

# Minería de Datos: Application Programming Interface

## (Sesión 10B)

Prof. Juan C. Correa, Ph.D.

Colegio de Estudios Superiores de Administración  
Bogotá - Colombia



## Objetivo de Aprendizaje

En la sesión 10A estudiamos, a través de un ejercicio guiado y un taller, la técnica de web scraping. En esta sesión, estudiaremos algunos de los fundamentos de una técnica semejante llamada Application Programming Interface (API). A través de un API, es posible recolectar datos disponibles en redes sociales de manera semi-automática. En esta sesión nos concentraremos en cómo usar la API de Twitter para recolectar datos frescos de los tweets de usuarios de Twitter. En la próxima sesión continuaremos con la idea de saber cómo extraer y guardar esos datos.



# 1

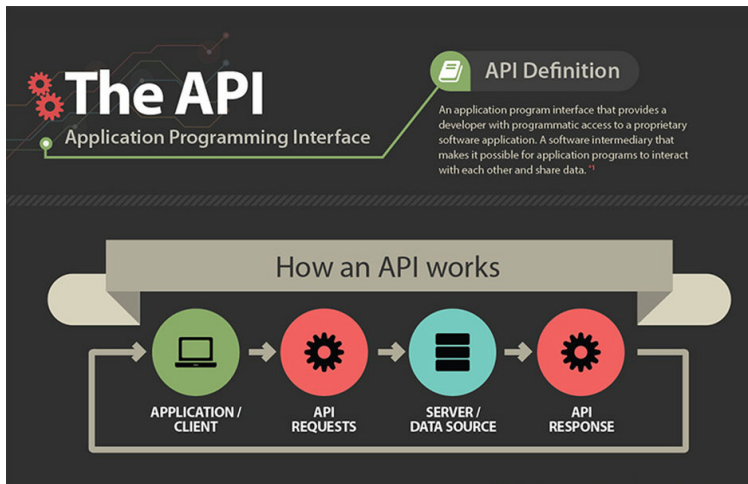
---

## Insumos Preliminares



## Insumo Conceptual:

Servidor Fuente de Datos (Sesión 3B) y Application Programming Interface (Chan, Chung, y Huang, 2019)



# 2

---

## Caso de Estudio



## Ideological Consumerism in Colombian Elections, 2015: Links Between Political Ideology, Twitter Activity, and Electoral Results

Juan C. Correa, PhD<sup>1</sup> and Jorge E. Camargo, PhD<sup>2</sup>

### Abstract

Propagation of political ideologies in social networks has shown a substantial impact on voting behavior. Both the contents of the messages (the ideology) and the politicians' influence on their online audiences (their followers) have been associated with such an impact. In this study we evaluate which of these factors exerted a major role in deciding electoral results of the 2015 Colombian regional elections by evaluating the linguistic similarity of political ideologies and their influence on the Twitter sphere. The electoral results proved to be strongly associated with tweets and retweets and not with the linguistic content of their ideologies or politicians' followers in Twitter. Finally, suggestions for new ways to analyze electoral processes are discussed.

**Keywords:** Twitter, Colombian elections, ideological consumerism

(Correa y Camargo, 2017)





El **Consumismo Ideológico** ocurre cuando las decisiones individuales de apoyar o rechazar alguna acción (comprar, vender) son guiadas por ideales basados en principios éticos o morales.



¿Existe consumismo ideológico en las elecciones regionales de Colombia 2015?  
y de ser así ¿Cómo se evidencia tal fenómeno?





52 candidatos a gobernadores; 52 colecciones de tweets; 69.202 tweets; número promedio de tweets por candidato; número promedio de retweets por tweet por candidato; ideología política; similaridad lingüística entre ideologías; resultados electorales (número de votos recibidos).

The screenshot shows the Colombia.com website interface. At the top, there's a navigation bar with links like 'Inicio', 'Iniciar Sesión', 'Registrarse', and a search bar. Below this is a banner for 'Elecciones' with icons for Fútbol, Deportes, Actualidad, Cine, and Entretenimiento. A central banner for 'betsson' promotes a football betting event: '¿QUIEN ANOTARA EL PROXIMO GOL? APUESTAS EN VIVO'. Below this is a large banner for '2015 ELECCIONES REGIONALES' with a 'VOTE' button. The main content area shows filters for 'GOBERNADOR', 'AMAZONAS', and 'Municipios'. It displays election results for the 'GOBERNADOR' position, dated October 25, 2015. A table shows the progress of the election, including the number of installed tables (134), informed tables (134, 100.00%), potential voters (44,201), and votes marked (1,177, 4.46%). A sidebar on the right features an advertisement for LATAM Airlines with the text 'NUEVOS VUELOS BOGOTÁ/LETICIA/BOGOTÁ'.

GOBERNADOR	
Boletín: 25	
Avance:	27
Fecha del avance:	25 de Octubre de 2015
Hora del avance:	06:20:30 pm
Departamento:	AMAZONAS

Mesas instaladas:	134
Mesas informadas:	134 100.00%
Potencial sufragantes:	44,201
Votos no marcados:	1,177 4.46%
Votos nulos:	533 2.02%

En esta **matriz término-documento** se captan las relaciones entre las palabras (filas) que aparecen en cada mensaje tuiteado por cada candidato (columnas), cuya ideología política se expresa con un color específico.

	Candidato 1	Candidato 2	Candidato 3	...
mejor	1	0	1	...
educación	1	1	0	...
para	1	0	1	...
Colombia	1	1	1	...
no	0	1	0	...
merece	0	1	0	...
mala	0	1	0	...
empleo	0	0	1	...
...	...	...	...	...



Si consideramos que todos los tuits de un candidato conforman un documento que se representa como un vector de 1 y 0, entonces su representación formal sería:

$$x_i = (x_{1,i}, x_{2,i}, x_{3,i}, \dots, x_{t,i}) \quad (1)$$

y

$$x_{t,i} = tft \cdot \frac{\log|D|}{|t \in i|} \quad (2)$$

Donde,  $x$  es un tuit,  $tft$  es la frecuencia de aparición de la palabra  $t$  en el documento  $i$ ,  $|D|$  es el número de tuits de un candidato y  $\log|D|/|t \in i|$  es el inverso de la frecuencia de los documentos que contienen la palabra  $t$ .

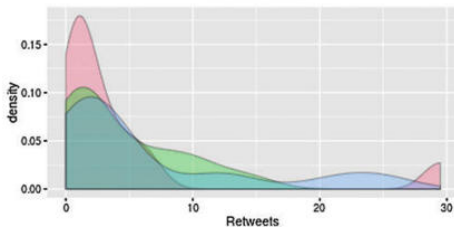
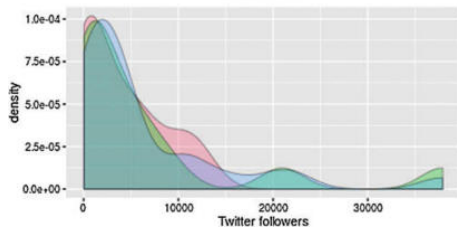
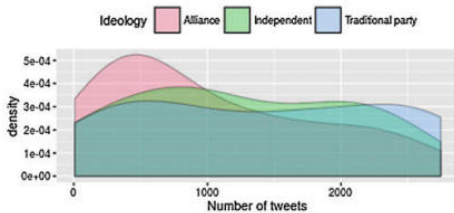
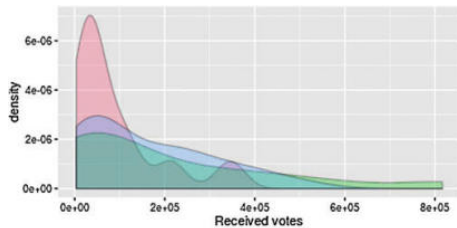


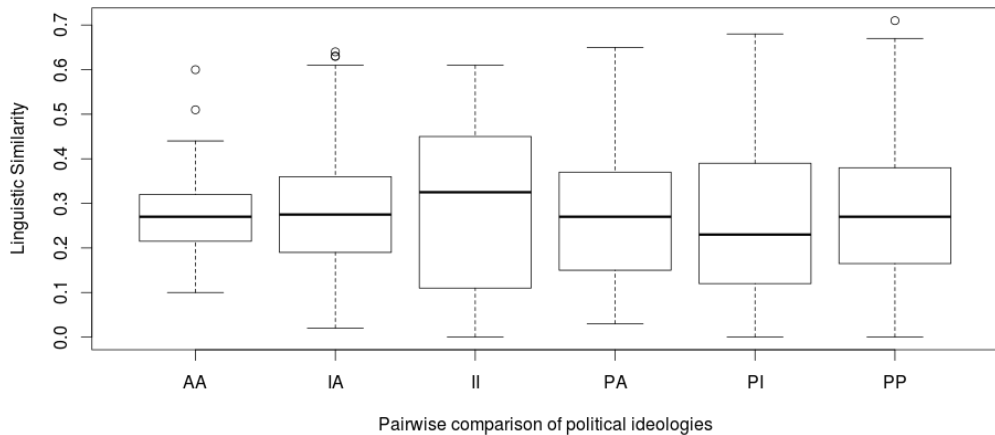
Asumiendo que  $x_{t,i}$  se puede pintar en un espacio cartesiano, entonces la **semejanza lingüística** entre cualquier par de vectores se puede estimar como:

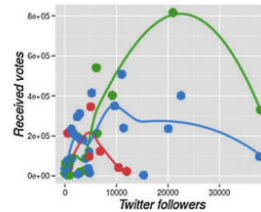
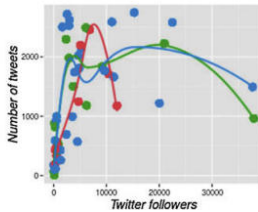
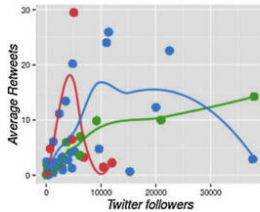
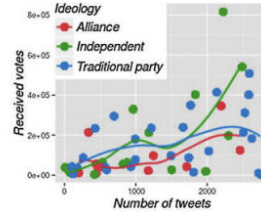
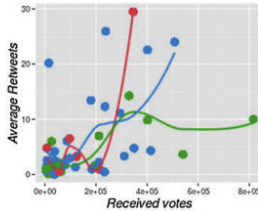
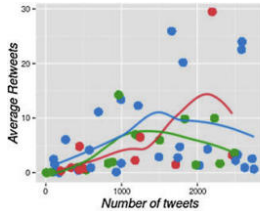
$$d(p, q) = \sqrt{(p - q)^2} \quad (3)$$

donde,  $d$  representa a la distancia Euclidea entre el conjunto de tuits publicados por el candidato  $p$  y el conjunto de tuits del candidato  $q$









# 3

---

## Pasos para usar una API





```
In [ ]: 1 pip install python-twitter
```

```
In [ ]: 1 pip install pytest
```

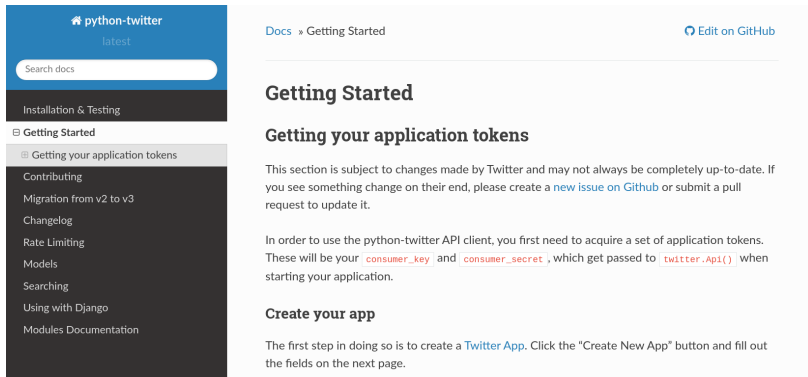
```
In [ ]: 1 pip install pytest-cov
```

```
In [1]: 1 import twitter
2 api = twitter.Api(consumer_key='[REDACTED]',
3 consumer_secret='60uHdUePGu8AcNBgy68uUqKUE4ymsb9Ws7ybo83WARAEnpVbrJ',
4 access_token_key='[REDACTED]',
5 access_token_secret='BJdEe21hxvkGvestC1wI3AAPwF17L8LhQORBsdztlWsCU')
```

```
In [2]: 1 statuses = api.GetUserTimeline(106483037)
2 print([s.text for s in statuses])
```

["RT @C3UNAM: This talk will show how machine learning, artificial intelligence, and natural language processing can help to trace people's b...", "RT @isidroaguiro: Ya se que debemos utilizar la cortesía académica y ser política mente correctos, pero ¿cómo actuar con un interlocutor ab...", '@cosnet bifi @BIFI Instituto @unizar @UccUnizar @ARA ID ES @GobAragon Congrats!', 'https://t.co/7jHt4A8JKJ', 'Wow. This interesting article by @gcabanac and colleague s presents a brand new perspective of scientific writing a... https://t.co/Y26MRzZDPm', 'RT @MyScienceWork: An interesting read from Interdisciplinary Scientist @jcorrean \nDoes limited access to publications really hinder sci...', 'RT @profjamesevans: The amazing Johan Chu and I published a piece out this week at PNAS (https://t.co/McJwKX BxvG) showing how the growing s...', 'RT @PLOS: How Open is your research? \n\nTake our short quiz to find out your s core and get tips on how you can develop your #OpenScience skill...', 'Interesante lectura https://t.co/y5p0gpinm8', 'https://t.co/Y9IHQIt6hB', 'In our last paper, we revealed that the trust-sales relationship is not consistent at all in several Latin American... https://t.co/0AyF345ah5', 'The Consistency of Trust-Sales Relationship in Latin Ame rican E-Commerce https://t.co/0lhBHYrnok', 'Our last paper presents new evidence on the relationship between Econo mic Complexity, Economic Growth, and CO2 Emis... https://t.co/n4MBLCayQL', 'RT @hadleywickham: Major updates to all tidyverse cheat thanks to our intern @avperny. Download them now! #rstats https://t.co/9Nl4tX9F3d', 'RT @Bibliomet raf: The Invisible Citation Commons. \n\nhttps://t.co/yiRPM0GdNj', 'RT @Piniisima: La evidencia más sólida de como trasciende un tuit de una red social poca representativa a una red nacional significativa; L...', 'RT @Marek Kwiek: A reliable bibliometric picture of the post-Soviet world in science, 15 countries. Interesting! \n\nhttps://t.co/5m RdwNCdDN', 'El próximo 21 de Septiembre de 2021, estaré encantado de compartir algunas ideas en el Primer Simposio Internacional... https://t.co/SURHkveoTS', 'RT @RibeiroHV: Our work "Association between productivity and journal im pact across disciplines and career age" has just been accepted for...', 'Our paper on the sci-hub effect has received a recent citation. This time from physical review research... https://t.co/3wkfQLTyGz']

# Paso 1: Crear una App de Twitter leyendo el manual



The screenshot shows the 'python-twitter' documentation page on ReadTheDocs. The left sidebar contains a navigation menu with the following items: 'python-twitter latest', 'Search docs', 'Installation & Testing', 'Getting Started' (selected), 'Getting your application tokens' (selected), 'Contributing', 'Migration from v2 to v3', 'Changelog', 'Rate Limiting', 'Models', 'Searching', 'Using with Django', and 'Modules Documentation'. The main content area is titled 'Getting Started' and 'Getting your application tokens'. It includes a link to 'Edit on GitHub', a paragraph about the section being subject to changes, and a paragraph explaining the need for application tokens, with code snippets for `consumer_key`, `consumer_secret`, and `twitter.Api()`. A section titled 'Create your app' follows, with instructions on how to create a new app.

python-twitter  
latest

Search docs

Installation & Testing

Getting Started

Getting your application tokens

Contributing

Migration from v2 to v3

Changelog

Rate Limiting

Models

Searching

Using with Django

Modules Documentation

Docs » Getting Started [Edit on GitHub](#)

## Getting Started

### Getting your application tokens

This section is subject to changes made by Twitter and may not always be completely up-to-date. If you see something change on their end, please create a [new issue on Github](#) or submit a pull request to update it.

In order to use the python-twitter API client, you first need to acquire a set of application tokens. These will be your `consumer_key` and `consumer_secret`, which get passed to `twitter.Api()` when starting your application.

#### Create your app

The first step in doing so is to create a [Twitter App](#). Click the "Create New App" button and fill out the fields on the next page.

[https://python-twitter.readthedocs.io/en/latest/getting\\_started.html](https://python-twitter.readthedocs.io/en/latest/getting_started.html)



Paso 1 (opción alternativa): Observe este video

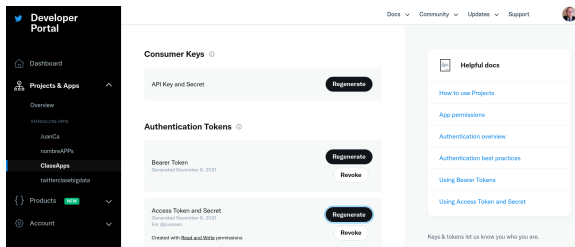


<https://youtu.be/XQ6X5DYcdzA>

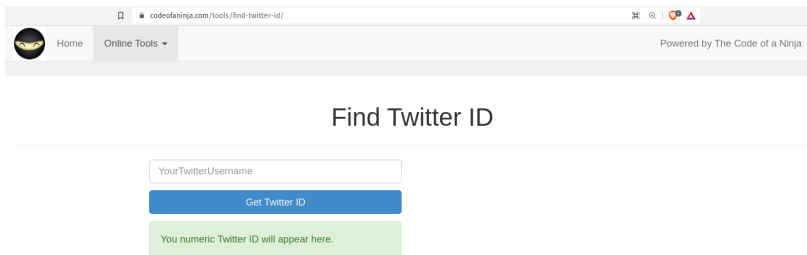


## Paso 2: Tome nota de las credenciales de su API:

- 1 **consumer\_key**
- 2 **consumer\_secret**
- 3 **access\_token\_key**
- 4 **access\_token\_secret**



## Paso 3: Consiga el ID numérico del usuario de Twitter cuyos datos quiere extraer.



The screenshot shows a web browser window with the address bar displaying 'codeofaninja.com/tools/find-twitter-id/'. The page has a header with a ninja logo, 'Home', 'Online Tools' (with a dropdown arrow), and 'Powered by The Code of a Ninja'. The main heading is 'Find Twitter ID'. Below it is a text input field labeled 'YourTwitterUsername', a blue button labeled 'Get Twitter ID', and a green box containing the text 'You numeric Twitter ID will appear here.'

<https://codeofaninja.com/tools/find-twitter-id/>



## Paso 4: Tome nota del ID numérico y úselo en la sintaxis

Get Twitter ID

Username: **@jcorrean**

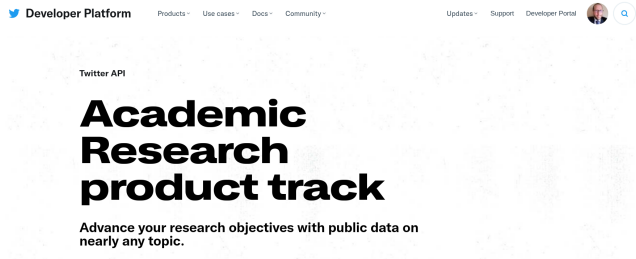
Twitter Numeric ID: **106483037**

Member since: **January 19 2010, 18:16:44 PM**

<https://codeofaninja.com/tools/find-twitter-id/>



## Sugerencias para estudio adicional



<https://developer.twitter.com/en/products/twitter-api/academic-research>



- Chan, J., Chung, R., y Huang, J. (2019). *Python API Development Fundamentals*. Birmingham, UK: Packt Publishing.
- Correa, J. C., y Camargo, J. E. (2017). Ideological consumerism in colombian elections, 2015: Links between political ideology, twitter activity, and electoral results. *Cyberpsychology, Behavior, and Social Networking*, 20(1), 37–43. doi: 10.1089/cyber.2016.0402





