

Nube dinámica de palabras basada en SenticNet para contexto político mexicano

K.M. Ramírez Vergara ^{1*}, L.A. Banda Peláez ², A. López Chau ³, R. Rojas Hernández ⁴, J. Bautista López ⁵

¹Universidad Autónoma del Estado de México, CU UAEM Zumpango, Zumpango, México
kramirezv003@alumno.uaemex.mx, ibandap001@alumno.uaemex.mx, alchau@uaemex.mx,
rojashe@uaemex.mx, jbautistal@uaemex.mx

Área de participación: Sistemas Computacionales

Resumen

Una de las primeras etapas en el proceso de análisis de sentimientos, consiste en un estudio exploratorio de los documentos. La nube de palabras es una de las herramientas más usadas para la visualización resumida de estos datos. Una de las principales desventajas de las nubes de palabras comunes, es que no permiten visualizar los cambios en la polaridad de las opiniones durante el tiempo; tampoco consideran teorías emocionales para una mejor comprensión del fenómeno a estudiar. En este artículo se propone una modificación a las implementaciones actuales de nubes de palabras, con el fin de obtener una representación gráfica de la dinámica de la evolución de la frecuencia y polaridad de palabras en documentos. Estas polaridades se calculan con SenticNet. A este modelo se le denominó nube dinámica de palabras. Como caso de estudio, se aplicó al contexto político mexicano actual.

Palabras clave: Análisis de sentimientos, SenticNet, nube de palabras.

Abstract

One of the first steps in the sentiment analysis process consists of an exploratory study of the documents. The word cloud is one of the most used tools for the summarized visualization of these data. One of the main disadvantages of common word clouds is that they do not allow to visualize changes in the polarity of opinions over time; neither do they consider emotional theories for a better understanding of the phenomenon to be studied. In this paper we propose a modification to the current implementations of word clouds, to obtain a graphical representation of the dynamics of the evolution of the frequency and polarity of words in documents. These polarities are calculated with SenticNet. This model was called dynamic word cloud. As a case study, it was applied to the current Mexican political context.

Key words: Análisis de sentimientos, SenticNet, nube de palabras.

Introducción

El progreso tecnológico ha influido en la forma de pensar, interactuar y relacionar a la sociedad, ya que permiten expresar las inquietudes y las defensas de los intereses sociales. Gracias a la tecnología, se está modificando la manera en cómo se desarrolla la política y la democracia, del mismo modo, la comunicación política en redes sociales se ha convertido en un proceso dinámico, en el cual la información presentada por el gobierno hacia los ciudadanos tiene una respuesta inmediata, generando un intercambio de opiniones a favor o en contra. [6]

Las redes sociales representan una de las herramientas más utilizadas por el gobierno y por ciudadanos, permitiendo la libertad de expresión en el ámbito político, fomenta la participación al diálogo y se obtiene un análisis de las reacciones, comentarios u opiniones, con ello su compromiso político se ve mayormente obligado [1]. Las redes sociales más utilizadas actualmente son Facebook, Twitter, YouTube e Instagram; por lo que, durante un movimiento político, los ciudadanos suelen realizar comentarios en estas redes sociales para difundir su opinión, solidaridad, queja o solicitud, generando así, una gran cantidad de datos e información respecto al movimiento político. [2]

La cantidad de información generada por un evento o circunstancia en las redes sociales ha sido de interés a la comunidad científica para la extracción de información relevante y la toma de decisiones de las figuras políticas, empresas y organizaciones gubernamentales [1]. Además, la importancia del estudio de las emociones y

sentimientos juega un papel importante en comprender los fenómenos sociales y eventos políticos, donde algunas técnicas de la inteligencia artificial pueden ayudar [3].

El análisis de sentimientos es una técnica que implica el procesamiento del lenguaje natural, la aplicación de técnicas de minería de datos [4] y aprendizaje automático; este tipo de análisis consiste en extraer características a partir de textos, audios y/o videos; que permitan clasificar los datos en categorías como “positivo”, “negativo”, o “neutral”. En los últimos años se ha trabajado bastante en el análisis de sentimientos con datos recolectados en Twitter, lo que ha permitido el desarrollo de investigaciones sobre temas internacionales y políticos que tienen un impacto en la sociedad actual [5]. Existen dos principales enfoques de solución para el análisis de sentimientos, el primer enfoque está basado en recursos léxicos, ya que se requiere de un diccionario de palabras asociadas con un sentimiento o tema, este enfoque se conoce comúnmente como análisis de sentimientos basado en lexicón. El segundo enfoque es basado en el aprendizaje supervisado, es decir, se requiere de un conjunto de datos previamente etiquetados [6]

La técnica de análisis de sentimientos se utiliza para examinar los sentimientos de un usuario en base al estudio de los textos. Para analizar textos se utiliza la lingüística, teorías emocionales, métodos estadísticos y el aprendizaje automático, con la finalidad de que las máquinas encuentren patrones en el lenguaje humano y puedan entender el significado del texto.[7] Los métodos de análisis de sentimientos utilizan lexicones, un lexicon se puede definir como un conjunto de palabras que tiene como objetivo obtener de cada elemento una etiqueta que indica el grado de orientación semántica o polaridad, ya sea positiva o negativa [6]. Las palabras de un lexicon pueden ser divididas en dos categorías, tipo base y tipo comparativo. Las de tipo comparativo se utilizan para aquellas opiniones que son comparativas y superlativas, por ejemplo, las palabras “más que”, “a diferencia de”, “mejor”, “peor”, “vencer”, etc. Mientras que las palabras de tipo base son clasificadas mediante una polaridad. Uno de los recursos léxicos que se ocupan para obtener la base de la polaridad de las palabras es SenticNet. Este recurso léxico, semántico y afectivo de acceso público para el análisis de opiniones y sentimientos a nivel de concepto, es capaz de asociar polaridad e información afectiva también a conceptos complejos. SenticNet asigna una puntuación de sentimiento en un rango entre -1 y 1, para 14,000 conceptos de sentido común. Este marco de trabajo se construye mediante la computación semántica, un paradigma que explota tanto las técnicas de inteligencia artificial, como las técnicas de la web semántica para procesar las opiniones en lenguaje natural [8]

Una de las primeras etapas en el proceso de análisis de sentimientos, consiste en un estudio exploratorio de los documentos, para lo cual la nube de palabras es una de las herramientas más comúnmente usadas. En una nube de palabras, la información suele presentarse a modo de figura abstracta, en la que las palabras con mayor frecuencia son representadas de un mayor tamaño.[9] La principal limitación de estas nubes es que no permiten visualizar los cambios en la polaridad de las opiniones durante el tiempo; tampoco consideran teorías emocionales para una mejor comprensión del fenómeno o evento a estudiar.

En este artículo, se presenta el concepto, implementación y aplicación una nube de palabras para visualizar de forma dinámica (respecto al tiempo), la frecuencia de las palabras en el ámbito político, considerando las polaridades de acuerdo con el modelo sentimental de SenticNet.

El resto del artículo está estructurado como se indica a continuación. En la primera sección se introducen las definiciones principales sobre la nube dinámica de palabras. En la segunda sección se presentan los materiales y métodos empleados para recolectar y procesar datos, así como los cálculos para la creación de una nube dinámica de palabras. En la tercera sección se muestra la aplicación del enfoque propuesto para publicaciones en Twitter sobre el contexto político mexicano. En la sección final se describe las conclusiones y futuras investigaciones.

Consideraciones

Se presentan las siguientes definiciones relacionadas con el desarrollo de este trabajo:

- **C**: Conjunto de documentos sobre algún tema (corpus).
- **S**: Subconjunto de documentos que pertenecen al corpus.
- **S'**: Conjunto de palabras perteneciente a **S**. Este conjunto se obtiene de pre-procesar los documentos existentes en **S**.

- **Nube de palabras $N(C')$** : Es una representación gráfica de la frecuencia de uso de cada palabra en un documento d . El tamaño de una palabra es proporcional al número de veces que se repite la misma en el conjunto C'
- **Documento fechado d_t** : Es un documento (texto) que ha sido publicado en cierto instante t . En este tipo de documentos, tanto el tiempo (fecha y hora) de publicación como el contenido, son características muy importantes. Las publicaciones en redes sociales (como Facebook) y en servicios de microblogging (como Twitter) son ejemplos de este tipo de documentos.
- **Nube de palabras de documentos fechados $N_{t_1}^{t_2}(S')$** : Es una nube de palabras construida a partir de un conjunto de documentos fechados entre un tiempo inicial t_1 y un tiempo final t_2 .
Basado en las anteriores definiciones, se presentan ahora las siguientes:
- **Nube dinámica de palabras $ND_{ta}^{tb}(C') = \{N_{ta}^{ta+1}(S'), N_{ta+1}^{ta+2}(S'), \dots, N_{tb-1}^{tb}(S')\}$** : Es una sucesión de nubes de palabras de documentos fechados. Esta nube permite visualizar la evolución temporal de las frecuencias de uso de las palabras en documentos fechados.
- **Nube dinámica de palabras basada en SenticNet $ND_{SenticNet}^{tb}(S')$** : Es una nube dinámica de palabras en la cual las frecuencias calculadas son ponderadas con las correspondientes puntuaciones de polaridad.

Metodología

Materiales

Se utilizó el API Rest que proporciona Twitter cuando se genera una cuenta como desarrollador, para permitir el acceso al núcleo de datos y acceder a los tweets recientes. Así mismo, se utilizó la biblioteca Tweepy que contiene la interfaz Cursor para iterar a través de distintos tipos de objetos (usualmente representados como objetos JSON). [4] Para el desarrollo del artículo, se utilizó SenticNet, con el objetivo de inferir la polaridad de las opiniones. SenticNet consta de un extractor de conceptos y de una biblioteca con hasta 14,000 conceptos clasificados según su positividad o negatividad, el resultado es un número flotante entre -1 y +1 (donde -1 es negatividad extrema y +1 es positividad extrema) [9]

A partir de la lista de palabras clave: ['amlo', 'ElPeorPresidenteDeLaHistoria', 'LopezBasuraPresidencial', 'obrador', 'FuerzaPresidente', '4taTransformación', 'AMLOVE', 'elcacasobrador', '4T', 'Nacional', 'AmloElFracasoPresidencial', 'ElPeorGobiernoDeLaHistoria'] se han descargado 137,801 tweets del 28/06/2022 al 11/07/2022. Finalmente, en la recopilación de los datos, se obtuvo la información presentada en la Tabla 1.

Tabla 1. Número de Tweets recopilados por cada palabra clave. Fuente: Elaboración propia.

Palabra clave	Tweets recolectados
amlo	41,910
ElPeorPresidenteDeLaHistoria	4,757
LopezBasuraPresidencial	1,230
obrador	14,355
FuerzaPresidente	12
4taTransformación	147
AMLOVE	129
elcacasobrador	847
4T	39,504
Nacional	26,262
AmloElFracasoPresidencial	3,891
ElPeorGobiernoDeLaHistoria	4,757
TOTAL	137,801

Métodos

La Figura 1 muestra un resumen gráfico de la metodología empleada, misma que a continuación se explica detalladamente.

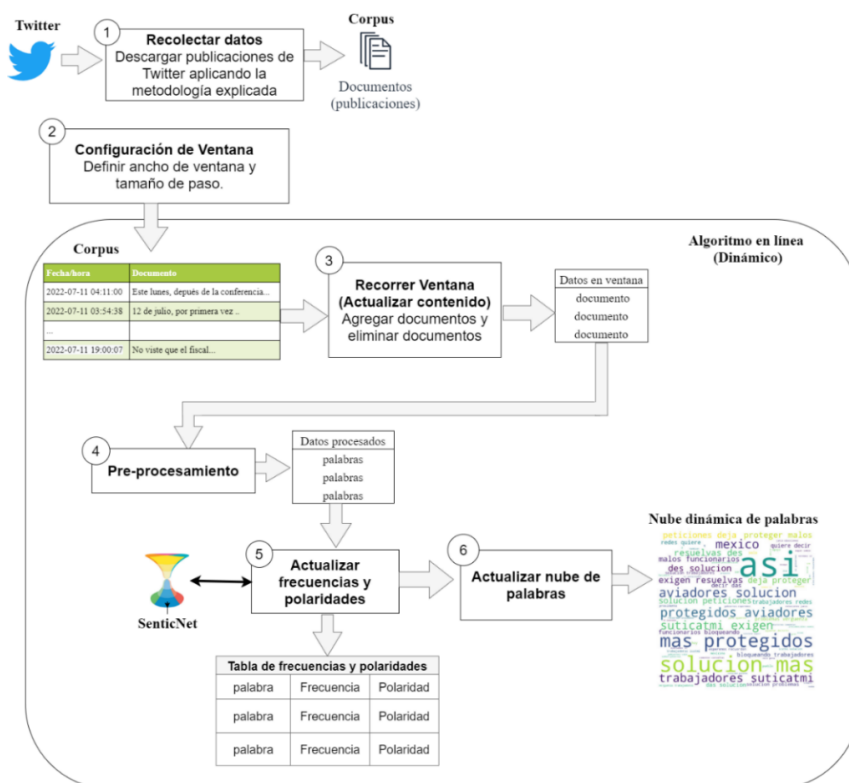


Figura 1. Resumen de la metodología aplicada. Fuente: Elaboración propia

Primer paso: Recolección de datos

La recolección de datos es una de las etapas más importantes de la metodología para la creación de la nube dinámica de palabras, el tema y los criterios de búsqueda dependerán del fenómeno que se desea estudiar [5], para el artículo, se eligió el contexto político mexicano actual, debido a las próximas elecciones en 2023. Se utilizó la biblioteca Tweepy, una de las bibliotecas para Python más populares para descargar documentos de Twitter.

Para realizar la colección de datos, se propone el siguiente proceso:

1. Comenzar con una lista de términos de búsqueda (hashtags y usuarios) relacionados con el tema, especificando el periodo y el lugar para recolectar datos.
2. Buscar para cada elemento de la lista en Twitter y descargar los documentos.
3. Aplicar un muestreo aleatorio simple a los documentos para elegir el tamaño adecuado del subconjunto. El tamaño del muestreo se calcula con la siguiente fórmula:

$$n = \frac{n}{1 + \frac{z^2 p(1-p)}{\varepsilon^2 N}}$$

Donde:

- $n = \frac{z^2 p(1-p)}{\varepsilon^2}$
- $z = 1.644854$
- P : porción de población ($P = 0.6$)

- ε : *margen error* ($\varepsilon = 10\%$)
 - N = *tamaño de población* (número de tweets descargados)
4. Leer los tweets de la muestra y seleccionar aquellos que sean relevantes para el estudio.
 5. Agregar a la lista las menciones y hashtags que aparecen en los archivos seleccionados, y que aún no están en la lista.
 6. Repetir los pasos del 2 al 5 tres veces.
 7. Usar la lista ampliada de hashtags y usuarios creada anteriormente para encontrar datos relevantes y descargar los documentos. Este conjunto de documentos se le denomina *Corpus*.

Las descargas de documentos con biblioteca Tweepy incluyen, además del texto, la fecha y la hora de la publicación, entre otros atributos.

Segundo paso: Configuración de la ventana

Con el objetivo de observar la evolución de la frecuencia de uso de palabras en el corpus y su polaridad, se propone generar nubes de palabras sucesivas a partir de subconjuntos del corpus. Cada subconjunto contiene documentos publicados entre dos tiempos (fecha y hora), a los que llamaremos t_a y t_b , donde:

- t_a tiempo de inicio de ventana: Los documentos publicados a partir del tiempo t_a y hasta t_b son considerados para generar la nube de palabras.
- t_b tiempo de fin de ventana: Los documentos publicados hasta el tiempo t_b pero desde t_a son considerados para generar la nube de palabras.

El ancho de ventana se define como:

$$W = |t_b - t_a|$$

Dado un corpus con la estructura siguiente:

$$C = \{(t_i, d_i): t_i \text{ tiempo}, d_i \text{ texto}, i = 1, \dots, N\}$$

El subconjunto $S_m \subseteq C$, es el conjunto de documentos que pertenecen a la ventana m , está definido como:

$$S_m = \{d_i: d_i \in C, t_a \leq t_i \leq t_b\} \quad m = 0, 1, 2, \dots$$

Donde:

$$\begin{aligned} t_a &= t_{1+m \times \text{step}} \\ t_b &= t_a + W \\ \text{step} &\subseteq \mathbb{Z}^+: \text{desplazamiento de la ventana en el tiempo} \end{aligned}$$

Los parámetros W y step son valores que se configuran manualmente por el usuario y deben de cumplir con la condición:

$$W > \text{step}$$

Tercer paso: Recorrer o desplazar la ventana

Una vez establecido el ancho de la ventana (W) y el tamaño del paso (step), se comienza con un proceso en línea de actualización de frecuencias de palabras y sus polaridades, con los elementos del conjunto actual S_m . Este último se genera cuando la ventana avanza (o retrocede) un paso. La Figura 2 ejemplifica el desplazamiento de la ventana en un paso.

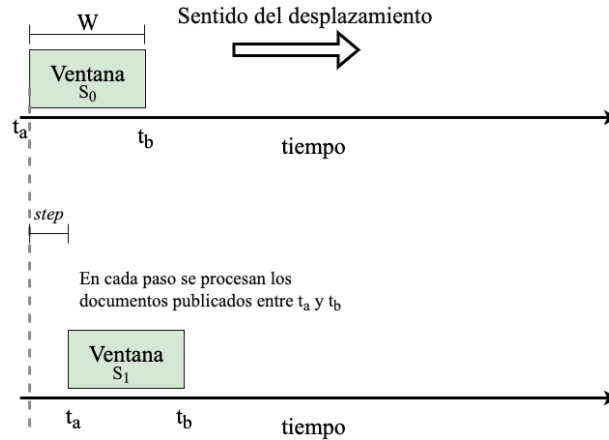


Figura 2. Desplazamiento de ventana. Fuente: Elaboración propia.

La idea es similar a una ventana que se desliza hacia los lados para permitir la entrada de documentos y rechazar los que ya han sido utilizados en la nube dinámica de palabras. El uso de una ventana deslizante garantiza que la nube de palabras sea vista a través del tiempo, es decir, visualizar gráficamente el análisis de sentimientos a través del tiempo.

Cuarto paso: Pre-procesamiento de documentos (textos)

El procesamiento de documentos consiste fundamentalmente en una serie de transformaciones que se realiza al documento (texto), con el objetivo de obtener datos uniformes y sin elementos que puedan afectar a la calidad de los resultados. De manera general, para el procesamiento de análisis de opiniones se eliminaron los números, acentos, símbolos, signos de puntuación y cadenas que contienen un URLs, usuarios y/o hashtags.[10]

Para cada documento fechado (d_t) del subconjunto S_m se realiza lo siguiente:

1. Eliminar caracteres especiales: Considerando que P es un conjunto de caracteres especiales $P = \{ | , " , ' , \$, \% , \& , / , (,) , = , ? , \grave{c} , * , - , + , : , ; , _ , < , > , [,] , \backslash , ^ , \textcircled{R} , \sim , ^\circ \}$, se realiza la operación mostrada en la ecuación (1)

$$d'_t = d_t - P \quad (1)$$

Donde:

d'_t : es el documento sin símbolos especiales.

2. Tokenizar: Se utiliza para dividir una frase en palabras, frases, símbolos u otros tokens significativos mediante la eliminación de los signos de puntuación [9], por lo que el documento modificado d'_t se separa cómo se observa en la ecuación (2)

$$T_{d'_t} = \{ \langle \text{palabra} \rangle_1, \langle \text{palabra} \rangle_2, \dots, \langle \text{palabra} \rangle_k \} \quad (2)$$

Donde:

$T_{d'_t}$: Conjunto de palabras o también llamado conjunto de tokens

3. Eliminación de palabras: Consiste en remover las palabras que son usuarios, hashtags, urls, y palabras vacías (*stopwords* en inglés). Las palabras vacías son aquellas palabras de uso común que no contribuyen de manera significativa en el texto. Dichas palabras se eliminan del conjunto de palabras antes del procesamiento ya que no son necesarias posteriormente. El conjunto de palabras es modificado como se muestra en la ecuación (3):

$$R_{d'_t} = (T_{d'_t} - U) \cup (T_{d'_t} - H) \cup (T_{d'_t} - Y) \quad (3)$$

Donde:

R_{d_t} : conjunto de palabras relevantes

$U = \{u | u \in \text{usuarios}\}$

$H = \{h | h \in \text{hashtags}\}$

$Y = \{y | y \in \text{stopwords}\}$

Finalmente, al conjunto de palabras del S_m , es expresado como se muestra en la ecuación (4):

$$S_m' = \left\{ \left(R_{d_t} \right) |_{i=1}^N R_{d_t} \in C \right\} \text{--- (4)}$$

Es importante mencionar que los documentos que se encuentran en la ventana (W) deben de pasar por un pre-procesamiento. Ya que cuando los documentos del subconjunto S_m sean pre-procesados, se obtendrá un conjunto de palabras S_m' . Este conjunto se utilizará posteriormente para la creación de las nubes de palabras.

Quinto paso: Actualización de frecuencias y polaridades

Consiste en obtener de cada palabra del conjunto S_m' una frecuencia y una polaridad, es decir, obtener una estructura numérica de la palabra para que posteriormente con dicha información el algoritmo pueda comprender y generar nubes de palabras de acuerdo con las emociones del modelo de SenticNet.

En este proceso existen tres aspectos a considerar:

1. Para generar las nubes de palabras emocionales solo se consideran las palabras que cuentan con una emoción proporcionada por SenticNet, por lo que inicialmente el conjunto S_m' es reducido.
2. Si $m = 0$ en el conjunto S_m' :
Se obtienen todas las frecuencias y polaridades del conjunto S_m'
3. Si $m \neq 0$ en el conjunto S_m' :
Se actualizan las frecuencias y polaridades del conjunto S_m'

Para el cálculo de las polaridades se utilizó la propiedad '*polarity_value*' del concepto de la palabra que proporciona SenticNet, este valor va de -1 a 1, siendo los valores negativos una polaridad negativa y los valores positivos una polaridad positiva. Así mismo, el cálculo de las frecuencias absolutas da información acerca de las veces que se repite una palabra en la ventana actual (S_m').

Sea:

$$E = \{e | e \in \text{modelo sentimental de SenticNet}\}$$

Para cada e^i se obtiene un conjunto de tuplas, que contendrán la información respectiva de las palabras ubicadas en la ventana m , su estructura es la siguiente (ecuación (5)):

$$D_i(W) = \{(w_j^i, f_j^i, p_j^i) \in e^i\} \text{--- (5)}$$

Donde:

W : ancho de la ventana

e^i : sentimiento i-esimo del conjunto E

w_j^i : palabra j de la ventana m correspondiente al sentimiento e^i

p_j^i : polaridad asignada a la palabra w_j^i de la ventana m correspondiente al sentimiento e^i

f_j^i : frecuencia asignada a la palabra w_j^i de la ventana m correspondiente al sentimiento e^i

Conforme transcurre el tiempo, la ventana m va incrementando, por lo que las frecuencias y las polaridades de $D_i(W)$ se deben actualizar, es decir, existirán palabras que serán removidas o añadidas a la ventana. La figura 3 ejemplifica esta actualización.

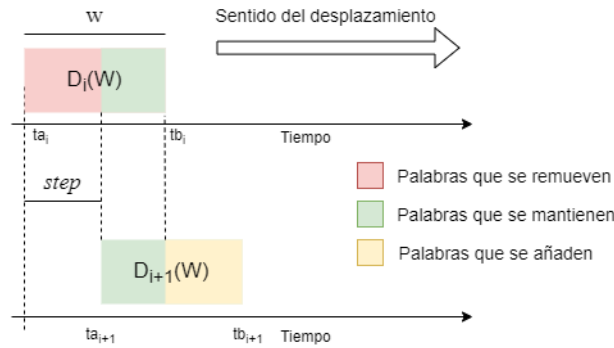


Figura 3. Actualización de palabras. Fuente: Elaboración propia.

Como se observa en la figura 3, las palabras w_j^i del conjunto $D_i(W)$ que se encuentran entre el periodo $|ta_{i+1} - ta_i|$ son las palabras que serán removidas de la ventana m . Por lo que las frecuencias de dichas palabras son disminuidas; matemáticamente se expresa:

$$f_2 = f_1 - k$$

Donde:

k : número de veces que se repite w_j^i en el periodo $|ta_{i+1} - ta_i|$

$$f_2: f_j^i \in D_{i+1}(W)$$

$$f_1: f_j^i \in D_i(W)$$

Posteriormente de remover las frecuencias, se debe verificar si alguna palabra w_j^i tiene una frecuencia igual a 0 ($f_j^i = 0$), si es el caso, la tupla (w_j^i, f_j^i, p_j^i) será totalmente removida del conjunto $D_{i+1}(W)$. De igual manera, las palabras w_j^i del conjunto $D_i(W)$ que se encuentran entre el periodo $|tb_{i+1} - tb_i|$ son las palabras que serán añadidas de la ventana m . Por lo que las frecuencias de dichas palabras son aumentadas; matemáticamente se expresa:

$$f_2 = f_1 + k$$

Donde:

k : número de veces que se repite w_j^i en el periodo $|ta_{i+1} - ta_i|$

$$f_2: f_j^i \in D_{i+1}(W)$$

$$f_1: f_j^i \in D_i(W)$$

Considérese que para cada palabra nueva w_j^i en $D_{i+1}(W)$, se debe calcular su polaridad.

Sexto paso: Actualizar la nube de palabras

Una vez obtenido el conjunto $D_i(W)$, cada frecuencia f_j^i para w_j^i será ponderada con su polaridad p_j^i , siguiendo la ecuación (6). El objetivo es ponderar la frecuencia de w_j^i con la intensidad emocional de la palabra para obtener una representación visual tanto en frecuencia como en polaridad.

$$fit(w_j^i) = |p_j^i| * f_j^i \quad \text{--- (6)}$$

Al final del cálculo de las ponderaciones se obtiene un conjunto de valores con la siguiente estructura:

$$FIT_i(D_i(W)) = \{(w_j^i, fit(w_j^i)) \in e^i\}$$

Donde:

$D_i(W)$: conjunto de tuplas de la ventana m correspondiente al sentimiento e^i

w_j^i : palabra j de la ventana m correspondiente al sentimiento e^i
 $fit(w_j^i)$: peso asignado a la palabra w_j^i

Finalmente, se utilizó la biblioteca wordcloud de Python para la creación de las nubes de palabras emocionales, indicando el color de las palabras basado en base el "Hourglass of emotions" de SenticNet, con los resultados de $FIT_i(D_i(W))$. Los sentimientos utilizados son la confianza (+), el miedo (-), la negatividad (-), tristeza (-), ira (-), sorpresa (+), positivo (+), asco (-), alegría (+) y anticipación (+). En este conjunto, los símbolos "-" y "+" se refieren a la polaridad negativa y positiva, respectivamente [11].

Resultados y discusión

Como ejemplo de la aplicación de la metodología explicada en este artículo, se utilizaron datos del contexto político mexicano actual. Para los resultados se descargó un corpus que corresponde a cada tendencia identificada en el contexto político, estas tendencias son: #4T, #4taTransformación, #amlo, #AmloElFracasoPresidencial, #AMLOVE, #FuerzaPresidente, #elcacasobrador, #ElPeorGobiernoDeLaHistoria, #ElPeorPresidenteDeLaHistoria, #LopezBasuraPresidencial, #Nacional y #obrador.

Para el análisis exploratorio de los corpus recolectados, se realizaron algunas tareas básicas de análisis de sentimientos con ayuda de SenticNet. Se calculó la polaridad de sentimiento en un rango de [-1,1], de esta manera se puede visualizar una distribución de las palabras según sus valores de frecuencia y polaridad.

El corpus con el que se realizó la nube dinámica de palabras es el conjunto de datos de la tendencia #amlo, cuyos parámetros a especificar fueron el tamaño de la ventana con 4 días y el paso de la ventana con 2 días. En total se generaron 4 evoluciones donde se identificaron 8 emociones: {#anger, #calmness, #disgust, #eagerness, #fear, #joy, #pleasantness, #sadness}. En la figura 4, se muestran algunas nubes de palabras emocionales correspondientes a cada evolución.

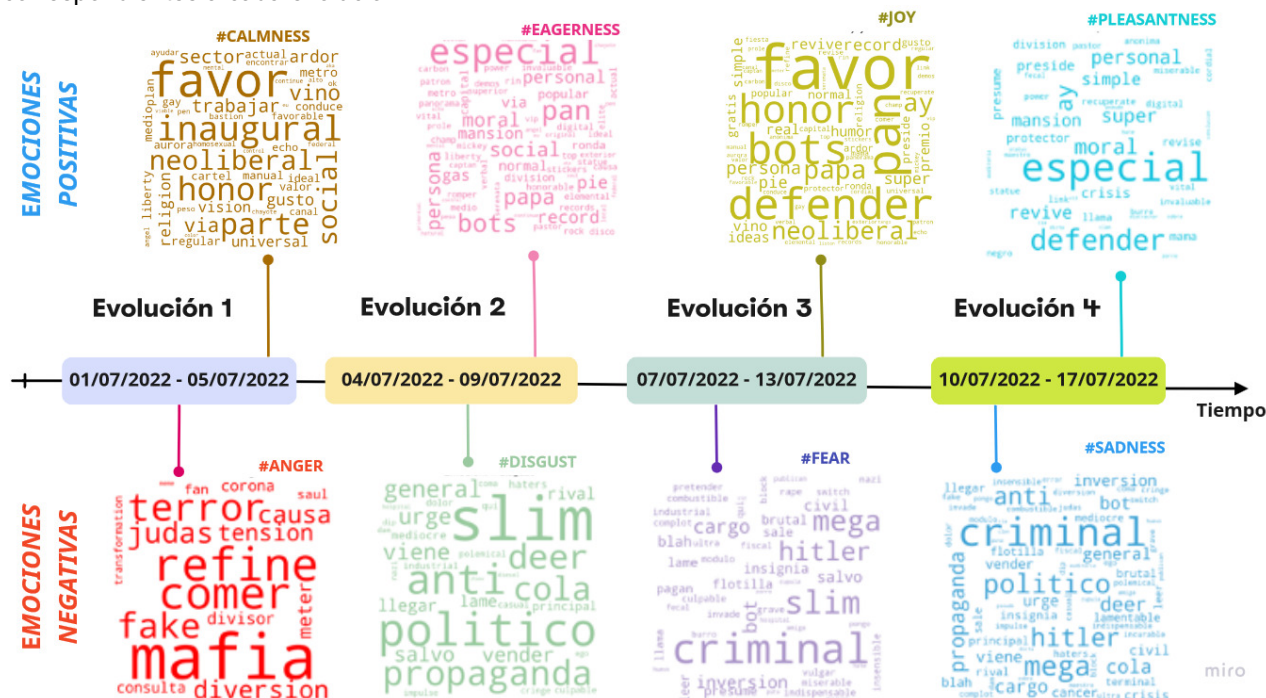


Figura 4. Nube dinámica de palabras basada en SenticNet. Fuente: Elaboración propia.

Con base en los resultados, se puede observar que, al hablar de política, se encuentran palabras negativas asociadas que predominan en la nube dinámica de palabras, como mafia, terror, judas, tensión, criminal y Hitler; igualmente las palabras positivas que predominan son: favor, especial, honor, neoliberal y especial. Se observa que en la nube de palabras de la emoción #calmness existe más palabras asociadas a este sentimiento, con respecto al resto de nube de palabras emocionales.

Para una interpretación más profunda de los resultados que se obtienen mediante la nube dinámica de palabras, se sugiere una colaboración interdisciplinaria con sociólogos, politólogos y expertos de las áreas de las ciencias

sociales. Lo anterior queda fuera del alcance de este artículo, ya que su principal objetivo es presentar la propuesta de diseño e implementación de la nube dinámica de palabras basada en SenticNet.

Trabajos futuros

Se continuarán explorando más técnicas para el análisis de sentimientos, con el fin de implementarlas en la nube dinámica de palabras. Por ejemplo, análisis automático de sarcasmos, negación simple o doble y análisis de los discursos de doble sentido, con el fin de complementar la representación gráfica de la evolución de las palabras y emociones a través del tiempo. Además, se contempla incorporar un análisis exploratorio a los datos con otras técnicas estadísticas.

Conclusiones

El análisis de sentimientos se ha convertido en la actualidad en uno de los campos más estudiados, debido a la importancia del estudio de las emociones y sentimientos, para la comprensión de un fenómeno social o evento político. En este artículo, se presenta la propuesta de una nube dinámica de palabras basada en SenticNet. Así mismo, se mostró una metodología formal para la recolección de datos, y para la creación de nubes dinámicas de palabras con polaridad emocional. Las metodologías expuestas pueden servir como una herramienta tecnológica para visualizar las polaridades de documentos publicados en redes sociales de una manera novedosa, igualmente ayudan a lograr una mejor comprensión de la realidad actual en beneficio a la sociedad.

Bibliografía

- [1] Jorge Youseth Herrera Aguilar and José Antonio O'Quinn Parrales, "El papel de las redes sociales en la gestión de un gobierno independiente: La comunicación política a través de Facebook del gobierno de Jaime Rodríguez Calderón," 2020.
- [2] P. de Fin, D. E. Grado, D. M. Burgos, G. Bello Orgaz, and J. De, "Análisis de la polarización del discurso de comunidades en redes sociales y blogs."
- [3] D. Valle-Cruz, A. Lopez-Chau, and R. Sandoval-Almazan, "How much do Twitter posts affect voters? Analysis of the multi-emotional charge with affective computing in political campaigns," in ACM International Conference Proceeding Series, Jun. 2021, pp. 1–14. doi: 10.1145/3463677.3463698.
- [4] Pech May Fernando, "Minería de datos en Twitter con Python," pp. 1–11, 2019.
- [5] A. López-Chau, R. Sandoval-Almazan, D. Valle-Cruz, and J. Ignacio Criado, "Analyzing polarization through social media with artificial intelligence: The mexican presidential election in 2018," in ACM International Conference Proceeding Series, Jun. 2019, pp. 502–503. doi: 10.1145/3325112.3328214.
- [6] A. H. Ambrocio, G. Ramírez De La Rosa, and V. Tello, "Un método para el análisis de sentimientos bajo un enfoque supervisado usando recursos léxicos A Sentiment Analysis Method using a Supervised Approach based on Lexical Resources," 2018. [Online]. Available: <https://www.socialbakers.com/statistics/facebook/>
- [7] A. López Chau, D. Valle-Cruz, and R. Sandoval-Almazán, "Sentiment Analysis in Crisis Situations for Better Connected Government," 2020, pp. 162–181. doi: 10.4018/978-1-7998-4570-6.ch008.
- [8] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "International Journal of Computer Science and Mobile Computing Sentiment Analysis of Events from Twitter Using Open Source Tool," 2016. [Online]. Available: <http://grouplens.org/datasets/movielens/>
- [9] Luque Sanchez Maria Alejandra and Cortés Díaz Luis Felipe, "Análisis etiquetado de textos para predicción de la polaridad, enfoque semi supervisado y etiquetado automático," pp. 1–44, 2020.
- [10] A. López-Chau, D. Valle-Cruz, and R. Sandoval-Almazán, "Sentiment Analysis of Twitter Data Through Machine Learning Techniques," 2020, pp. 185–209. doi: 10.1007/978-3-030-33624-0_8.
- [11] D. Valle-Cruz, V. Fernandez-Cortez, A. López-Chau, and R. Sandoval-Almazán, "Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study of the H1N1 and the COVID-19 Periods," Cognitive Computation, vol. 14, no. 1, pp. 372–387, Jan. 2022, doi: 10.1007/s12559-021-09819-8.