

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. Para fines ilustrativos, en esta sesión nos concentraremos en cómo usar la API de Twitter para recolectar datos frescos de los tweets de personajes públicos.



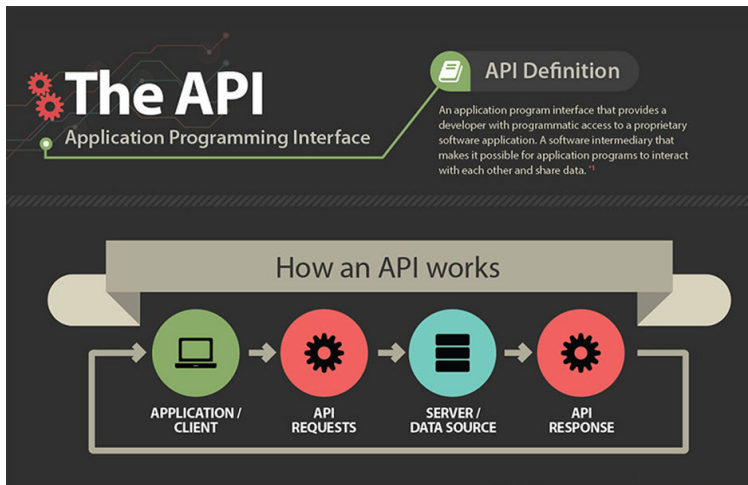
1

Insumos Preliminares



Insumo Conceptual:

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,¹ and Jorge E. Camargo, PhD²

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 features a betsson advertisement with the text '¿QUIEN ANOTARA EL PROXIMO GOL? APUESTAS EN VIVO' and a 'LEER MÁS' button. Below the banner is a section for '2015 ELECCIONES REGIONALES' with a navigation menu including 'Inicio', 'Noticias', 'Resultados', 'Programas de Gobierno', 'Cifras', '¿Cómo votar?', 'Glosario', 'Encuestas', and 'Multimedia'. The main content area shows the results for 'GOBERNADOR' in 'AMAZONAS'. It includes a table with election details and a bar chart showing the percentage of votes for each candidate.

Colombia.com | Elecciones 2015 | Resultados Electorales / GOBERNADOR

GOBERNADOR AMAZONAS -- Municipios --

GOBERNADOR
ELECCIONES REGIONALES - OCTUBRE 25 DE 2015
*Datos suministrados por la Registraduría Nacional del Estado Civil de Colombia

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

Municipios	
Mesas instaladas:	134
Mesas informadas:	134 100.00%
Potencial sufragantes:	44.201
Votos no marcados:	1.177 4.46%
Votos nulos:	533 2.02%

NUEVOS VUELOS BOGOTÁ/LETICIA/BOGOTÁ*
A PARTIR DEL 1 DE OCTUBRE 11 VUELOS SEMANALES

LATAM AIRLINES

En esta matrix 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 .

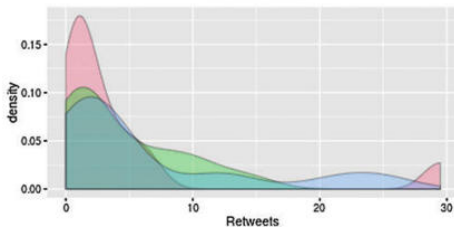
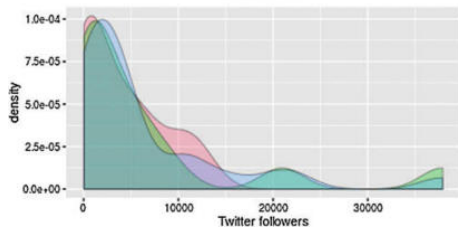
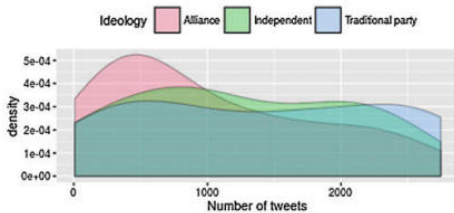
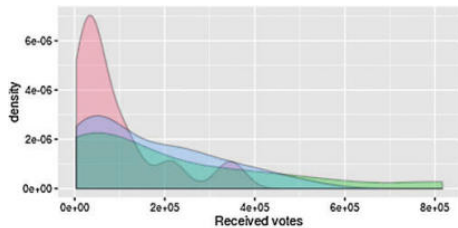


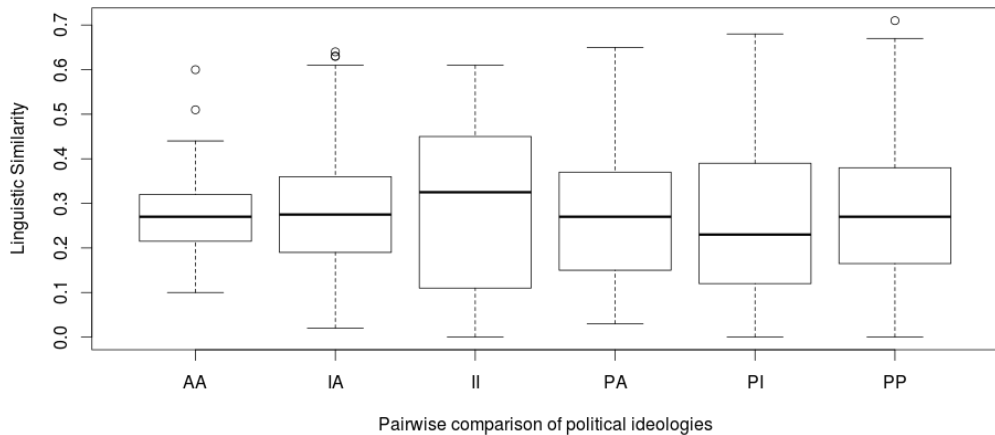
Asumiendo que $x_{t,i}$ se puede pintar en un espacio cartesiano, entonces la **semejanza lingüística** entre cualquier par de vectores se puede estimar como:

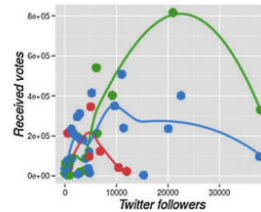
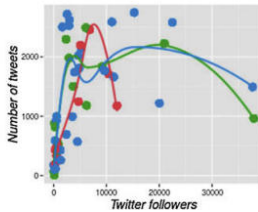
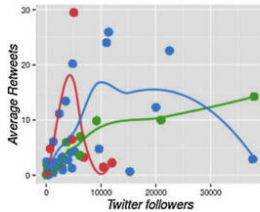
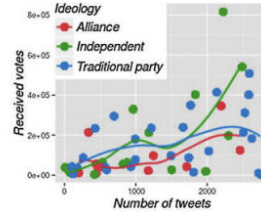
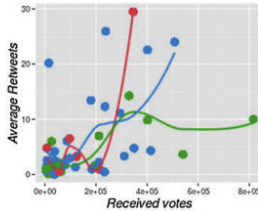
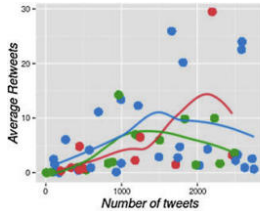
$$d(p, q) = \sqrt{(p - q)^2} \quad (3)$$

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



- 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



