



ROBERTO QUINDE

Enunciado:

- Diseñe y desarrolle un sistema recopilador que permita obtener las noticias, Facebook, twitter de los alcaldes dentro de una base de datos orientados a grafos

Para la solución del examen he usado la técnica de webscraping usando el lenguaje Phyton y la herramienta **Selenium**, el código esta dividido en 3 partes y la cuarta parte corresponde al proceso de cargar la información en la base de datos:

1. La primera parte lo que hace es acceder a la pagina de Twitter para realizar la búsqueda “alcalde yunda”, luego de esto avanza mediante el scroll avanza en la página para que se carguen todos los twits relacionados con la búsqueda.

```
In [37]: #WEBSCRAPING TWITTER
from selenium import webdriver
from selenium.webdriver.support.ui import Select
from selenium.webdriver.support.ui import WebDriverWait as EC
import time
import csv
import pandas as pd

website = 'https://www.twitter.com/login'
path = '/webdriver/chrome/chromedriver'

driver = webdriver.Chrome(path)
driver.get(website)

loginUSR = driver.find_element_by_xpath('//*[@id="react-root"]/div/div/div[2]/main/div/div/div[2]/form/div/div[1]/label/div/div[2]/input')
loginKey = driver.find_element_by_xpath('//*[@id="react-root"]/div/div/div[2]/main/div/div/div[2]/form/div/div[2]/label/div/div[2]/input')
loginButon = driver.find_element_by_xpath('//*[@id="react-root"]/div/div/div[2]/main/div/div/div[2]/form/div/div[3]/div')
loginButon.click()

time.sleep(5)
website = 'https://twitter.com/search?q=alcalde%20yunda&src=typed_query'
driver.get(website)
time.sleep(5)

# CARGA TODAS LAS PAGINAS DE LA BUSQUEDA
# Get scroll height
last_height = driver.execute_script("return document.body.scrollHeight")

while True:
    # Scroll down to bottom
    driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")

    # Wait to Load page
    time.sleep(5)

    # Calculate new scroll height and compare with last scroll height
    new_height = driver.execute_script("return document.body.scrollHeight")
    if new_height == last_height:
        break
    last_height = new_height
```



Examen Practico

27/05/2021

2. La segunda parte del código va extrayendo la información de cada twitt buscándolos por su tag: 'article' (**aux = tuit = driver.find_element_by_tag_name('article').text**) luego de extraer los datos elimina el elemento mediante un script (**driver.execute_script("return document.getElementsByTagName('article')[0].remove();")**) y repite el proceso para los siguientes elementos.

```
In [38]: #EXTRAE LOS 1000 PRIMEROS TUIITS
tuits = []
tuits.append(['NOMBRE', 'USUARIO', 'FECHA', 'MENSAJE'])
for x in range(0, 1000):
    aux = tuit = driver.find_element_by_tag_name('article').text
    info = aux.split(sep='\n')

    tuitinfo = {'autor':info[0], 'usuario':info[1], 'fecha':info[3], 'mensaje':info[4]}
    print(tuitinfo)
    driver.execute_script("return document.getElementsByTagName('article')[0].remove();")

    tuits.append([info[0], info[1], info[3], info[4]])
    #time.sleep(0.10)
```

3. La tercera parte lo que hace es exportar los datos a un archivo csv para con este importarlos a NEO4J.

Jupyter ExamenIntercicloRobertoQ Last Checkpoint: hace 18 horas (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [44]: #EXPORTA LOS DATOS A UN ARCHIVO CSV
f = open("tuitsaux.csv", "a", newline="", encoding="utf-8")
writer = csv.writer(f)

for line in tuits:
    writer.writerow([line[0], line[1], line[2], line[3]])

f.close()

df = pd.read_csv("C:/Users/Roberto Quinde/tuitsaux.csv")
print(df)
df.to_csv('C:/examenIA/tuitsfinalfinal3.csv', sep=',', index=False)
```

	NOMBRE	USUARIO	FECHA	\
0	Primicias	@Primicias	13m	
1	Primicias	@Primicias	37m	
2	Fermin Antonio Vaca	@FerminVaca	May 31	
3	Radio Elite 99.7 fm	@Radioelite997fm	Jun 1	
4	danilo Camacho	@daniloCamacho2	May 31	
...	
1411	Lolos Lolos	@Lolos2017	May 29	
1412	Jean Cano	@jean_cano	May 31	
1413	danilo Camacho	@daniloCamacho2	May 31	
1414	Carlos Chamorro Arturo	@cchamorro17019	May 28	
1415	Edgar P Jaramillo S	@jaramilloeps	May 28	

MENSAJE

	MENSAJE
0	El alcalde Jorge Yunda pondrá a prueba su capa...
1	"No vamos a permitir que pocos concejales vuln...
2	Tres concejales correistas podrian decidir la ...
3	Un mes después de que se llamó a juicio al alc...
4	No se puede tener como alcalde a una persona c...
...	...
1411	Esperemos que para entonces el alcalde con gri...
1412	Que el PSC pacte con Correa, malo. ¿Que alguno...
1413	No se puede tener como alcalde a una persona c...
1414	Por el bien de la ciudad de Quito, y sus habit...
1415	La Conision d Mesa del Cabildo de Quito aprobó...

[1416 rows x 4 columns]

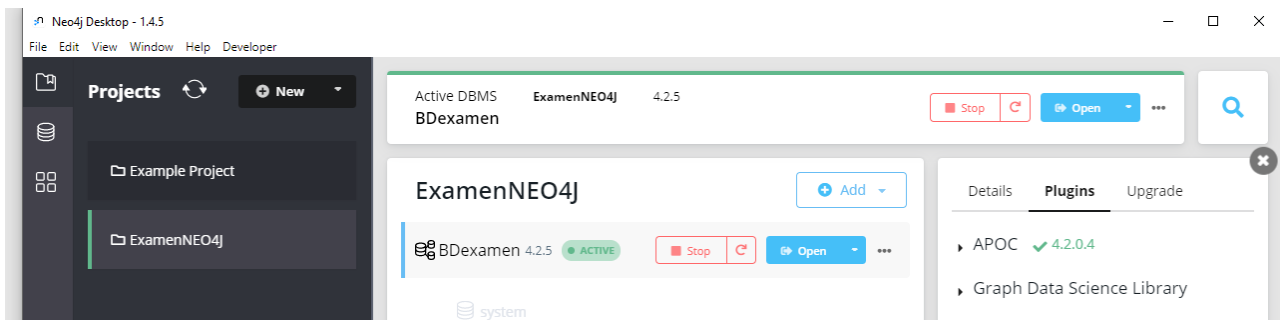
```
In [40]: print(len(tuits))
X = 1
#time.sleep(0.10)
```



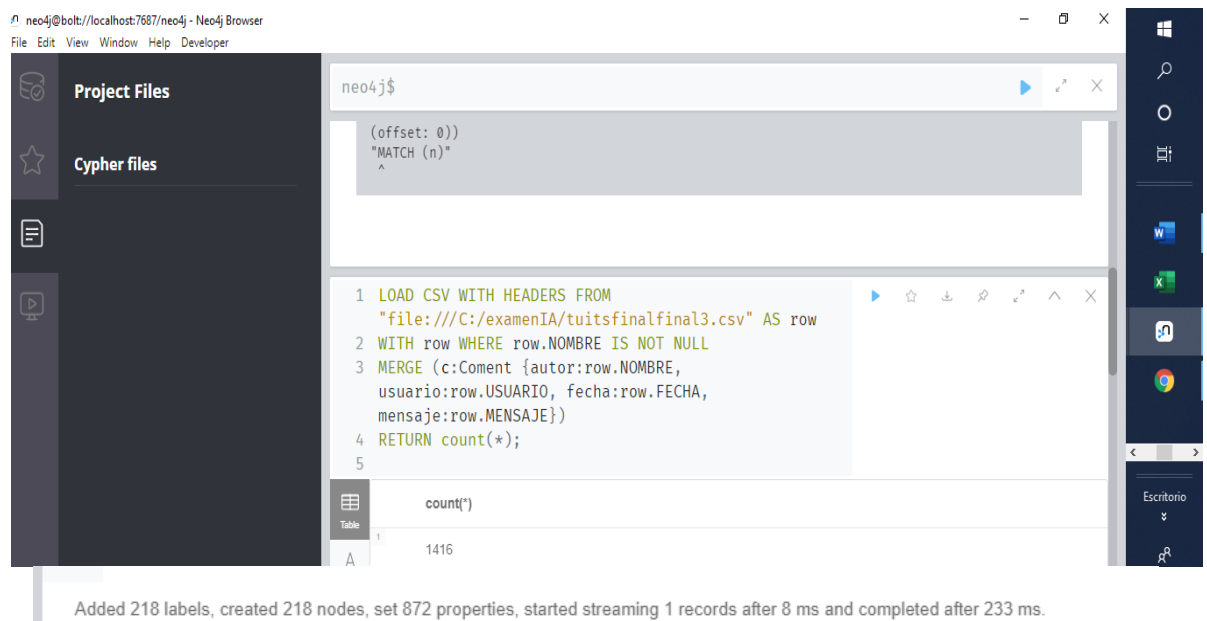
Examen Practico

27/05/2021

4. La parte final consiste en el proceso de cargar los datos en NEO4J, para lo cual procedí de la siguiente manera:
- Generar un proyecto nuevo en NEO4J con el nombre “ExamenNEO4J”
 - Crear una base de datos nueva con el nombre “BDexamen”



- Procedo a abrir la base de datos e importar el archivo csv





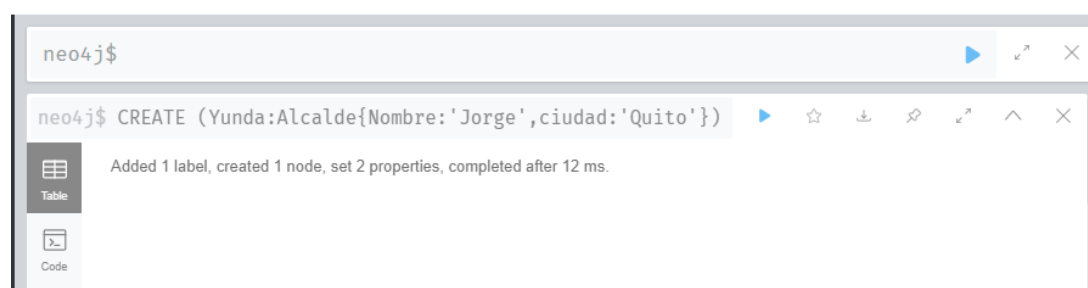
Examen Practico

27/05/2021

- Hago una consulta para verificar que se cargaron los datos



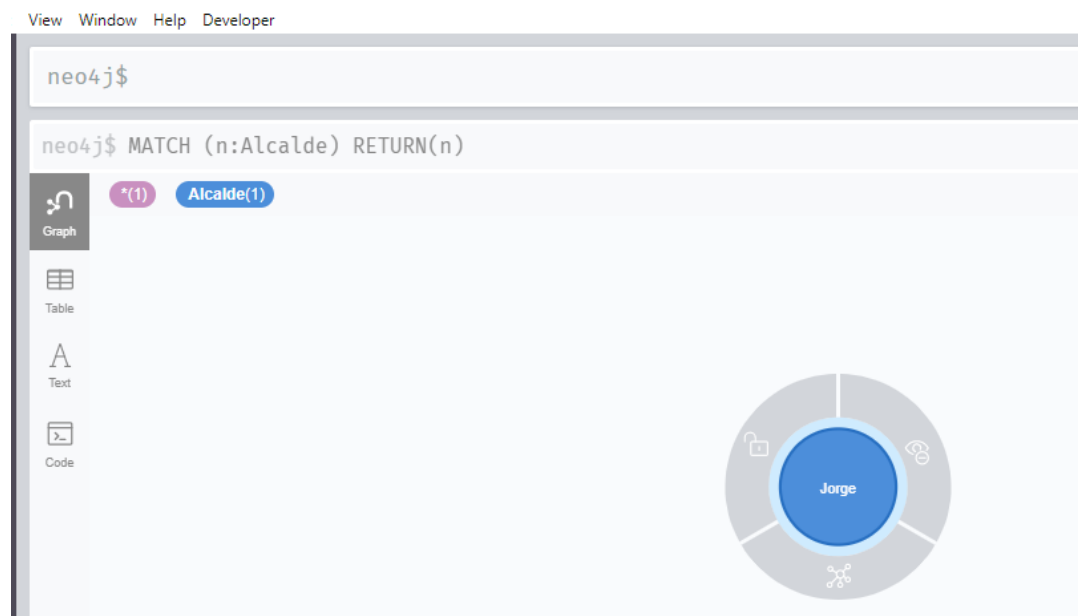
- Creo el nodo Yunda con etiqueta Alcalde, y propiedades nombre y ciudad.



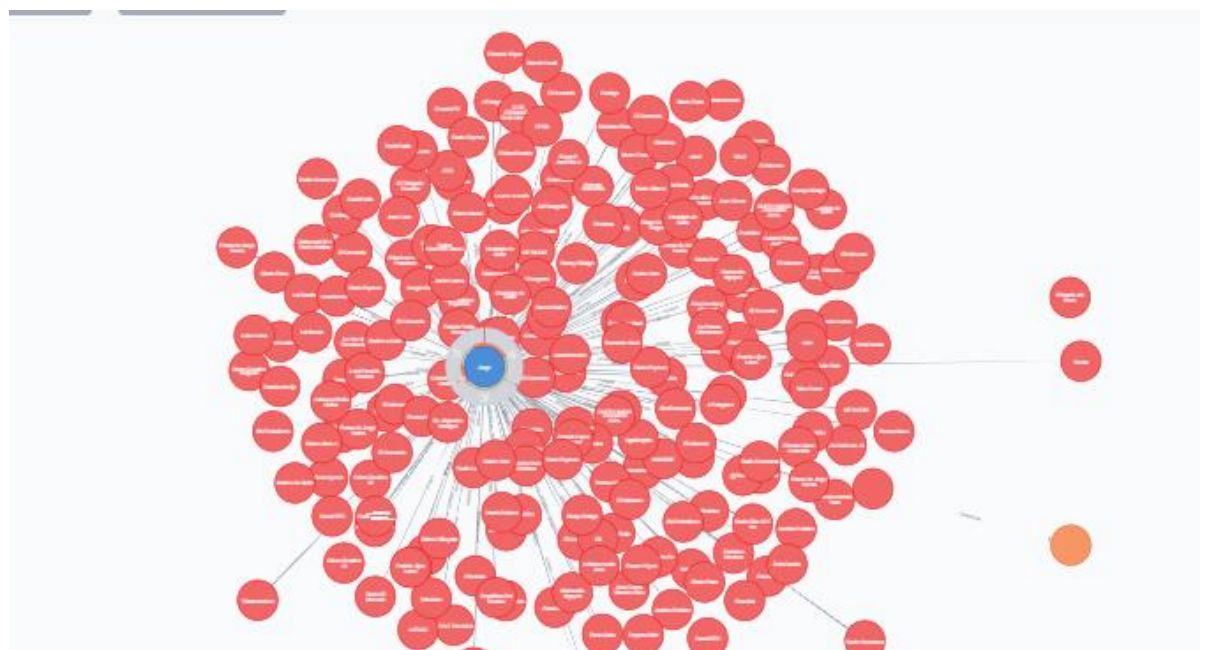


Examen Practico

27/05/2021



- Creo la relación COMENTAN de todos los comentarios con Alcalde.

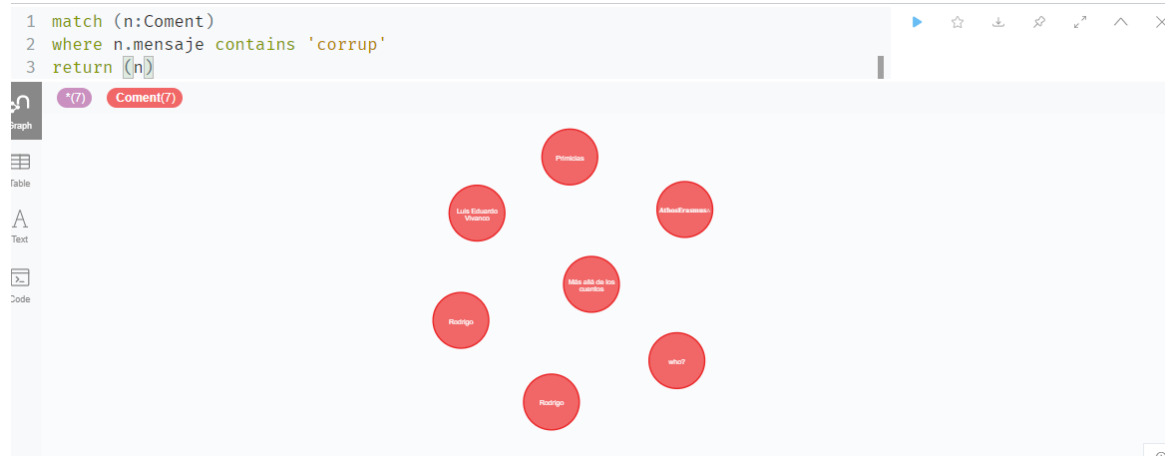




Examen Practico

27/05/2021

- Consulta cuantos mensajes nombran corrupción.



- Consulta twits que contengan la palabra fuera en el mensaje



ANALISIS

En primera instancia podemos apreciar que se crearon 218 nodos después de importar el archivo csv, lo que nos indica que los 1415 Twitts recopilados, todos fueron emitidos por 218 usuarios.

Se puede decir que el 3.2% de los mensajes están relacionados con la palabra corrupción.

Se puede decir que el 0.92% de los mensajes están relacionados con la palabra fuera.



CONCLUSION

Se puede concluir que el alcalde Jorge Yunda ha sido mencionado una gran cantidad de veces los últimos días, se podría decir que en gran parte debido a la situación polémica actual.

RECOMENDACIÓN

Para un análisis con mayor profundidad de los comentarios se podría ahondar en el contenido de los mensajes y plantear nuevas relaciones entre los mismos.