron la mayor cantidad de tweets.

Para finalizar, creemos que se puede observar información interesante en cómo reaccionaron los espectadores del evento ante ciertas partes el mismo. También creemos que sería interesante escalar el proyecto a futuro, y orientarlo a realizar un análisis de sentimientos. Con esto podría observarse mejor cuáles eran las expectativas de los usuarios ante ciertos elementos y cuáles fueron las opiniones al final.

Guia de Uso

El proyecto se divide en dos partes: la obtención de tweets y el procesamiento de ellos. Ambas partes se procesan en scripts diferentes. Mientras que la obtención de tweets se realiza usando el script *get_tweets.py*, el procesamiento de ellos se realiza en el script *procesing.py*.



La primera parte corresponde a la obtención de tweets. Para ello debes hacer uso de tu consola de comandos y acceder a la ubicación del archivo *get_tweets.py*. Para ello debes ejecutar el siguiente comando:

Windows: *python get_tweets.py*

Linux: *python3 get_tweets.py*

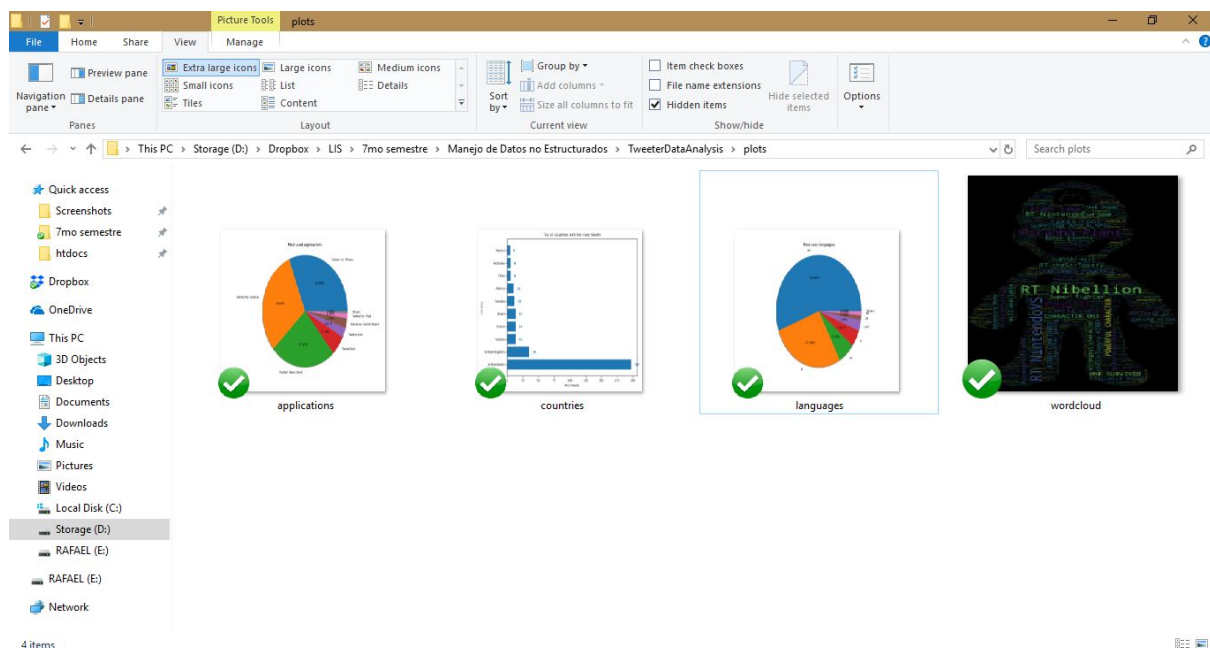
La consola empezará a mostrar tweets de manera continua. El script recibe los tweets y los guarda en un servidor en línea (Atlas) de MongoDB. Es necesario estar

registrado en el servidor y proporcionar tus datos en el script antes de poder ejecutarlo.

El script se mantendrá activo hasta que pierda la conexión o pares el proceso. En caso de desconexión, simplemente vuelve a ejecutar el comando en consola mostrado anteriormente.

La segunda parte consiste en el procesamiento. Si lo que quieres es ver el funcionamiento como tal, este es el script que debes ejecutar.

procesing.py recupera los tweets de la base de datos y realiza distintos procesamientos sobre ellos. Al final lo que genera es una serie de imágenes correspondientes a distintas gráficas. Estas son: una nube de palabras con los términos más usados en inglés, una gráfica de pastel con los siete idiomas más usados, una gráfica de barras con los diez países que más realizaron tweets, y una gráfica de pastel con las siete aplicaciones más usadas para realizar los tweets. Las gráficas se muestran a continuación:



Para ejecutar el script, es necesario que abras una consola de comandos y te ubiques en la carpeta donde se encuentra el script *procesing.py*. Para ejecutarlo debes escribir el siguiente comando:

Windows: *python procesing.py*

Linux: *python3 procesing.py*

```
rafael@rafael: ~/Dropbox/LIS/7mo semestre/Manejo de Datos no Estructurados/TweeterDataAnalysis$ python3 procesing.py

La base de datos "nintendo" existe!
La coleccion "tweets" existe!

Generando nube de palabras...

Generando gráfica de dispositivos...

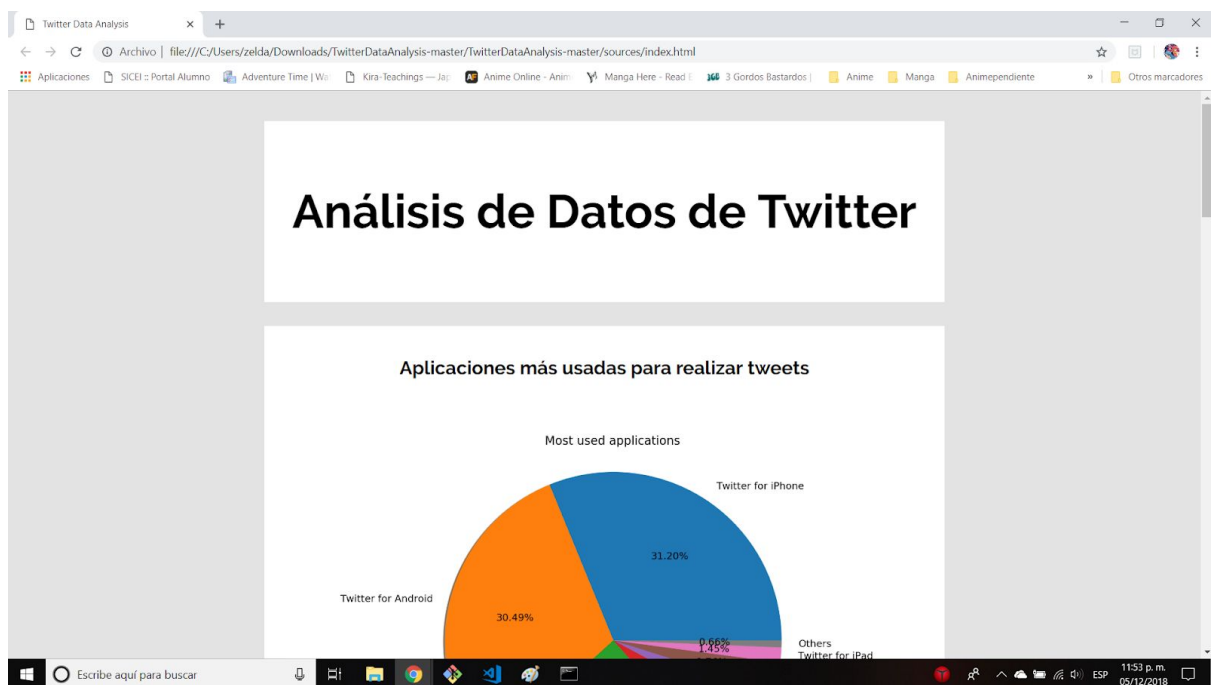
Generando gráfica de países...

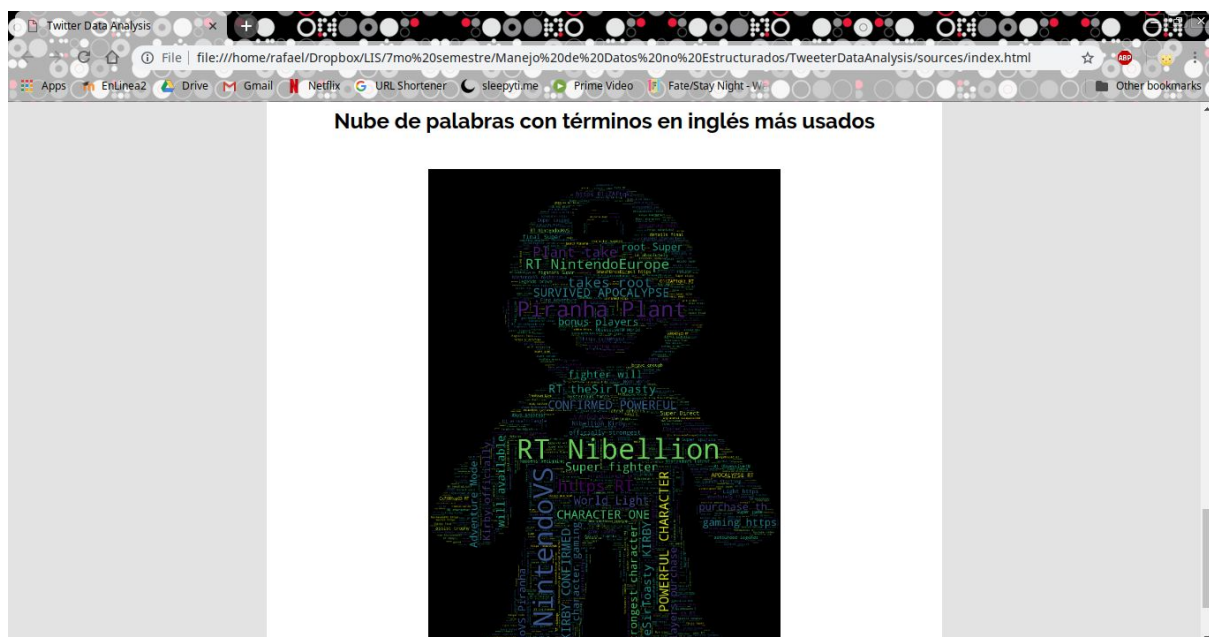
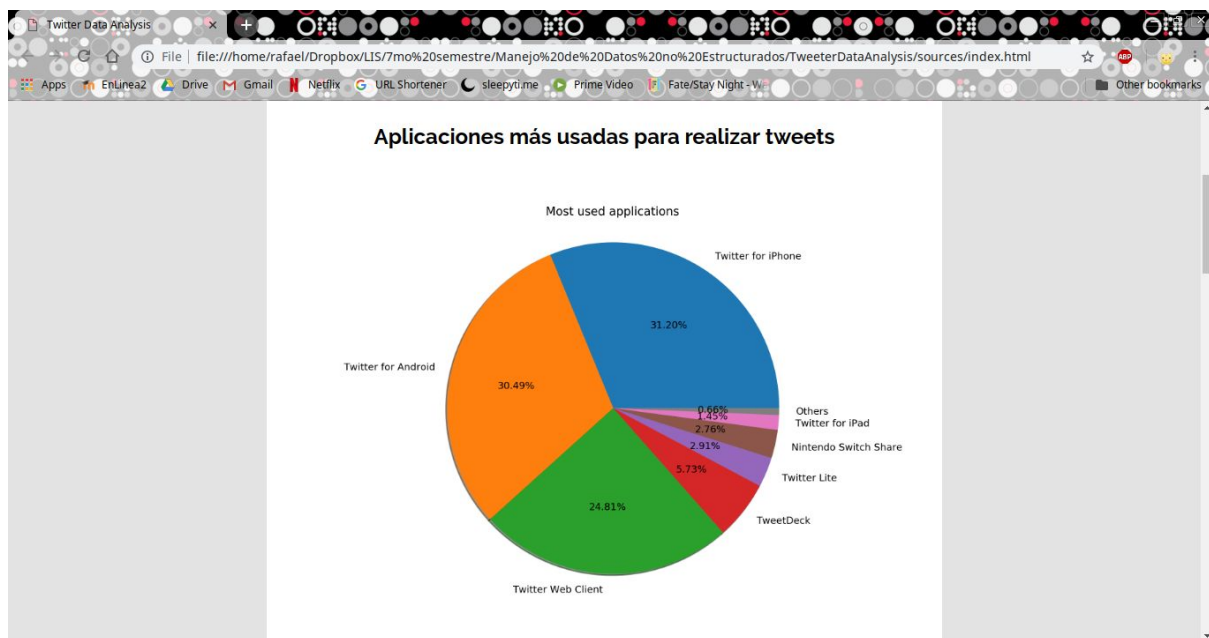
Generando gráfica de idiomas...

main.go:63: openFile: "/home/rafael/Dropbox/LIS/7mo semestre/Manejo de Datos no Estructurados/TweeterDataAnalysis/sources/index.html"
main.go:93: desktop file: /usr/share/applications/google-chrome.desktop

--- 97.97560238838196 seconds ---
rafael@rafael: ~/Dropbox/LIS/7mo semestre/Manejo de Datos no Estructurados/TweeterDataAnalysis$
```

Al terminar la ejecución del script, se abrirá una ventana en tu navegador web por defecto. El navegador cargará una página que contiene las gráficas generadas por el script.





Las imágenes que se generan se ubican en la ruta `/plots/` dentro de la misma carpeta del proyecto.

Si tienes problemas para ejecutar verifica que:

- Tienes una versión de Python 3.6 instalada
- Tienes las bibliotecas necesarias instaladas
- Tienes acceso a internet, y por lo tanto, a la base de datos

- Los datos que proporcionaste para conectarte a la base de datos (dentro del código del script) son correctos.

Revisa el manual técnico para verificar que tienes todo correcto.