

# MODELAMIENTO Y VISUALIZACIÓN DEL CONTEXTO GEOTEMPORAL DE EVENTOS EXTRAÍDOS DE REDES SOCIALES EN LÍNEA

En la actualidad, las redes sociales en línea tienen un papel importante en el consumo y propagación de información. En estas plataformas, los usuarios intercambian millones de mensajes al día, los cuales no sólo contienen conversaciones sin importancia sino también información acerca de lo que pasa en el mundo. Sin embargo, extraer conocimiento valioso de esta gran cantidad de datos no es una tarea fácil dado su volumen y diversidad. Además, presentarla de una manera entendible para personas, y no sólo para los computadores, tampoco lo es. En este artículo hablaremos de los enfoques actuales para entender el qué, cuándo y dónde de un evento noticioso extraído de las redes sociales, incluyendo algunas herramientas visuales que permiten un mejor entendimiento de estos datos.





## VANESSA PEÑA

Estudiante de Doctorado en Ciencias mención Computación, Universidad de Chile; Ingeniería Civil en Computación, Universidad de Chile. Su trabajo se enfoca en Minería de Datos y Visualización de Eventos Extraídos de Redes Sociales en Línea, y Educación en Computación, con énfasis en cómo atraer mujeres a Ingeniería y Ciencias.

[vpena@dcc.uchile.cl](mailto:vpena@dcc.uchile.cl)

## CONOCIMIENTO ÚTIL DESDE LAS REDES SOCIALES EN LÍNEA

Aunque muchas veces las redes sociales en línea son usadas solo para conversaciones triviales entre usuarios, también lo son para comunicar que algo está pasando en el mundo [11]. Esto se debe a que cuando sucede un evento noticioso los usuarios de redes sociales en línea reaccionan a él compartiendo y propagando información acerca de tal evento. De esta manera, los usuarios han pasado de tener un rol pasivo a uno activo en la producción de información, generando millones de mensajes por día. Este gran volumen de datos es muy heterogéneo y al no tener una estructura definida se hace complejo extraer de manera ordenada lo que los usuarios están compartiendo. Existen muchas aplicaciones en las cuales organizar, procesar y analizar estos datos puede aportar algo útil. Algunas de ellas son la detección de sismos [12], inferencia del valor y confiabilidad del contenido de los mensajes compartidos en Twitter [9], análisis geográficos [5], análisis de sentimiento [4], etc.

## EXTRAYENDO EL DÍA, LA HORA Y LAS COORDENADAS DE UN TWEET

Para entender una noticia que pasa en el mundo, es importante saber cuándo y dónde pasó. Es común que los mensajes de las redes sociales en línea tengan asociadas el día y la hora en que fueron creados, por lo que obtener el cuándo de un mensaje no es difícil. Haciendo un análisis del conjunto de mensajes encontrados en una red social en línea que se refieren a un evento, es posible tener una idea de cuándo ocurrió. Otra forma de extraer esta información es por el contenido mismo de los mensajes: Status Calendar<sup>1</sup> es una herramienta Web que extrae eventos populares que ocurrirán en el futuro desde Twitter. Los autores de esta herramienta buscan palabras como “concierto”, para saber cuál será el evento, y “el siguiente sábado”, para identificar el momento en que ocurrirá [10].

Por otra parte, aunque la mayoría de la plataformas de redes sociales en línea guardan las coordenadas desde dónde se compartió un mensaje o una fotografía, saber dónde pasó un evento no es trivial. Esto se ve dificultado aún más por

<sup>1</sup> <http://statuscalendar.com>



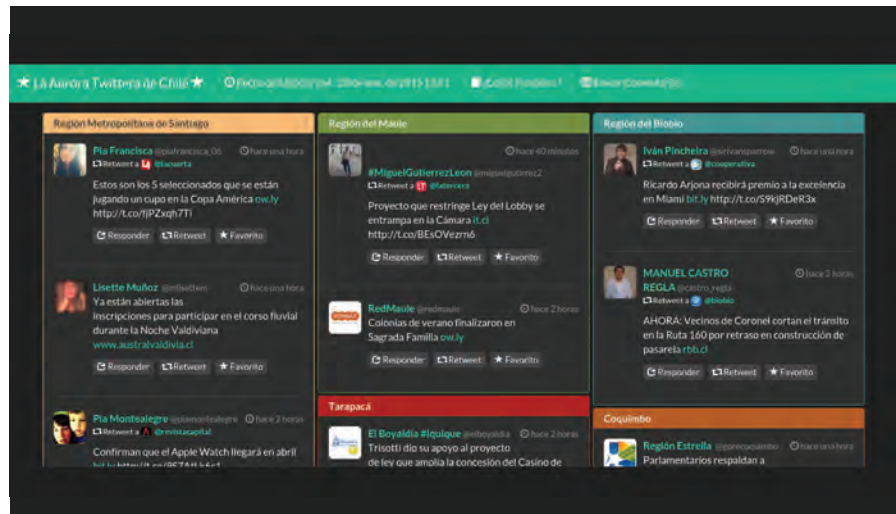
el hecho que el porcentaje de usuarios que permiten que se guarde su ubicación geográfica exacta es bastante bajo. En Twitter por ejemplo, alrededor del 3% de los mensajes posee coordenadas exactas<sup>2</sup>, por lo que es necesario buscar más fuentes de información como el perfil del usuario, su zona horario, su idioma, su red de contactos, entre otras.

El contenido de los mensajes compartidos en las redes sociales en línea también aporta información acerca del lugar de donde viene el usuario. La Aurora Twittera<sup>3</sup> es un sitio web cuyo objetivo es facilitar el acceso de información con diversidad geográfica dado el problema de centralización de nuestro país. Para lograr esto, los autores realizaron un estudio para determinar las palabras frecuentemente usadas para referirse a distintos lugares de Chile y así poder clasificar los mensajes de Twitter de acuerdo a la ubicación desde la que hablan [3]. La interfaz de esta herramienta puede verse en la **Figura 1**.

## ¿CÓMO USAR ESTA INFORMACIÓN?

Usando la ubicación desde donde fueron emitidos mensajes extraídos de redes sociales en línea y el momento en que fueron creados, se puede obtener conocimiento acerca de eventos que motivan a los usuarios a compartir estos mensajes. De manera general se puede decir que hay dos enfoques principales. El primero, es identificar dónde pasó un evento detectado y, el segundo, es el de extraer eventos por ubicación. El trabajo de Sakaki et al. [12], es un ejemplo del primer caso. Éste es un sistema que detecta y reporta sismos en tiempo real usando a los usuarios de Twitter como sensores, identificando el epicentro de éste y su trayectoria. Para esto, el sistema hace una búsqueda en tiempo real en el stream de Twitter para extraer todos los mensajes referentes a sismos, tifones y otros desastres naturales. Usando las coordenadas de los tweets o la ubicación del usuario

2 <http://blogs.law.harvard.edu/internetmonitor/2013/06/17/tweeters-geography-visualized-and-explained/>  
3 <http://auroratwittera.cl/>



**FIGURA 1.** CAPTURA DE PANTALLA DE LA AURORA TWITTERA.

desde su perfil, aplican algoritmos y técnicas de estimación para obtener el epicentro.

Para el segundo caso, podemos hablar del trabajo de Abdelhaq et al. [1], quienes presentan un sistema que permite detectar en tiempo real eventos reales geolocalizados desde Twitter. Su metodología se basa en seleccionar un conjunto de palabras que se repitan frecuentemente, obteniendo aquellos mensajes que las contengan y que además tengan coordenadas geográficas. Usando el conjunto resultante de mensajes se crean grupos de palabras por su similitud espacial y a cada grupo se le asigna un puntaje para seleccionar aquellos que representen eventos del mundo real que estén localizados.

## VISUALIZANDO LAS REDES SOCIALES EN LÍNEA

Dada la enorme cantidad de datos cuando se trabaja con redes sociales en línea, es importante proveer herramientas para que los humanos, y no solo los computadores, puedan entenderlos y hacer uso de ellos. El área de *Visual Analytics* crece cada vez más dada la importancia que tienen los grandes datos. Su

objetivo es el estudio de interfaces visuales para el razonamiento analítico. Existen muchas herramientas que permiten visualizar las redes sociales en línea. La principal motivación al visualizar un evento desde las redes sociales en línea en su componente geográfico, usualmente es mostrar desde dónde vienen los mensajes que hablan acerca de él. Esto significa que es posible ver la propagación de un evento, pero no necesariamente entender dónde efectivamente sucedió. Por ejemplo, si consideramos el terremoto en Chile de 2010, dado que las telecomunicaciones funcionaban de manera intermitente, la frecuencia de mensajes en Twitter provenientes de Chile fue más baja de la esperada en las primeras horas después del terremoto [9]. Esto significa que, en primera instancia, la distribución geográfica de tweets podría no dar indicios de que la noticia ocurrió efectivamente en Chile. Algunas herramientas visuales que se enfocan en mostrar la propagación de mensajes son *TwitInfo* [8] y *mTrend* [6].

*SensePlace2* [7] tiene un enfoque diferente e intenta resolver este problema. *SensePlace2* es una herramienta Web enfocada en permitir el análisis geovisual de eventos extraídos de Twitter, cuya interfaz puede verse en la **Figura 2**. La gran diferencia con la mayoría de las otras herramientas es que no sólo usa la información del lugar desde dónde se emitieron los tweets sino también de los lugares



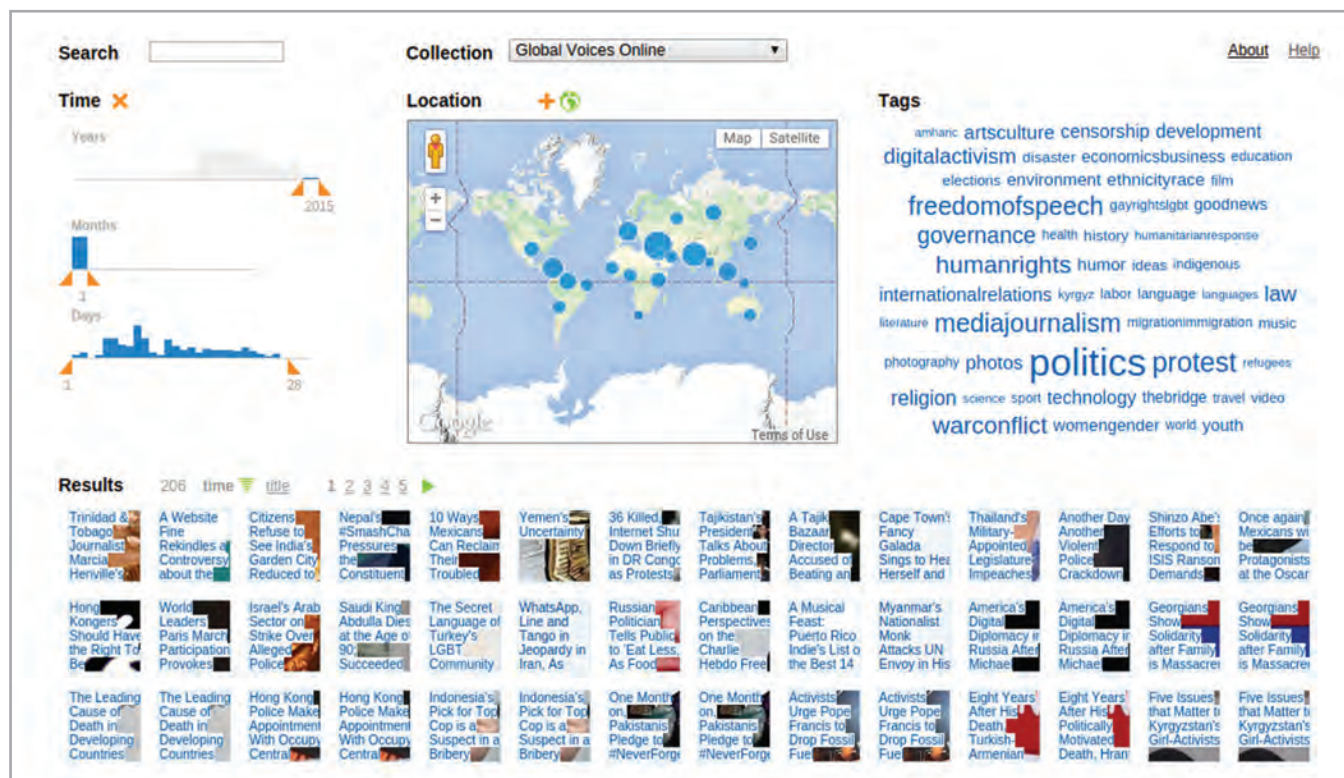
mencionados en ellos. Esto permite entender visualmente dónde sucedió un evento además de cómo se propagó a través de Twitter. Una gran desventaja de la herramienta es que no se enfoca en identificar eventos sino en agrupar tweets, por lo que es difícil identificar de manera correcta dónde pasó una noticia.

Finalmente es importante mencionar a Visgets[2]<sup>4</sup> ya que es una de las pocas herramientas que están disponibles para su uso en la Web. Esta herramienta visual, que puede observarse en la **Figura 3**, provee una interfaz que permite explorar entidades por tres características: tiempo, países y palabras. Algo muy interesante de esta herramienta es que permite distintas fuentes de información, como los proceedings de la conferencia WWW ACM, Flickr o Global Voices Online<sup>5</sup>. Es decir, podemos observar tanto papers de conferencia, imágenes o noticias y entenderlos por su ubicación geográfica.



**FIGURA 2.** CAPTURA DE PANTALLA DE SENSEPLACE2. IMAGEN EXTRAÍDA DESDE [HTTPS://WWW.YOUTUBE.COM/WATCH?V=YW7SHBEURXU](https://www.youtube.com/watch?v=YW7SHBEURXU).

<sup>4</sup> <http://mariandoerk.de/visgets/demo>  
<sup>5</sup> <http://globalvoicesonline.org/>



**FIGURA 3.** CAPTURA DE PANTALLA DE VISGETS DESDE LA DEMOSTRACIÓN EN LÍNEA.



## CONCLUSIÓN

LOS MENSAJES DE REDES SOCIALES EN LÍNEA NO SON SÓLO CONVERSACIONES TRIVIALES ENTRE USUARIOS, SINO QUE PUEDEN CONTENER INFORMACIÓN VALIOSA ACERCA DE LO QUE PASA EN EL MUNDO REAL. SIN EMBARGO, ESTOS REPRESENTAN DATOS MUY DIVERSOS Y SIN UNA ESTRUCTURA DEFINIDA, POR LO QUE ES DIFÍCIL SACAR CONOCIMIENTO VALIOSO DE ELLOS. DADO UN EVENTO DETECTADO EN LAS REDES SOCIALES EN LÍNEA, EXTRAER EL CUÁNDO Y EL DÓNDE SUCEDIÓ PUEDE AYUDAR A SU ENTENDIMIENTO. LOS METADATOS ASOCIADOS A LOS MENSAJES DE LAS REDES SOCIALES O EL CONTENIDO DE LOS MISMOS, SON USADOS FRECUENTEMENTE PARA ESTE PROPÓSITO. NO OBSTANTE, AUNQUE EL PROCESAMIENTO SEA ENTENDIBLE PARA LOS COMPUTADORES, ES NECESARIO CREAR INTERFACES INTUITIVAS Y QUE SEAN ACCESIBLES PARA LOS HUMANOS, EN PARTICULAR PARA AQUELLOS QUE NO ESTÁN FAMILIARIZADOS CON LA MINERÍA DE DATOS. POR ESTA RAZÓN, EXISTEN HERRAMIENTAS VISUALES QUE ORGANIZAN ESTOS DATOS Y LOS MUESTRAN DE MANERA QUE SU EXPLORACIÓN SEA FÁCIL Y PROVECHOSA. AÚN ASÍ, DADO QUE LA MAYORÍA DE ESTAS HERRAMIENTAS NO ESTÁN DISPONIBLES PARA SU DESCARGA O USO, Y NO TODAS SE ENFOCAN EN EL CONTEXTO DE UN EVENTO DEL MUNDO REAL, AÚN QUEDA MUCHO QUE HACER EN EL ÁREA. ■

## REFERENCIAS

- [1] H. Abdelhaq, C. Sengstock, and M. Gertz, "EventTweet: online localized event detection from twitter", *Proceedings of the VLDB Endowment*, vol. 6, no. 12, pp. 1326–1329, 2013.
- [2] M. Dork, S. Carpendale, C. Collins, and C. Williamson, "Visgets: Coordinated visualizations for web-based information exploration and discovery", *Visualization and Computer Graphics, IEEE Transactions on*, vol. 14, no. 6, pp. 1205–1212, 2008.
- [3] E. Graells-Garrido and B. Poblete, "#Santiago is not #Chile, or is it? A model to normalize social media impact", in *Proceedings of the 2013 Chilean Conference on Human-Computer Interaction*, 2013.
- [4] A. Hannak, E. Anderson, L. F. Barrett, S. Lehmann, A. Mislove, and M. Riedewald, "Tweedin' in the Rain: Exploring Societal-Scale Effects of Weather on Mood", in *ICWSM*, 2012.
- [5] K. Y. Kamath, J. Caverlee, K. Lee, and Z. Cheng, "Spatio-temporal dynamics of online memes: a study of geo-tagged tweets", in *Proceedings of the 22nd International Conference on World Wide Web*, 2013, pp. 667–678.
- [6] K. S. Kim, R. Lee, and K. Zetsu, "mTrend: Discovery of Topic Movements on Geo-Microblogging Messages (Demo Paper)", in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, New York, NY: ACM, 2011.
- [7] A. M. MacEachren, A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford, "Senseplace2: Geotwitter analytics support for situational awareness", in *Visual Analytics Science and Technology (VAST)*, 2011 IEEE Conference on, 2011, pp. 181–190.
- [8] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller, "Twitinfo: aggregating and visualizing microblogs for event exploration", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 227–236.
- [9] M. Mendoza, B. Poblete, and C. Castillo, "Twitter Under Crisis: Can we trust what we RT?", in *Proceedings of the first workshop on social media analytics*, 2010, pp. 71–79.
- [10] A. Ritter, O. Etzioni, S. Clark, and others, "Open domain event extraction from twitter", in *ACM SIGKDD*, 2012.
- [11] R. Rogers. Debanalizing Twitter: The transformation of an object of study. URL [http://www.govcom.org/publications/full\\_list/rogers\\_debanalizingTwitter\\_websci13.pdf](http://www.govcom.org/publications/full_list/rogers_debanalizingTwitter_websci13.pdf).
- [12] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors", in *Proceedings of the 19th international conference on World Wide Web*, 2010, pp. 851–860.



### **MAGÍSTER**

- Tecnologías de la Información (vespertino)

### **DIPLOMAS DE POSTÍTULO**

- Calidad de Software
- Ciencia e Ingeniería de Datos
- Gestión Informática
- Ingeniería de Software
- Ingeniería y Calidad de Software
- Seguridad Computacional
- Tecnologías de Información

### **PROGRAMAS CORPORATIVOS**


- Cursos que se adaptan a las necesidades de capacitación de su empresa

# **PROGRAMA DE EDUCACIÓN CONTINUA**

## DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

Sigue avanzando



 [facebook.com/pec.dcc](https://facebook.com/pec.dcc)

 [capacita@dcc.uchile.cl](mailto:capacita@dcc.uchile.cl)

[www.dcc.uchile.cl/pec](http://www.dcc.uchile.cl/pec)

2 2978 4965





# Bits

DE CIENCIA

[www.dcc.uchile.cl](http://www.dcc.uchile.cl)  
[revista@dcc.uchile.cl](mailto:revista@dcc.uchile.cl)



**fcfm**

Ciencias de la  
Computación  
FACULTAD DE CIENCIAS  
FÍSICAS Y MATEMÁTICAS  
UNIVERSIDAD DE CHILE