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.

VANESSAPEÑA

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

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[1] 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

1 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[2], 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[3] 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**.

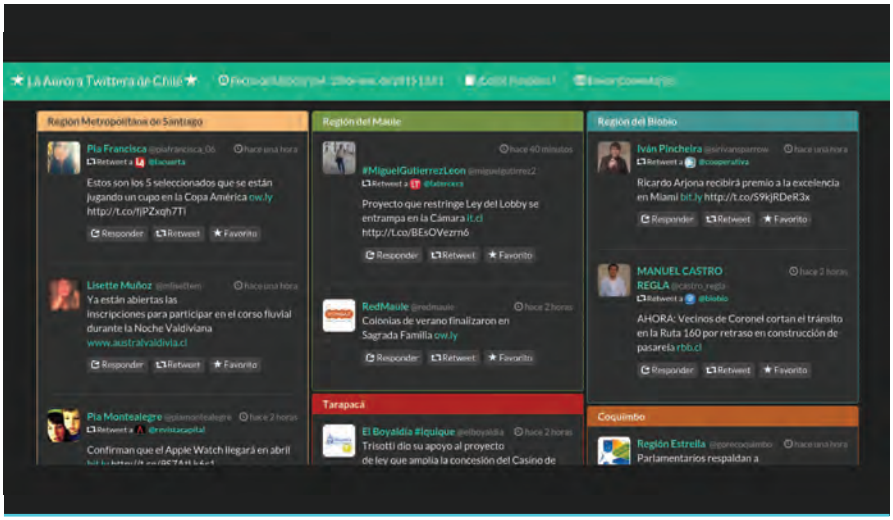

FIGURA 1.
CAPTURA DE PANTALLA DE LA AURORA TWITTERA.

¿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 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

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

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][4] 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[5]. 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.

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

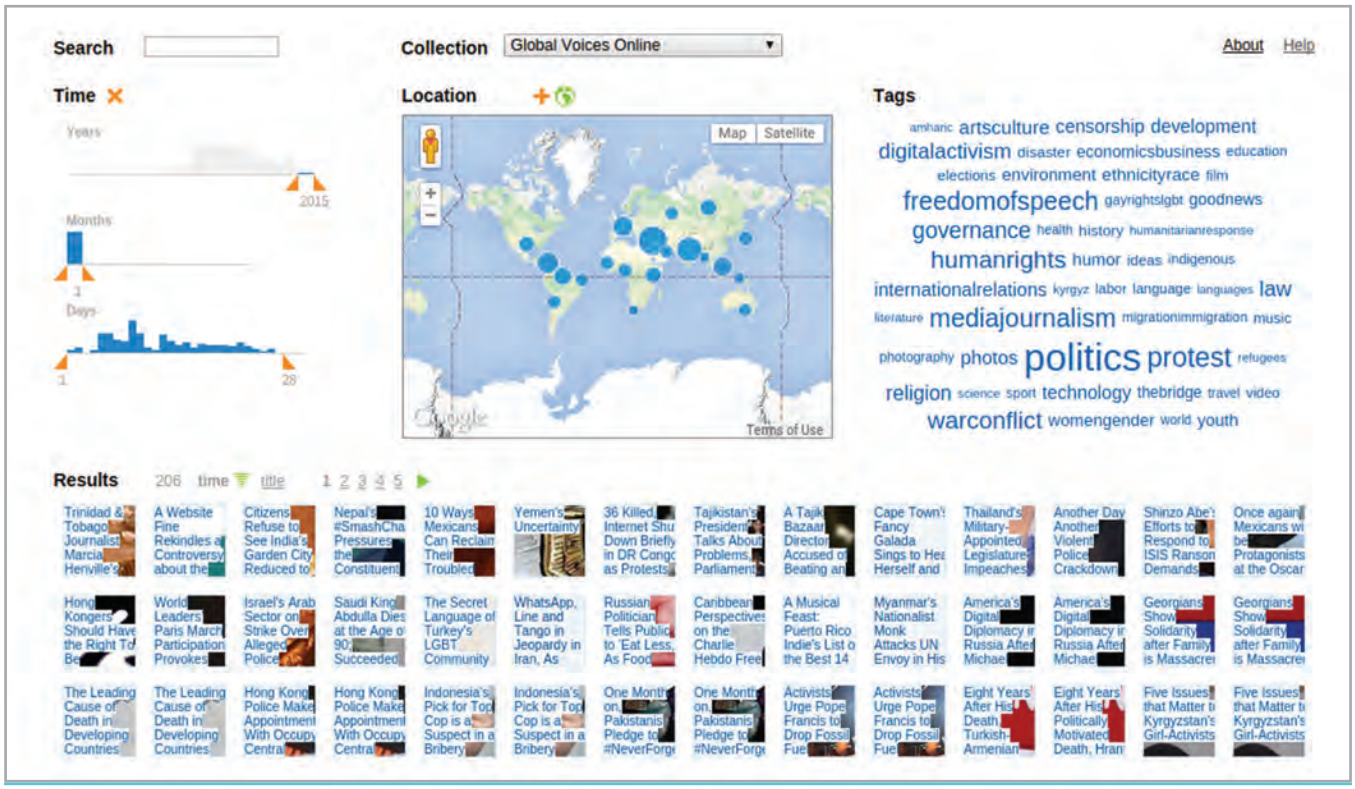

FIGURA 3.
CAPTURA DE PANTALLA DE VISGETS DESDE LA DEMOSTRACIÓN EN LÍNEA.