

SISTEMA PARA EL MONITOREO DE OPINIÓN CENTRADO EN ENTIDADES A PARTIR DE TWITTER

José Alejandro Reyes Ortiz

Universidad Autónoma Metropolitana

jaro@azc.uam.mx

Ezra Saucedo Vargas

Universidad Autónoma Metropolitana

al2102010913@azc.uam.mx

Ángeles Belém Priego Sánchez

Universidad Autónoma Metropolitana

abps@azc.uam.mx

Resumen

El análisis de la reputación en línea es una tarea que está atravesando un gran momento, debido a que actualmente existe un fuerte interés en la gran cantidad de opiniones publicadas, ya sea positivas o negativas, que se generan alrededor de un evento o entidad con el fin de conocer su prestigio en las redes sociales. La minería de opiniones se enfoca en determinar la polaridad de las publicaciones de una entidad con la finalidad de monitorear opiniones en línea. Este artículo presenta un sistema para el monitoreo de la opinión centrado en entidades utilizando textos de Twitter en español generados en línea. El proceso completo involucra, primero decidir la polaridad de un texto, determinando si el contenido tiene implicaciones positivas, negativas o neutras, después, obtener un monitoreo global (opinión colectiva) de una entidad. Se utiliza un promedio de las polaridades de cada palabra del mensaje, y luego, un promedio de las polaridades de los mensajes generados para una entidad.

Palabras Clave: Análisis de opinión centrado en entidades, análisis de textos en español, minería de opiniones, monitoreo de medios.

Abstract

Online reputation is a very important task gaining a great interest because of the huge amount of opinions published, positive or negative, about a particular event or entity with the aim of knowing the status in social networks. Opinion mining is a research area focused in analyzing texts in order to determine its polarity in order to monitor the online opinions about entities. This paper presents a system for the monitoring of entity-centered opinion using Spanish twitter texts generated on line. The process is two-fold, first to decide the polarity of a text, verifying whether or not the content has positive, negative or neutral implications, thereafter, to determine the global monitoring (collective opinion) of an entity. The process requires to calculate the average of the message word polarities, and then, to obtain the average of the polarities of messages generated by a given entity.

Keywords: *Entity-centered, media monitoring, opinion analysis, opinion mining, Spanish text analysis.*

1. Introducción

Con el crecimiento acelerado de los medios de comunicación en línea, como el caso de las redes sociales Twitter y Facebook, los usuarios han tenido la libertad de expresar sus sentimientos u opiniones sobre entidades como personas, empresas, organizaciones, productos, servicios, entre otros. Estas opiniones, que en el caso de Twitter no exceden los 140 caracteres, generadas por los usuarios de las redes sociales han controlado y conducido la toma de decisiones por parte de empresas u organizaciones. Por ello, se tiene la necesidad de generar herramientas o sistemas de monitoreo de las redes sociales para analizar las opiniones de una entidad basada en los mensajes generadas sobre ellas en las redes sociales.

El análisis de las opiniones está atravesando un gran momento, debido a que actualmente existe un fuerte interés en conocer la polaridad de las publicaciones que se generan alrededor de un evento o entidad, ya sea positivas o negativas; con el fin de conocer el prestigio de dicha entidad en las redes sociales. Dentro de los estudios centrados en la clasificación de opiniones, positivas o negativas, se

encuentra el presentado en [Pang, 2002]. El cual utiliza datos de críticas de películas encontradas en la Web, éstos son empleados en tres algoritmos de clasificación (Naïve Bayes, entropía máxima y máquinas de soporte vectorial), superando los *baselines* producidos manualmente por un humano.

Las opiniones, fabricadas mediante una conversación o comentario, es posible consultarlas a través de foros, blogs o redes sociales, estas últimas siendo la novedad y teniendo el mayor auge actual. Un trabajo basado en una colección de entradas de blogs, es el presentado en [Fernández, 2011] que realiza el análisis de sentimientos y minería de opiniones en dichas entradas, mostrando la relevancia de los sistemas de aprendizaje automático como recurso para la detección de información de opinión. En [Carrillo-de-Albornoz, 2016], se propone una metodología para evaluar resúmenes en el contexto del monitoreo de la reputación en línea, ésta beneficia los informes de reputación.

Actualmente, muchos investigadores del ámbito de la recuperación de información y la lingüística computacional han focalizado sus investigaciones en las redes sociales, especialmente en Twitter, dada la gran cantidad de publicaciones que se generan día a día. De esta manera, se hace necesaria la creación de herramientas capaces de gestionar y manipular toda esta información disponible. Además de categorizarla de acuerdo al contenido generado por el usuario, las opiniones, con el fin de identificar los puntos de opinión vinculados a la posición de los usuarios con respecto a algún tema. Esto se basa en el constante monitoreo de los mensajes producidos en las redes sociales. Hasta el 2014, estos sistemas eran evaluados en el marco de la competencia RepLab [Amigó, 2013] para tuits en inglés y en español. A pesar de incluir el idioma español, dentro de la competencia, la mayoría de los trabajos reportados en la literatura se centran en el idioma inglés. En este sentido, los diferentes métodos aplicados para la clasificación de opiniones de tuits en inglés, han sido aplicados para el idioma español [Fernández, 2013]. Estos métodos han considerado desde la utilización de *n-gramas* de palabras, la reducción de éstas a su raíz e incluso su sustitución. Sin embargo, no se obtuvieron los resultados esperados dado que los tuits son muy difíciles de tratar, sobre todo debido a su brevedad y falta de contexto.

En [Sidorov, 2012] se presenta un enfoque, para la minería de opinión de tuits en español, basado en el funcionamiento y diferentes configuraciones de algoritmos de aprendizaje automático. A pesar de que los algoritmos empleados presenten buenos resultados para el idioma inglés, en este trabajo se muestra como los diferentes tamaños de *n-gramas*, la longitud del corpus, el número de clases de sentimientos, el corpus balanceado con respecto al corpus no balanceado y los diferentes dominios (configuraciones) afectan la precisión del algoritmo.

La generación de léxicos de palabras, que se encuentren anotadas con su correspondiente polaridad, es otro enfoque en el que se han orientados diferentes investigadores y que ha colaborado al monitoreo de la opiniones en los comentarios. En el caso del español, ejemplos de estos enfoques son los presentados en [Pérez, 2012] y [Brooke, 2009].

Debido a la longitud de los tuits, actualmente, éstos incluyen abreviaturas del tipo mensajes cortos (SMS, acrónimo del inglés *Short Message Service*). Además de variantes léxicas, letras repetidas, uso de emoticones, empleo de mayúsculas para añadir énfasis, etc. Lo que ha provocado que existan enfoques basados en la normalización de este tipo de textos cortos, la mayor parte de enfoques están centrados en el inglés [Han, 2011]. Para el caso del español, la Sociedad Española de Procesamiento de Lenguaje Natural¹ está fomentando esta área de investigación, con la realización del sitio *Tweet-norm*², un taller centrado en la tarea de normalización de tuits [Alegria, 2013].

Además de las investigaciones realizadas, también se pueden encontrar herramientas web que monitorean las opiniones en redes sociales. Herramientas similares al sistema presentado en este artículo. Tal es el caso de *SWB Social* [SemanticWebBuilder, 2017] y *Keyhole* [Keyhole, 2016]. En el primer caso, se trata de una herramienta que utilizando la tecnología de semántica e implementando el análisis, monitoreo de mensajes y sentimientos puede identificar cuando un comentario es positivo, negativo o neutro. Además, esta herramienta permite que la aplicación aprenda a identificar los mensajes y a mejorar su desempeño en

¹ Para más información sobre la Sociedad Española de Procesamiento de Lenguaje Natural, consultar el sitio <http://www.sepln.org/>

² *Tweet-norm* disponible en <http://komunitatea.elhuyar.eus/tweet-norm/>

clasificación y atención. Pero a diferencia de la propuesta es una herramienta de uso comercial, por lo que tiene un costo su utilización. En el caso de la segunda herramienta, ésta es una aplicación web que se asemeja mucho a esta propuesta. El diseño es claro, evalúa y grafica los tuits, pero al mismo tiempo brinda resultados de la red social Instagram, llegando a ser un poco compleja al interpretar los resultados. Tiene el inconveniente de ser una herramienta de uso limitado, solo por 3 días, una vez terminado este período es necesario registrarse y pagar para continuar utilizándola. En este sentido, la propuesta que se presenta será más sencilla y clara al mostrar los resultados.

Como se expone anteriormente, existen herramientas de análisis de opiniones que consideran mensajes de Twitter en diversos lenguajes. En el caso del español se tiene una brecha tecnológica en el análisis de medios, lo que genera una carencia de recursos de análisis de textos, uno de ellos el monitoreo de la opinión centrado entidades a partir de los mensajes generados por usuarios en dicha red social.

Hacer uso de los datos generados en Twitter es una gran oportunidad para ganar tiempo en las decisiones tomadas, debido a que en esta red social éstos pueden ser adquiridos en tiempo real. Con ello, las decisiones tomadas estarán fundadas en datos actuales, generados casi de manera instantánea. A partir de los datos adquiridos se puede realizar un análisis automático y generar estadísticas sobre la opinión colectiva (positiva o negativa) de un producto, servicio o persona. Dicho análisis es de gran utilidad para los analistas de medios, desde la disminución de tiempos hasta la disminución de costos, por ejemplo en los estudios manuales se nota un consumo inmenso de tiempo y costos.

Por ello, en este artículo se presenta un sistema para el monitoreo de la opinión en línea centrado en entidades mencionadas en mensajes de Twitter para el idioma español. Estos mensajes son textos cortos, de una longitud de máximo 140 caracteres.

El sistema involucra dos procesos primordiales:

- La obtención de la polaridad de un mensaje, el cual se obtiene mediante un promedio de la polaridad individual de cada palabra, ésta forma parte de un lexicón ponderado para el análisis de sentimientos.

- El cálculo de una polaridad global de la entidad mediante el promedio de las polaridades de todos sus mensajes relacionados. Adicionalmente, el sistema presenta la opinión colectiva de la entidad en cuestión de manera comprensible para el analista de datos, presentando una aportación importante en la solución al problema de la carencia de recursos de análisis de textos en español.

El resto de este artículo está organizado como sigue. La Sección 2, presenta el desarrollo del sistema de monitoreo de opinión, sus procesos de cálculos de polaridades de mensajes y la presentación de la opinión colectiva. La Sección 3, presenta la experimentación de la herramienta para el monitoreo de opinión centrado en entidades para tres dominios. En la sección 4 se presenta la discusión de los resultados obtenidos en la experimentación del sistema. Finalmente, en la Sección 5 se exponen las conclusiones y el trabajo futuro.

2. Métodos

El sistema completo de monitoreo en línea de entidades, a partir de sus mensajes de Twitter, involucra tres etapas. La adquisición de datos, que se encarga de obtener los mensajes de Twitter para una entidad en cuestión; la obtención de la polaridad, que incluye una limpieza de los mensajes, pesado y la identificación de la polaridad; finalmente, un monitoreo, que consiste en una aplicación que visualiza los resultados de la polaridad de los mensajes y con ello proporcionar un punto de referencia de la opinión colectiva de la entidad. Estas etapas se describen de manera detallada en las siguientes secciones.

Adquisición de Datos

La adquisición de los datos, se realiza mediante la Interfaz de Programación de Aplicaciones de Twitter denominada “*Streaming API*” [Twitter, 2017], que proporciona acceso al flujo global de datos de Twitter con una baja latencia. La idea principal de esta interface de comunicación es que una aplicación solicitante estará recibiendo mensajes producidos en Twitter de manera constante.

Esta interface de comunicación permite obtener mensajes de este flujo global. Cada mensaje tiene una estructura en formato de un objeto de Java (JSON), el cual contiene información inherente al mensaje como la fecha de creación, un identificador único, hora en que fue creado, localización geográfica, datos del usuario (nombre de usuario).

En la tabla 1 se puede observar el código en JavaScript para la extracción de mensajes de Twitter y su código JSON generado utilizando la *Streaming API*. Para este proceso es necesario verificar las credenciales y permisos (*conexion.verifyCredentials*) para conectarse a la red social, después, se establecen los valores de la frase a analizar (*conexion.stream*) para iniciar la conexión (*stream.on*). Cabe destacar que este trabajo está interesado en mensajes de Twitter en español, por ello, se valida dicha pertenencia (*data.lang === 'es'*) para finalmente, obtener los datos del mensaje, tales como el nombre de usuario (*data.user.name*), texto del mensaje (*texto: data.text*), y foto del usuario (*data.user.profile_image_url*).

Tabla 1 Extracción de mensajes de Twitter y código JSON de su estructura.

Código JavaScript para la extracción de mensajes	Estructura de un mensaje
<pre>// verificación de credenciales conexion.verifyCredentials(function (data) stream = conexion.stream('statuses/filter', {'track': fraseaAnalizar}, function (stream) //Encendido del canal de extracción stream.on('data', function (data) // Extracción de mensajes en español if (data.lang === 'es') { var mensaje = {nombre: data.user.name, texto: data.text, foto: data.user.profile_image_url}; }</pre>	<pre>{ created_at: 'Tue Mar 21 01:24:10 +0000 2017', id: 843996623072124900, id_str: '843996623072124928', text: 'Tiene México el "cementerio" clandestino más grande de AL https://t.co/rzLsUAI4bM', source: 'Twitter for Android', user: { id: 1096406485, id_str: '1096406485', name: 'jorge nava alvarado'}</pre>

La interfaz de programación “*Streaming API*” hace posible la obtención personalizada de un conjunto de mensajes. En este trabajo, la personalización se ha llevado a cabo con base en el idioma (español). Con ello, la adquisición de

datos para la entidad en cuestión fue posible gracias a dicha personalización, esto provee como resultado un listado de mensajes en español, que para la sección de pruebas, serán 100 mensajes, como mínimo, para cada entidad.

Identificación de la Polaridad

El proceso de identificación de la polaridad, consiste en determinar si el contenido de un mensaje tiene implicaciones positivas, negativas o neutras. La idea es encontrar la polaridad de la opinión colectiva de una entidad basada en su conjunto de mensajes. Este proceso implica una serie de tareas sobre los mensajes, las cuales son descritas a continuación.

Pre-procesado de los Mensajes

Los mensajes son pre-procesados con la finalidad de mejorar el desempeño de la tarea para la identificación de la polaridad en un mensaje. Para ello, el texto de los mensajes es normalizado, es decir, se convierte a minúsculas y se obtiene la raíz de cada palabra. De esta manera, el algoritmo de identificación de polaridad encontrará mayores coincidencias de las palabras del mensaje con la lista de palabras pesadas con polaridades.

Ponderación del Mensaje

Cada mensaje de Twitter es representado como una lista de palabras pre-procesadas. Por su parte, el diccionario de palabras AFINN [Nielsen, 2011] es utilizado para determinar la polaridad del mensaje. Esta lista es pre-procesada de la misma manera que los mensajes para encontrar mayor coincidencia de las palabras.

La lista de palabras AFINN está conformada por 2477 términos y frases únicas, en su versión más actual denominada AFINN-111 [Nielsen, 2011]. Estos términos y frases son traducidos al español y revisados manualmente, por personas bilingües, para su utilización en este artículo. Cada elemento de esta lista tiene asignado un peso que corresponde a su carga positiva o negativa, utilizando un rango de puntuación de -5 (muy negativo) a +5 (muy positivo).

El objetivo principal de esta etapa es obtener un vector característico de cada mensaje con los valores de puntuación obtenidos de la lista AFINN. Para ello, se lleva a cabo una comparación de todas las palabras del mensaje con las palabras de la lista AFINN y se obtiene el valor de puntuación o polaridad de cada una, aquellas palabras que no tienen una referencia en la lista AFINN son ignoradas con la finalidad de no afectar la puntuación promedio del mensaje y por ende, la decisión de la polaridad (positivo, negativo o neutro) del mismo.

Decisión de la polaridad

En esta etapa se decide la polaridad total del mensaje, la cual depende de la puntuación promedio del mensaje, para calcular esta puntuación promedio se utiliza la ecuación 1.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, x_i \in L \quad (1)$$

El promedio (\bar{x}) considera solo aquellas palabras del mensaje (x) que tienen una referencia en la lista de palabras (L) del conjunto AFINN. Una vez obtenido la ponderación promedio del mensaje completo, es necesario obtener su polaridad. Para ello, se utiliza la fórmula de decisión que se presenta en la ecuación 2.

$$Polaridad(m_i) = \begin{cases} Positiva, & \bar{x} \geq +0.5 \\ Negativa, & \bar{x} \leq -0.5 \\ Neutra, & \text{otro caso} \end{cases} \quad (2)$$

Esta polaridad indicará la carga positiva o negativa de cada mensaje que pertenece a la entidad en cuestión. Sin embargo, el objetivo es monitorear la opinión colectiva de dicha entidad, actividad que se lleva a cabo mediante un sistema de visualización.

Monitoreo de Opinión Centrado en Entidades

La polaridad de cada mensaje es considerada para la visualización y análisis de la opinión en línea de una entidad en cuestión. La visualización, consiste en obtener el porcentaje correspondiente. Este porcentaje está basado en el número de mensajes que han sido decididos en cada categoría de polaridad: positiva, negativa y neutra.

Para esta tarea es necesario proporcionar el nombre de la entidad en cuestión y el sistema recolectará una lista de mensajes a partir del flujo global de datos de Twitter que mencionen a dicha entidad. Con este conjunto de mensajes, el sistema analiza y visualiza la opinión global, ver figura 1. En la cual se hace énfasis en la entidad (a), muestra la opinión colectiva utilizando una gráfica de dona (b), muestra un conteo de los mensajes para cada categoría (c) y finalmente, despliega la lista de mensajes para cada categoría “positivos”, “negativos” o “neutros” (d).

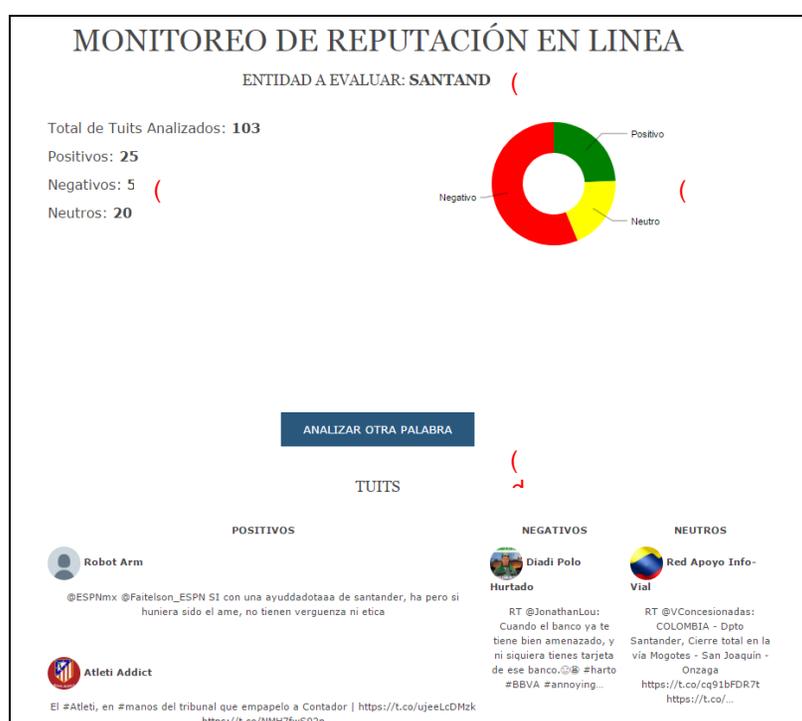


Figura 1 Sistema de monitoreo de opinión en línea.

3. Resultados

La evaluación del sistema de monitoreo de opinión en línea, se llevó a cabo con entidades de tres dominios para mensajes en español. Estos dominios son bancario, música y automóviles, los cuales representan una variación en cuanto al lenguaje utilizado por los usuarios para expresar sus opiniones al respecto. Inclusive, son los dominios proporcionados en la competencia de polaridad de la reputación de entidades en español denominada RepLab [Amigó, 2013].

Para cada dominio de interés, se selecciona un conjunto de entidades para el monitoreo de su opinión en línea. Esta selección ha sido desempeñada mediante un análisis estadístico de mayor cantidad de textos o mensajes generados en la red social según los datos proporcionados en la competencia RepLab [Amigó, 2013]. Mediante este análisis estadístico de entidades relevantes, se han seleccionado tres entidades por dominio de interés. La tabla 2, muestra las entidades seleccionadas por cada dominio.

Tabla 2 Entidades seleccionadas por dominio.

Dominio	Entidades seleccionadas
Automóviles	#ferrari, #volkswagen, #fiat
Bancario	#hsbc, #bbva, #santander
Música	#Shakira, #JustinBieber, #JenniferLopez

El monitoreo de la opinión en línea, para cada una de las entidades de los tres dominios, se realizó durante el mes de mayo de 2017. Los resultados del monitoreo de cada entidad se presentan por el número de mensajes generados en Twitter clasificados en positivo, negativo y neutro. Por ello, la tabla 3 muestra los resultados de cada categoría para todas las entidades.

Tabla 3 Resultados del monitoreo para las entidades.

Dominio	Entidad	Número de mensajes		
		Positivo	Negativos	Neutro
Marcas de automóviles	#ferrari	98	0	4
	#volkswag	0	117	4
	#fiat	78	19	4
Entidades bancarias	#hsbc	7	45	3
	#bbva	69	128	3
	#santander	25	58	20
Músicos	#Shakira	365	0	0
	#JustinBieb	0	37	64
	#JenniferL	22	33	67

Además de obtener el número de mensajes para cada entidad, el sistema de monitoreo de opinión en línea, también, muestra el resultado en una gráfica de tipo dona, en la cual se puede apreciar la opinión de la entidad en cuestión de una manera colectiva.

La figura 2, muestra el resultado del monitoreo para la entidad #fiat, del dominio automovilístico, mediante el gráfico y una lista resumida de mensajes clasificados.

En dicha figura se puede apreciar que esta entidad tiene una opinión global positiva, al menos, en el periodo de tiempo al que pertenecen los mensajes. Esto se puede concluir gracias a que tiene por encima del 77% de mensajes positivos.



Figura 2 Resultados gráficos y lista de mensajes para la entidad #fiat.

La figura 3, muestra el resultado del monitoreo para la entidad #bbva, del dominio bancario, mediante el gráfico y la lista de mensajes clasificados. El monitoreo de esta entidad ha mostrado una opinión global negativa, ya que el 64% de los mensajes son opiniones negativas.



Figura 3 Resultados gráficos y lista de mensajes para la entidad #bbva.

Finalmente, la figura 4, muestra el resultado del monitoreo para la entidad #JenniferLopez, del dominio musical. El gráfico expresa una opinión colectiva

dividida, el 18% de las opiniones son positivas, 27% de opiniones negativas y 55% de opiniones neutras que no reflejan un sesgo en la opinión de la entidad.

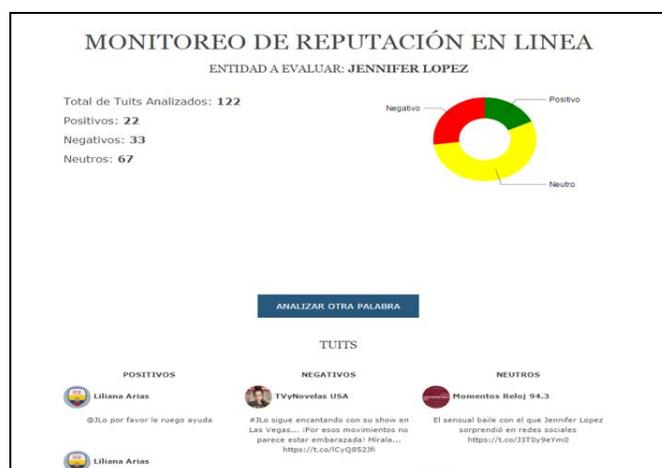


Figura 4 Resultados gráficos y lista de mensajes para la entidad #JenniferLopez.

4. Discusión

El sistema de monitoreo de opinión en línea centrado entidades, se ha evaluado en tres dominios, a saber: marcas de automóviles, entidades bancarias y artistas (músicos). Para cada dominio, tres entidades fueron seleccionadas, las cuales son estadísticamente significativas, es decir, entidades que generan 100 o más mensajes de twitter en un lapso de una hora. En estos tres dominios, los resultados fueron alentadores debido a que el sistema extrae de manera correcta los mensajes de la red social Twitter en un periodo de tiempo determinado (una hora). Además, monitorea la opinión en línea de las entidades en cuestión, mostrando los resultados de manera gráfica. Este monitoreo se ha realizado sin ningún problema para las 9 entidades (tres por dominio).

Cabe destacar que los resultados de la opinión colectiva dependen en gran medida de la entidad en cuestión y de la fecha en que se este llevando a cabo el análisis. Además, la mención de la entidad debe estar escrita de la manera esperada con o sin el símbolo *hashtag* (#).

El sistema propuesto es una aportación en el problema de carencia de recursos para el análisis de textos en español. Este sistema se puede extender a dominios adicionales a los experimentados, como la política, en el cual se puede monitorear

la opinión colectiva de los usuarios de redes sociales centrada en candidatos políticos que están participando en un proceso de campañas electorales.

5. Conclusiones

En este artículo se ha presentado un sistema de monitoreo en línea para entidades de tres dominios (automovilístico, bancario y musical) usando los mensajes generados por usuarios de la red social Twitter en el idioma español. El sistema abarca tres etapas. La primera tarea es la adquisición de los datos, la cual se lleva a cabo mediante una Interfaz de Programación de Aplicaciones, que se conecta a la red social Twitter y obtiene los mensajes en tiempo real, es decir, son generados casi al instante. Estos mensajes de texto son pre-procesados, ponderados y un algoritmo de decisión de polaridad es aplicado en esta segunda etapa. Finalmente, en la tercera etapa se visualizan los resultados del monitoreo en una aplicación Web utilizando un gráfico de tipo dona y mostrando la lista de mensajes positivos, negativos y neutros.

El algoritmo de decisión considera una ponderación promedio de mensaje, la cual es obtenida mediante los valores de polaridad de cada palabra del mensaje. Con ello se obtiene una ponderación global del mensaje y se decide si es positivo (mayor o igual a 0.5), negativo (menor o igual a -0.5) o neutro (en otro caso).

Las principales aportaciones de este trabajo se listan a) la obtención de una ponderación global por cada mensaje de la red social Twitter en español; b) el algoritmo de decisión mediante el cual se determina si un mensaje es positivo, negativo o neutro; c) el sistema completo de adquisición de mensajes, ponderación y visualización global de la opinión para una entidad específica.

Es importante destacar que este artículo hace una contribución importante en la carencia de recursos de análisis de textos para el español, desde que presenta el sistema de monitoreo en línea para entidades a partir de mensajes en español. Este aporte reduce el reto importante que tienen los analistas de medios de comunicación al llevar a cabo un monitoreo manual de entidades.

Como trabajo futuro, es posible experimentar con algoritmos de aprendizaje automático para la obtención de la polaridad de un mensaje, utilizando

clasificación supervisada de textos con la finalidad de determinar si un mensaje es categorizado como positivo, negativo o neutro. Adicionalmente, se puede realizar un aprendizaje profundo para encontrar los valores que mejor clasifiquen a un conjunto de datos, esto se verá reflejado en resultados precisos en el monitoreo de opinión en línea de una entidad en español. El sistema propuesto se puede extender a otros dominios como la política, para monitorear la opinión colectiva de candidatos políticos en un proceso de campañas electorales.

6. Bibliografía y Referencias

- [1] Alegria, I., Aranberri, N., Fresno, V., Gamallo, P., Padró, L., San Vicente, I., Turmo, J., & Zubiaga, A. Introducción a la tarea compartida Tweet-Norm, Normalización léxica de tuits en español. Tweet Normalization Workshop at SEPLN 2013: An overview. Proceedings of the Tweet Normalization Workshop co-located with 29th Conference of the Spanish Society for Natural Language Processing (SEPLN 2013). Vol. 1086, pp. 1-9, Septiembre 2013.
- [2] Amigó, E., Carrillo de Albornoz, J., Chugur, I., Corujo, A., Gonzalo, J., Meij, E., de Rijke, M., & Spina, D. Overview of RepLab 2013: Evaluating Online Reputation Monitoring Systems. P. Forner, H. Müller, R. Paredes, P. Rosso, B. Stein (eds.) CLEF. Lecture Notes in Computer Science. Vol. 8138, pp 333 – 352, 2013.
- [3] Brooke, J., Tofiloski, M., & Taboada, M. Cross-Linguistic Sentiment Analysis: From English to Spanish. RANLP, pp. 50-54, 2009.
- [4] Carrillo-de-Albornoz, J., Amigó, E., Plaza, L., & Gonzalo, J. Tweet Stream Summarization for Online Reputation Management. Ferro N. et al. (eds) Advances in Information Retrieval. ECIR 2016. Lecture Notes in Computer Science, vol. 9626. Springer, Cham, pp. 378-389, 2016.
- [5] Fernández, A., Nuñez, L., Morere, P., & Santos, A. Sentiment Analysis and Topic Detection of Spanish Tweets: A comparative Study of NLP Techniques. Procesamiento del Lenguaje Natural, Revista nº 50, pp. 45-52, 2013.

- [6] Fernández, J., Boldrini, Gómez, E. J. M., & Martínez-Barco, P., Análisis de sentimientos y minería de opiniones: el corpus EmotiBlog. *Procesamiento del Lenguaje Natural*, Revista nº 47, pp. 179-187, Septiembre de 2011.
- [7] Han, B., & Baldwin, T., Lexical Normalisation of Short Text Messages: Makn Sens a# twitter. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, pp. 368-378, 2011.
- [8] Keyhole. Hashtag Tracking for Twitter, Instagram and Facebook – Keyhole, Keyhole.co, 2016: <http://keyhole.co/preview>.
- [9] Nielsen, F. Å., A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts'*, Heraklion, Grece, pp. 93-98, Mayo 2011.
- [10] Pang, B., Lee L., & Vaithyanathan, S., Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10*. Association for Computational Linguistics, pp. 79-86, 2002.
- [11] Pérez-Rosas, V., Banea, C., & Mihalcea, R. Learning Sentiment Lexicons in Spanish. *LREC*, pp. 3077-3081, 2012.
- [12] SemanticWebBuilder, 2017: <http://www.semanticwebbuilder.org.mx/swb/swb/SWBSocial>.
- [13] Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbuck, A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., & Gordon, J., Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets. *Lecture Notes in Artificial Intelligence LNAI*, vol 7629, pp. 1-14, 2012.
- [14] Twitter Inc. Twitter Developer Documentation, 2017 Twitter Inc: <https://dev.twitter.com/streaming/overview>.