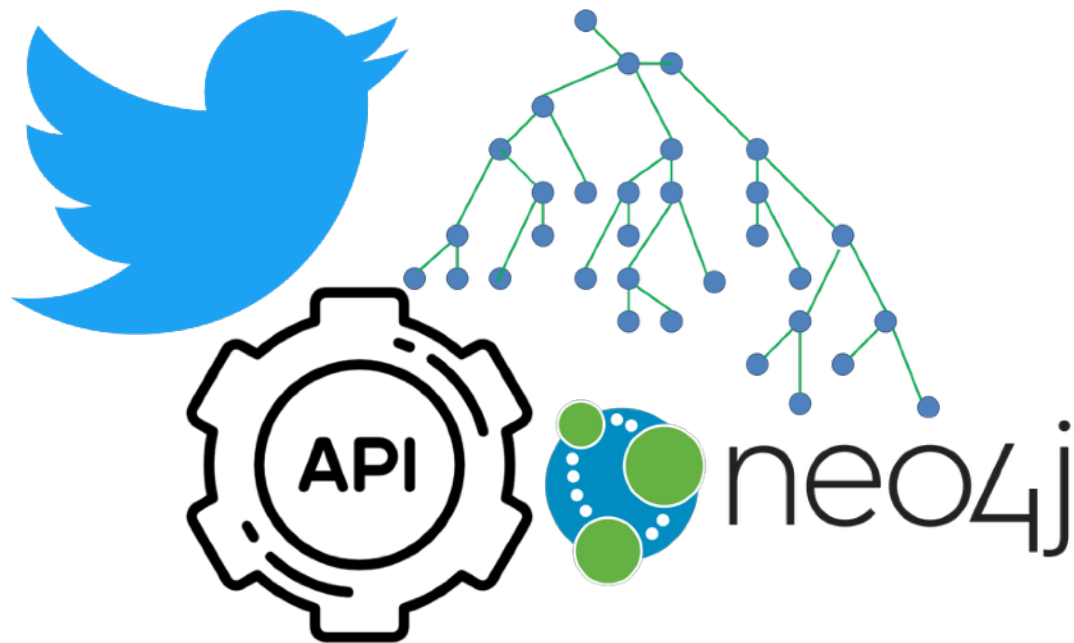


Plataforma para el análisis sociopolítico de comunidades y núcleos de población



1. Introducción	3
2. Adquisición de los datos	4
3. Volcado de los datos	5
4. Análisis de los datos	6
5. Usos	8

1. Introducción

Analizando la actualidad se puede observar cómo un gran porcentaje de la población expresa sus ideales y pensamientos de manera concisa en las diferentes redes sociales (En especial Twitter), lo que genera las convierte en una muy interesante fuente de información a explotar para proyectos de diferentes finalidades.

La mayoría de estas plataformas proporcionan APIs que facilitan la explotación de manera sencilla de este conjunto de datos. Ya que gracias a su flexibilidad son fácilmente adaptables las distintas necesidades de negocio.

Motivación

En la actualidad se observa un gran volumen de exaltación en cuanto a temas políticos, ya que prácticamente la totalidad de los usuarios exponen de manera directa o indirecta su pensamientos, por lo que sería de gran intereses el hecho de almacenar esta información con el objetivo de realizar distintos análisis sobre el conjunto de información.

2. Adquisición de los datos

Para la extracción de los datos se ha usado Python en conjunto a las librerías de Tweepy (facilitar la tarea de conexión con la API de twitter) y Pandas (para la manipulación de los datos extraídos)



Para el proceso de extracción ha sido necesario conectarse a la API de Streaming haciendo uso de varias de las distintas herramientas que proporciona Tweepy aplicando sobre ellas las configuraciones necesarias. Algunas de las modificaciones que se han aplicado serían.

- Palabras clave utilizadas para capturar el contenido tweets de contenido político. (necesario especificar una lista equitativa y representativa para evitar sesgo).
- Periodo de generación de archivos.
- Filtros de idiomas en los que capturar los tweets realizados,

Una vez pasado el periodo especificado para la generación de los distintos archivos se irán generando los distintos archivos como resultados de la extracción. Estos archivos se generaran en formato CSV tanto como XLSX y constarán de la siguiente estructura.

user_name	Name	Text	country	place_name
Nick del usuario (unico)	Nombre del usuario	Texto del tweet	Pais de origen	Ubicación registrada por Twitter como origen

Puede observar el código utilizado para la extracción de datos [aquí](#).

3. Volcado de los datos

Para realizar el volcado de los datos previamente es separar lo usuario por si hay columnas repetidas, aunque haciendo uso de MERGE esto no seria necesario. Pero por mayor limpieza hemos procedido a separarlos ya que se trata de una tarea simple que no requiere mucho tiempo.

Para separar los datos hacemos uso de panda con sentencia básicas como serian drop duplicates y filtrar columnas.

Una vez tenemos los datos preparados para su carga en la base de datos podemos cargarlos en esta haciendo uso de la sentencia LOAD CSV, para ello usaremos las siguiente sentencias.

```
CREATE CONSTRAINT ON (user: User) ASSERT user.user_name IS UNIQUE;  
CREATE CONSTRAINT ON (place: Place) ASSERT place.place_name IS UNIQUE;  
CREATE INDEX ON :Place(name);
```

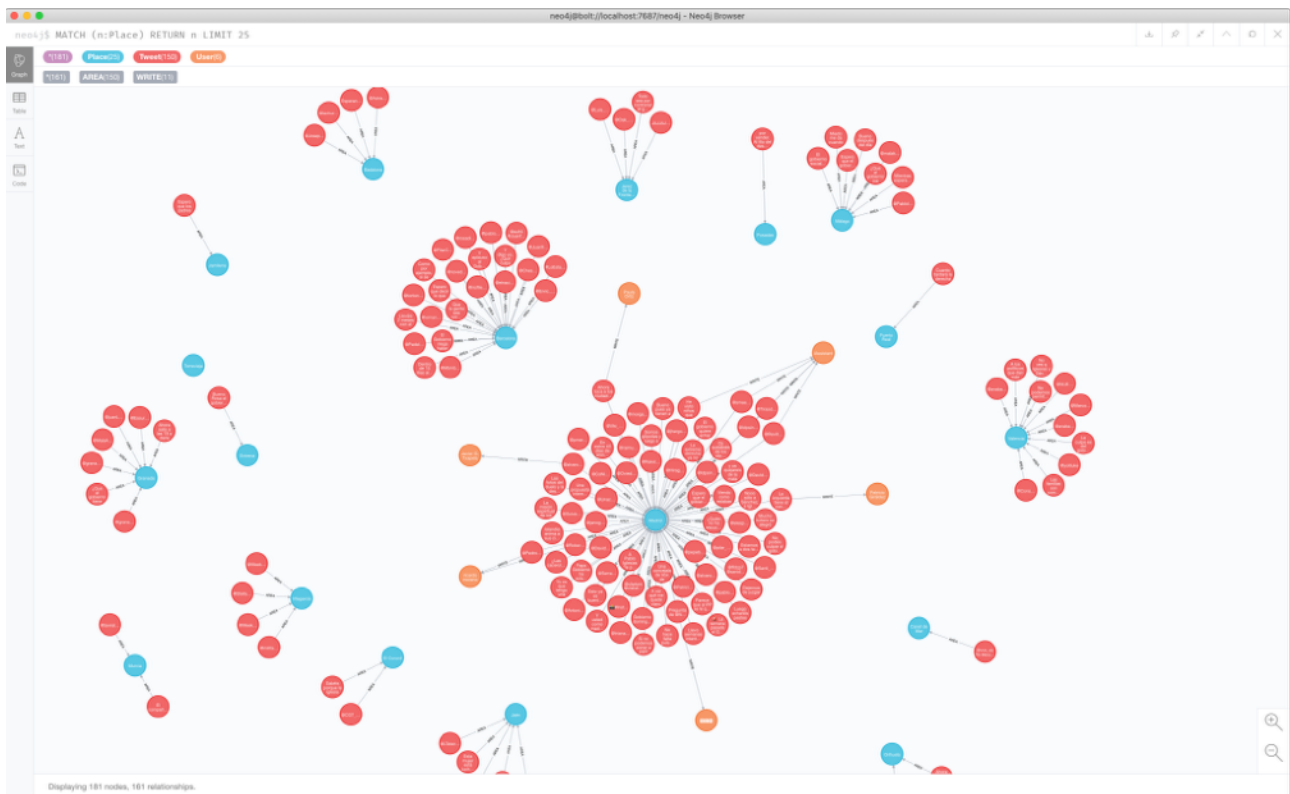
Es necesario crear las constraint de valores únicos en los campos que no se van a poder repetir. Una vez creamos las restricciones alocando lo archivos en el directorio import podemos cargarlos a la base de datos con los siguientes comandos.

```
# Create user  
LOAD CSV WITH HEADERS FROM "file:///users_1.0.csv" AS csvLine  
MERGE (user:User {user_name: csvLine.user_name, name: csvLine.name})  
  
# Ceate Tweets and Place and relation to their user  
LOAD CSV WITH HEADERS FROM "file:///tweets_1.0.csv" as csvLine  
MERGE (place: Place {name: csvLine.place_name, country: csvLine.country})  
CREATE (tweet:Tweet {text: csvLine.text})  
    WITH place, tweet, csvLine  
    MATCH (user: User {user_name: csvLine.user_name})  
    CREATE (user)←[:WRITE]-(tweet)-[:AREA]→(place)
```

4. Análisis de los datos

Una vez cargado el conjunto de datos podemos realizar unas queries generales para observar la distribución de estos.

Podemos observar como se distribuyen los distintos núcleos de población y los tweets se relacionan con estos



Aquí podemos observar los distintos núcleos de población y los tweets pertenecientes a estos. Si accedemos más en detalle a un nodo de ubicación podemos observar todas sus relaciones con tweets y usuarios.

Aquí observamos en mayor detalle todos los tweets extraídos de la comunidad de Madrid y los distintos usuarios a los que pertenecen.

5. Usos

Se podría aplicar distintos métodos de análisis con técnicas de ML cómo sentiment analysis y clustering sobre la plataforma establecida para la generar distintas fuentes de valor. Algunas de ellas serian las siguientes.

1 Campañas electorales personalizadas por ubicación

Se podrá utilizar esta aplicación como base para realizar estudios políticos por núcleo de población. Esto permitiría conocer qué temas se valoran más entre los usuarios permitiendo enfocar campañas con el objetivo de captación con el objetivo de captar nuevos votantes.

2 Detección de núcleos radicales y eventos organizados

En la actualidad debido a la alta inestabilidad política se producen congregaciones en ocasiones ilegales, lo que conllevan altercados públicos que en consecuencia generan desperfectos y perdidas tanto sociales como económicas.

En multiples casos estas agregaciones se convocan a través de redes sociales como twitter ya que se trata de un medio con alta expiación y con poca supervisión ya que tiene un trafico descomunal difícil de controlar. A través de esta base se podrían analizar los datos con el objetivo de detectar previamente este tipo de convocatorias.

3 Evaluación de la repercusión social de distintos políticos y políticas

A través de esta app y distintas técnicas de análisis se podría analizar la repercusión social de distintos políticos. Esto permitiría evaluarlos generando un valor y analizando puntos claves para mejorar su imagen eliminando conductas públicamente criticadas.