

# Analizando opiniones en las redes sociales\*

## *Analysing Opinions in Social Networks*

Javi Fernández, Fernando Llopis, Patricio Martínez-Barco,  
Yoan Gutiérrez, Álvaro Díez

Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante  
Apdo. de Correos 99, E-03080, Alicante, Spain  
{javifm,llopis,patricio,ygutierrez,adiez}@dlsi.ua.es

**Resumen:** La Web 2.0 ha focalizado la importancia de la información, no en unos pocos expertos en un tema, sino en una multitud de opiniones vertidas por usuarios a través de diversos medios en las redes sociales. Debido a ello, han cobrado un mayor interés los sistemas que son capaces de determinar qué es lo que piensan los usuarios sobre un determinado concepto, agregando diferentes fuentes de datos y aplicando cálculos de polaridad de las opiniones, que permiten determinar y comparar esos conceptos con otros similares. En este artículo describimos *Social Analytics*, nuestra visión sobre cómo deberían funcionar este tipo de sistemas, con una interfaz simple y optimizada que permita responder las necesidades de los usuarios.

**Palabras clave:** Análisis de opiniones, minería de opiniones, recuperación de información, redes sociales, turismo

**Abstract:** Web 2.0 has focused the importance of information, not on a few experts on a topic, but on a multitude of opinions expressed by users through various media on social networks. Due to this, there has been a increasing interest in systems that are able to determine what users think about a certain concept, by adding different sources of data and applying polarity calculations, to determine and compare these concepts with similar ones. In this paper we describe *Social Analytics*, our vision on how these kind of system should work, with a simple and optimized interface to meet the needs of the users.

**Keywords:** Sentiment analysis, opinion mining, information retrieval, social networks, tourism

## 1 Introducción

En la película *The Naked Jungle* se relataba la historia de una joven americana, Eleanor Parker, que se iba a casar con un terrateniente instalado en América del Sur, Charlton Heston. La joven llegaba con cierto temor a una jungla ante el peligro que podían suponer enormes animales salvajes. Lo que ella ignoraba es que el peligro real lo iban a suponer unas diminutas hormigas, que agrupadas en forma de millones podían suponer una marabunta que arrasara todo a su paso.

En la misma línea, con la Web 2.0 todo ha cambiado. Están empezado a perder valor las opiniones de grandes expertos en un tema concreto frente a las de cientos de usuarios que, con menos experiencia y conocimientos, opinan sobre los mismos temas. Si hablamos sobre teoría de la relatividad, sin duda alguna acudiríamos a los escritos de Einstein o discípulos para tratar de obtener más información. Es probable que nadie, o al menos casi nadie, pusiera en duda sus postulados frente a las opiniones de cientos de estudiantes de primero de universidad. Pero en cada vez más ámbitos estos cientos de opiniones pueden pesar más ante la toma de decisiones.

El sector turístico es un sector donde los conceptos web 2.0 se han impuesto. En una búsqueda del tipo «visitar iglesias de alicante capital» provocaría un resultado parecido al

\* Este trabajo ha sido parcialmente financiado por el Ministerio de Educación, Cultura y Deporte (MECD FPU014/00983), y la Universidad de Alicante, la Generalitat Valenciana y el Gobierno Español a través de los proyectos TIN2015-65136-C2-2-R, TIN2015-65100-R, PROMETEOII/2014/001 y FUNDACIONBBVA2-16PREMIOI.

mostrado en la Figura 1. La primera y la tercera opción propuesta corresponden a páginas oficiales, la primera al organismo de turismo de la ciudad y la tercera a la página del Ayuntamiento. Sin embargo, la segunda es la habitualmente seleccionada en un segundo click. Esta segunda corresponde a la famosa web de reservas e información turística *TripAdvisor*<sup>1</sup>. Dicha web contiene pequeñas opiniones de muchos usuarios. Los responsables de hoteles ya temen más unas cuantas opiniones negativas en una web de ese tipo que un análisis demoledor realizado por un experto y publicado en un diario o revista especializada.

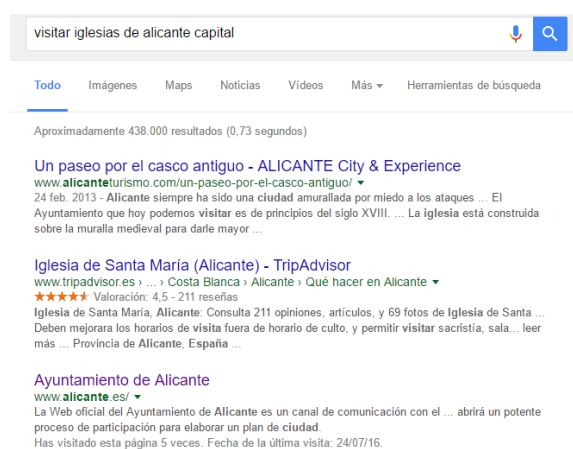


Figura 1: Retorno de un buscador de Internet

Al mismo nivel, aparecen las opiniones que los usuarios pueden verter habitualmente en las redes sociales. No hay mejor publicidad para un destino turístico que un usuario de una red social publique un comentario positivo sobre un lugar, y más todavía si es acompañado de una fotografía que inmortaliza el momento. Una playa, un atardecer, un copioso plato de arroz, un helado junto a un comentario positivo realizado por una persona a la que sigues en redes sociales es la mejor publicidad. De la misma forma, un comentario negativo acerca de la suciedad o inseguridad pueden ser auténticamente demoledores.

Pero al igual que ocurría en *The Naked Jungle*, una sola hormiga no produce ningún temor. De hecho, la aparición de una de ellas puede responder a un hecho anecdótico. La unión de ellas en forma de marabunta supone todo lo contrario. En el mundo de la Web 2.0 cada vez es más importante conocer en tiempo real las opiniones de la gente sobre lo

<sup>1</sup><https://www.tripadvisor.es>

que les interesa, les gusta o desprecian.

Por ejemplo, en el contexto del mundo del turismo (aunque sería aplicable a muchos más ámbitos) existe una doble visión: por una parte, la del usuario que pretende realizar una reserva; y por otro lado la del gestor del producto, que desea vender el suyo. Los objetivos del primero serían:

- No quiero equivocarme, ni pagar de más.
- Quiero conocer muchas opiniones, pero no tener que leerlas todas.
- Quiero que esas opiniones provengan del mayor número de fuentes posible.
- Quiero opiniones de expertos, pero también de gente como yo o en mis mismas circunstancias.

Y los objetivos del segundo:

- Quiero saber cómo conservar a mis clientes y ganar nuevos.
- Quiero conocer lo que piensa mi cliente sobre mí y sobre mi competencia.
- Quiero descubrir nuevos aspectos sobre mis potenciales clientes.

Ese conocimiento debe circunscribirse a la más inmediata actualidad, de nada me sirve saber que hace tres años un hotel era muy cómodo o que en el mismo período de tiempo un cliente pensara que la limpieza del local no era óptima. Como decía Lope de Vega «Lo que cuenta no es mañana, sino hoy. Hoy estamos aquí, mañana tal vez, nos hayamos marchado». Además, otro aspecto también de mucho interés es el concepto de medición y valoración. Como indica J.P. Rayo «Lo que no se mide no se puede gestionar». Necesitamos unos valores que nos determinen si mejoramos con respecto a días anteriores y cuál es nuestra posición con respecto a nuestra competencia.

En este artículo se presenta el sistema *Social Analytics*, que en base a las anteriores premisas analiza las redes sociales *Twitter*<sup>2</sup> e *Instagram*<sup>3</sup>, obteniendo mediciones de reputación de ciertos conceptos parametrizables en base al número de opiniones positivas y negativas recibidas. El sistema determina la valoración en base a una serie de fórmulas

<sup>2</sup><https://twitter.com>

<sup>3</sup><https://www.instagram.com>

que tienen en cuenta los aspectos de positividad y negatividad de las menciones, así como la influencia de los que las realizan. La información se agrupa de forma gráfica y en texto resumido.

El artículo se estructura en los siguientes apartados. El apartado 3 describirá la arquitectura del sistema *Social Analytics*. En el apartado 4 se realizará una descripción de modelo utilizado para determinar si un comentario se puede calificar como positivo, negativo o neutral. En el apartado 5 se detallarán las fórmulas utilizadas para determinar cómo se mide la reputación de un concepto. La sección 6 describirá la evaluación realizada. Por último, el apartado 7 recoge una serie de conclusiones y trabajos que se están realizando actualmente para mejorar el sistema.

## 2 Trabajos relacionados

Existen otros sistemas para visualizar datos de redes sociales, muchos de ellos centrados únicamente en estadísticas o únicamente en análisis de opiniones (Marcus et al., 2011; Hao et al., 2011; Wang et al., 2012). Nuestro sistema contiene visualizaciones de los dos tipos y además ofrece un valor de reputación para poder realizar estudios en el tiempo y poder realizar comparaciones. Entre los datos que extraemos podemos destacar: el texto del mensaje, la fecha de publicación, el autor, los usuarios mencionados, el lugar donde se ha escrito el mensaje, y el lugar de origen del autor. Este último dato es de formato libre y los usuarios pueden rellenarlo con lugares mal escritos o que no representen lugares reales. Obtener un lugar real a partir de estos datos es un problema muy complejo (Hecht et al., 2011; Peregrino, Tomás, y Llopis, 2013). Nuestra aproximación para este problema, junto a toda su arquitectura, se describe en la sección 3.

Por otro lado, es fundamental detectar si es positivo, negativo o si no se puede obtener un detalle claro de esa polaridad, en muchas ocasiones porque no la tiene. Se han utilizado diferentes técnicas tanto el análisis y estudio de esa polaridad en texto. El modelo propuesto por Hu y Liu (2004), se basa en un conjunto de adjetivos base o semilla que amplía utilizando las relaciones de sinonimia y antonimia proporcionadas por *WordNet* (Miller, 1995). *WordNet-Affect* es un modelo descrito en (Strapparava y Valitutti, 2004) y que ya utiliza seis categorías básicas de emociones

(alegría, tristeza, miedo, sorpresa, ira y disgusto) que también son expandidas utilizando *Wordnet*. El mayor problema que tienen estas propuestas es que no utilizan el contexto local para definir esa polaridad. *Social Analytics* utiliza un módulo de cálculo de la polaridad basado en el uso de n-gramas ampliado, con el fin de añadir contexto a las palabras y mantener parte de su secuencialidad. Este enfoque se describe en el apartado 4.

## 3 Arquitectura del sistema

El sistema se ha dividido en tres módulos principales: *escucha*, *procesamiento* y *presentación*, que a su vez utilizan tres bases de datos diferentes (ver Figura 2). Describiremos cada una de estas partes con detalle a continuación:

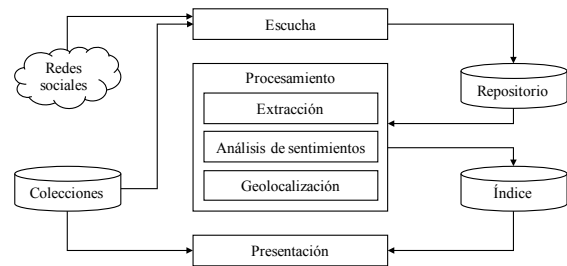


Figura 2: Arquitectura el sistema

### 3.1 Bases de datos

Primero describiremos las bases de datos utilizadas en nuestro sistema. Esto nos servirá para entender mejor cómo hemos representado los conceptos a analizar en las redes sociales y la información que finalmente queremos obtener.

- Colecciones.** En esta base de datos se almacenan los términos de interés para los usuarios. Cuando hablamos de términos nos referimos a conjuntos de palabras, hashtags o menciones a usuarios de las redes sociales. Ya que varios términos pueden representar a un mismo concepto, los agruparemos en lo que hemos llamado *entidades*. Al mismo tiempo, también agruparemos las entidades en *coleccion*, que son agrupaciones de entidades dentro de una misma categoría.

Por ejemplo, si estuviéramos interesados en lo que se comenta en las redes sociales sobre el Partido Popular (PP) y el Partido Socialista (PSOE) (dos partidos políticos conocidos en el panorama

político español), crearíamos dos entidades, una para cada partido. Estas entidades podrían tener diferentes términos por los que se las podría identificar, como “partido popular”, #partidopopular y @ppopular para el primero, o psoe, #psoe y @psoe para el segundo. A su vez, estas entidades, se podrían agrupar en una colección a la que podríamos llamar *Política*.

- **Repositorio.** Esta es una base de datos temporal en la que se almacenan los mensajes y comentarios encontrados en las redes sociales. Aquí, los datos se almacenan rápidamente y sin procesar, y se eliminan una vez procesados.
- **Índice.** Esta también es una base de datos de comentarios generados en redes sociales. Pero, a diferencia de la base de datos anterior, los datos se almacenan procesados e indexados de manera que realizar estadísticas y análisis sea lo más óptimo y eficiente posible. Cabe destacar que no se almacenan los mensajes completos, sólo se indexan.

### 3.2 Módulo de escucha

Este módulo es el encargado de la descarga de mensajes de las redes sociales. Se realizan búsquedas periódicas de los términos añadidos previamente utilizando la *API*<sup>4</sup> de cada red social para obtener mensajes que contengan dichos términos. La frecuencia de búsqueda dependerá de los límites de la API y del número de términos que tengamos en el sistema. Algunas APIs permiten obtener estos mensajes en *streaming*, esto es, ofreciéndonos los mensajes según se van publicando, sin necesidad de hacer búsquedas periódicas. En ambos casos, los mensajes obtenidos se almacenan el anteriormente mencionado *Repositorio*.

### 3.3 Módulo de procesamiento

Este módulo realiza toda la parte de extracción de datos, detección de la localización y análisis de opiniones de los mensajes descargados mediante el módulo de escucha. Como hemos comentado antes, extraemos el texto del mensaje, la fecha de publicación, el autor, los usuarios mencionados, el lugar donde se ha escrito el mensaje, y el lugar de

origen del usuario. Para este último punto hemos utilizado una aproximación muy simple, indexando una base de datos de lugares (*Geonames*<sup>5</sup>), realizando búsquedas en ella y quedándonos con el mejor resultado.

Utilizando el texto también obtenemos su polaridad, es decir, si el autor está escribiendo de forma positiva, negativa o neutral. Describiremos la aproximación utilizada con detalle en la sección 4.

### 3.4 Módulo de presentación

Este módulo se refiere a la interfaz de usuario. En la Figura 3 se puede ver la vista principal de un tablero de mando típico de Social Analytics. En el ejemplo podemos ver una comparativa de la reputación de los principales partidos políticos en España durante las elecciones de 2016.

De un sólo vistazo se puede acceder a los datos desde diferentes puntos de vista:

- Número de menciones que se han realizado sobre cada entidad. También es posible ver la evolución de este número en el tiempo.
- La audiencia a la que han llegado esas menciones (el número de seguidores de los autores de los mensajes).
- La reputación de cada entidad, según el número de menciones positivas, negativas y neutras (ver sección 5).
- Las palabras y hashtags más repetidos.
- Los usuarios que más han publicado.
- Los usuarios más mencionados.
- Los lugares en los que se han escrito los mensajes.
- Los lugares de origen de los usuarios que han publicado más mensajes.

Una de las grandes ventajas del sistema es que toda la información a considerar se puede parametrizar en base a diferentes aspectos, como el rango de fechas, la polaridad, o la fuente (*Twitter*, *Instagram* o ambas). El usuario también puede elegir en el mismo tablero si quiere cambiar alguna aspecto del informe y considerar solamente algunos elementos, como ver solo menciones positivas o negativas, ver las menciones de un usuario,

<sup>4</sup>Siglas de *Application Programming Interface*, o *Interfaz de programación de aplicaciones*

<sup>5</sup><http://www.geonames.org>



Figura 3: Cuadro de mando principal de Social Analytics

ver las menciones donde aparezcan ciertas palabras o hashtags, el lugar donde se ha escrito el mensaje, etc. Podemos ver otro ejemplo en la Figura 4.



Figura 4: Palabras, hashtags y lugares más repetidos

Además, el sistema incorpora un modo «versus» que permite contrastar la reputación de una o varias entidades frente a otras, con similar información, tal como se muestra en la Figura 5. Así, de un vistazo podemos saber quién va ganando en una confrontación de parámetros, o qué se dice de uno y otro en el mismo período de tiempo.

#### 4 Análisis de las menciones

Para detectar la polaridad de los mensajes hemos utilizado una aproximación híbrida (Fernández et al., 2013), que crea un lexicón a partir de un corpus etiquetado, y construye un clasificador utilizando técnicas de aprendizaje automático. Hemos evitado el uso de herramientas lingüísticas, con el fin de minimizar la propagación de errores externos a nuestra aproximación. A continuación detallaremos esta aproximación con más detalle.

##### 4.1 Extracción de términos

Como no queremos perder información subjetiva del texto original, realizamos una normalización muy básica. Utilizar una normalización más compleja puede inducir más errores que se propagarían a los resultados finales. Comenzamos convirtiendo los textos a minúsculas, eliminando nombres de usuario y URLs. Después, realizamos una eliminación de caracteres repetidos parcial: si el mismo carácter se repite más de tres veces, el resto de repeticiones se elimina. De esta manera, las palabras se normalizan, pero todavía es posible reconocer si las palabras originales tienen caracteres repetidos.

Una vez los textos están normalizados, extraemos las palabras que contiene. Después, obtendremos nuevos términos agrupándolos según la posición en la que aparecen en el texto. La mayoría de aproximaciones utilizan *n-gramas* para mantener parte de la secuencialidad de los textos, pero de manera muy estricta, ya que obligan a que los términos

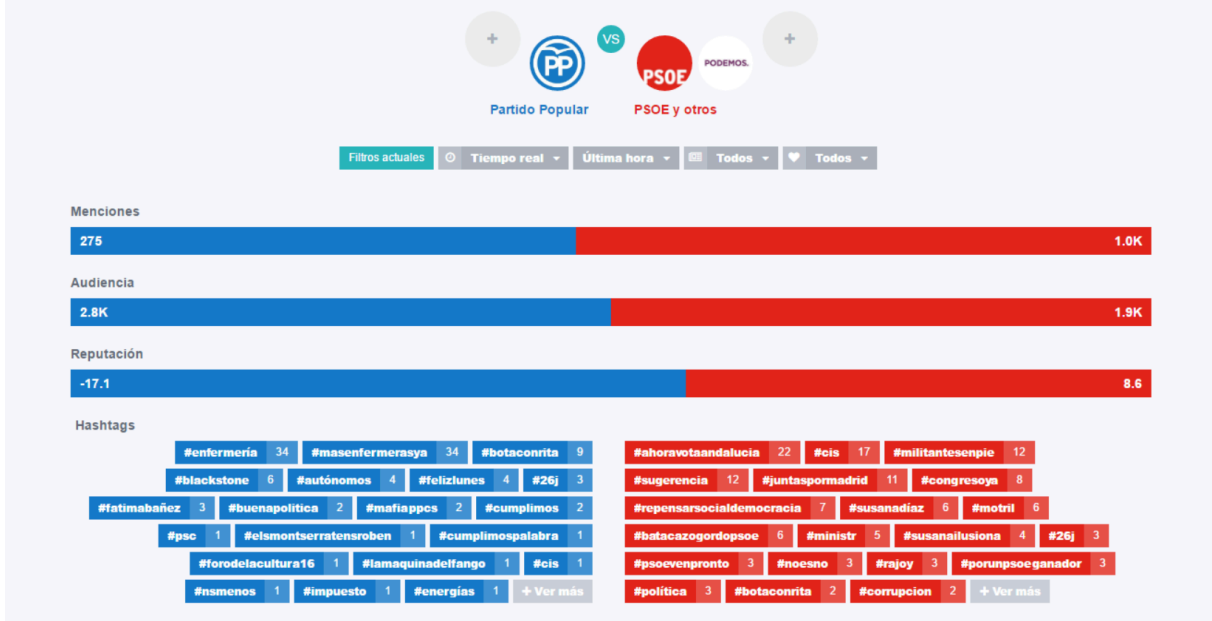


Figura 5: Modo «versus» para comparar pares de entidades

aparezcan siempre juntos. Por eso hemos decidido utilizar una aproximación basada en *skipgrams*, en la que los términos no tienen que aparecer juntos, sino que se permite que haya una cierta distancia entre ellos. Más específicamente, en un *k-skip-n-gram* se obtienen *n*-gramas en los que se permite una distancia máxima entre términos de *k*.

## 4.2 Puntuación de los términos

Nuestro lexicón consiste en una lista de skipgrams, donde cada uno tiene asociados diferentes valores de polaridad, indicando cómo es de fuerte la relación entre ese término y cada polaridad. También calcularemos una puntuación general para ver cómo es de importante (o frecuente) ese término en el corpus. Para construir este lexicón, necesitamos un corpus donde cada texto esté anotado con su polaridad. El peso de los skipgrams dependerá del número de veces que el skipgram aparece en textos de cada polaridad, y la distancia entre los términos de esas apariciones.

La ecuación 1 muestra la forma de calcular la puntuación general, donde  $T$  representa el conjunto de textos del corpus,  $t$  es un texto del conjunto  $T$ ,  $o_{s,t}$  representa una ocurrencia del skipgram  $s$  en el texto  $t$ , y  $skip$  es una función que nos indica la distancia entre los términos de  $o_{s,t}$ .

$$score(s) = \sum_{t \in T} \sum_{o_{s,t} \in t} \frac{1}{skip(o_{s,t}) + 1} \quad (1)$$

En la ecuación 2 se calcula la relación entre cada término y cada polaridad. Aquí,  $p$  representa una polaridad y  $T_p$  es el conjunto de textos del corpus etiquetados con esa polaridad  $p$ .

$$score(s, p) = \sum_{t \in T_p} \sum_{o_{s,t} \in t} \frac{1}{skip(o_{s,t}) + 1} \quad (2)$$

## 4.3 Aprendizaje supervisado

Utilizamos técnicas de aprendizaje automático para clasificar la polaridad de nuevos textos. Cada uno de los mensajes del corpus se utilizan como *instancias* de entrenamiento, y las polaridades etiquetadas se utilizan como *categorías*. Sin embargo, al contrario que las mayoría de aproximaciones, empleamos también las polaridades como *características* del modelo. El peso para estas características para cada texto se calcula según se especifica en la ecuación 3, donde  $S_t$  es el conjunto de skipgrams encontrados en el texto  $t$ .

Para construir nuestro modelo hemos elegido las máquinas de soporte vectorial, y más específicamente, la implementación por defecto de *LibSVM*<sup>6</sup>.

## 5 Medición de la reputación

A partir de la cantidad de opiniones positivas, neutras y negativas detectadas, y la

<sup>6</sup>www.csie.ntu.edu.tw/~cjlin/libsvm/

$$weight(p, t) = \sum_{o_{s,t} \in S_t} \left( \frac{1}{k(o_{s,t}) + 1} \cdot \frac{score(s, p)}{score(s, p) + 1} \cdot \frac{score(s, p)}{score(s)} \right) \quad (3)$$

cantidad de gente a la que han llegado estas opiniones, podemos calcular un valor de reputación para cada entidad en un período de tiempo concreto. Esta puntuación es un valor numérico, acotado dentro del intervalo  $[-1, +1]$ , donde  $-1$  sería la peor valoración y  $+1$  sería la mejor valoración dada por el sistema, que se calcula utilizando la ecuación 7.

$$pos = \sum_{p \in P_{e,t,+}} a_p \quad (4)$$

$$neg = \sum_{p \in P_{e,t,-}} a_p \quad (5)$$

$$neu = \sum_{p \in P_{e,t,0}} a_p \quad (6)$$

$$v_{e,t} = \frac{2 \cdot pos + neu - 2 \cdot neg}{2 \cdot (pos + neu + neg) + d_t} \quad (7)$$

Donde  $e$  es la entidad que estamos valorando;  $t$  es un período de tiempo determinado;  $P_{e,t,+}$ ,  $P_{e,t,0}$  y  $P_{e,t,-}$  representan los conjuntos de publicaciones que contienen una mención a la entidad  $e$  en el periodo de tiempo  $t$  con una polaridad positiva, neutral y negativa respectivamente;  $a_p$  es la audiencia o número de usuarios a los que ha llegado la publicación  $p$ ; y  $d_t$  es la duración en milisegundos del periodo  $t$ .

La suma de la duración del período en el denominador de la ecuación ( $d_t$ ) es una forma de dar una mayor valoración a las publicaciones que hayan llegado a más seguidores. Por ejemplo, si una entidad tiene menciones negativas que ha llegado a 100 personas en un minuto, su valoración sería de  $-3 \cdot 100 / (3 \cdot 100 + 60000) = -0,005$ . Sin embargo, si las publicaciones ha llegado a 10.000 personas, la valoración sería de  $-3 \cdot 10000 / (3 \cdot 10000 + 60000) = -0,333$ .

Cabe destacar que en nuestro sistema consideramos las menciones neutrales ( $p \in P_{e,t,0}$ ) como algo positivo, ya que el hecho de que una entidad sea mencionada en las redes sociales aumenta su valoración. De esta forma, si todas las valoraciones son neutrales, la reputación será mayor que cero, mientras que si no hay ninguna mención sería exactamente cero.

## 6 Evaluación

La evaluación del sistema se he enfocado en dos entornos diferentes. Por un lado se evaluó la eficacia del sistema a la hora de determinar si la polaridad detectada era la correcta. Las pruebas se realizaron a través de los datos de la tarea abierta *Sentiment Analysis in Twitter* del workshop *TASS 2015* (Villena-Román et al., 2015). Los resultados obtenidos en la misma se pueden observar en la Tabla 1, comparándolos con los resultados obtenidos con bigramas ( $n = 2$ ) y trigramas ( $n = 3$ ).

	Parámetros	F1
Bigramas	$n = 2$	0.636
2-skip-k-grams	$n = 2, k = 1$	0.642
	$n = 2, k = 2$	0.646
	$n = 2, k = 3$	<b>0.647</b>
	$n = 2, k = max$	<b>0.647</b>
Trigramas	$n = 3$	0.624
3-skip-k-grams	$n = 3, k = 1$	0.623
	$n = 3, k = 2$	0.630
	$n = 3, k = 3$	0.637
	$n = 3, k = max$	<b>0.639</b>

Tabla 1: Resultados de la evaluación

Se puede observar un beneficio en la utilización del modelo de skipgrams. La mejor puntuación se ha obtenido con  $n = 2$  y  $k = 3$  (o  $k = max$ ) comparando con los resultados con bigramas, con una mejora del 1,7%, y con  $n = 3$   $k = max$  respecto a los resultados con trigramas, con una mejora del 2,4%. Por lo tanto, se puede intuir que existen expresiones que identifican polaridad, cuyas palabras no aparecen explícitamente juntas en los textos, que el modelo de n-gramas no ha podido descubrir pero sí el modelo de skipgrams.

Por otro lado se realizó una prueba a nivel experiencia del usuario. Para ello se convocó a los responsables de turismo de las localidades de la provincia de Alicante. Se realizó un pequeño seminario de uso de la herramienta y se permitió a los diferentes usuarios utilizarla durante un tiempo. Entre las principales conclusiones de nuestro experimento podemos destacar las siguientes:

- El sistema era lo suficientemente intuiti-



vo y los datos que se mostraban eran de gran interés y utilidad para su trabajo.

- Los usuarios no entraron en valorar la calidad del clasificador de polaridad, sólo necesitaban una visión global de las opiniones de los usuarios.
- Había especial interés en los cambios bruscos de polaridad, y gracias a la lista de palabras y hashtags más repetidos, era posible adivinar cuál era la causa de dicho cambio.
- Su mayor preocupación era como afectaban determinados eventos o sucesos de forma positiva o negativa a un destino. Uno de los ejemplos más llamativos fue un accidente de autobús con numerosas víctimas mortales que se produjo en las inmediaciones de la ciudad de Valencia. Esto focalizó una gran cantidad de menciones con polaridad negativa a la ciudad e incluso a la festividad que tenía lugar en ese momento (Fallas de Valencia).

## 7 Conclusiones y trabajos futuros

*Social Analytics* es una herramienta que ya se está utilizando por el gobierno valenciano para monitorizar los avances de cada uno de los municipios que forman la comunidad en el apartado turístico. El sistema ha fomentado la competitividad y el deseo de mejorar las campañas, al mostrar información sobre la reputación de las localidades comparadas.

Como trabajo futuro incorporaremos más fuentes de información, algo que cada vez es más complicado, ya que las plataformas de las redes sociales son cada vez más conscientes del valor que tiene la información de la que disponen y se muestran cada vez menos predispuestas a permitir el acceso libre a la misma. También nos planteamos mejorar la evaluación de nuestra medida de reputación, comparando nuestros resultados con diferentes sondeos y encuestas realizadas en otros medios.

## Bibliografía

- Fernández, J., Y. Gutiérrez, J. M. Gómez, P. Martínez-Barco, A. Montoyo, y R. Muñoz. 2013. Sentiment Analysis of Spanish Tweets Using a Ranking Algorithm and Skipgrams. En *XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural (SEPLN 2013)*, páginas 133–142.
- Hao, M., C. Rohrdantz, H. Janetzko, U. Doyal, D. A. Keim, L.-E. Haug, y M.-C. Hsu. 2011. Visual sentiment analysis on twitter data streams. En *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, páginas 277–278. IEEE.
- Hecht, B., L. Hong, B. Suh, y E. H. Chi. 2011. Tweets from justin beiber’s heart: the dynamics of the location field in user profiles. En *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, páginas 237–246. ACM.
- Hu, M. y B. Liu. 2004. Mining and summarizing customer reviews. En *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, páginas 168–177. ACM.
- Marcus, A., M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, y R. C. Miller. 2011. Twitinfo: aggregating and visualizing microblogs for event exploration. En *Proceedings of the SIGCHI conference on Human factors in computing systems*, páginas 227–236. ACM.
- Miller, G. A. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- Peregrino, F. S., D. Tomás, y F. Llopis. 2013. Every move you make i’ll be watching you: geographical focus detection on twitter. En *Proceedings of the 7th Workshop on Geographic Information Retrieval*, páginas 1–8. ACM.
- Strapparava, C. y A. Valitutti. 2004. Wordnet affect: an affective extension of wordnet. En *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, volumen 4, páginas 1083–1086.
- Villena-Román, J., J. García-Morera, M. A. García-Cumbreras, E. Martínez-Cámara, M. T. Martín-Valdivia, y L. A. Urena-López. 2015. Overview of tass 2015. En *TASS 2015 Workshop on Sentiment Analysis at SEPLN*, volumen 1397, páginas 13–21.
- Wang, H., D. Can, A. Kazemzadeh, F. Bar, y S. Narayanan. 2012. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. En *Proceedings of the ACL 2012 System Demonstrations*, páginas 115–120. Association for Computational Linguistics.