



Sistemas Abiertos X Sistemas Abiertos S.C., saXsa.

25/10/2016

## Social Media Mining



En la actualidad las redes sociales son una de las principales herramientas de comunicación e información. Siendo conscientes de esta tendencia, en Saxsa nos hemos dado a la tarea de desarrollar tecnologías Big Data enfocadas al Social Media Mining, que se refiere al análisis de datos obtenidos de social media.

Para mostrar lo anterior, se da una breve introducción del tipo de análisis que el equipo Saxsa lleva a cabo en la Red Social Twitter para conocer el impacto que alcanza cualquier tema que se elija, en este caso, la Lucha Libre mexicana.

Dado que los datos en Twitter se generan cada segundo y a nivel mundial, es conveniente crear una ventana de tiempo que recoja datos cada cierto intervalo y así limitar el volumen de archivos que se van recolectando en tiempo real. Esta

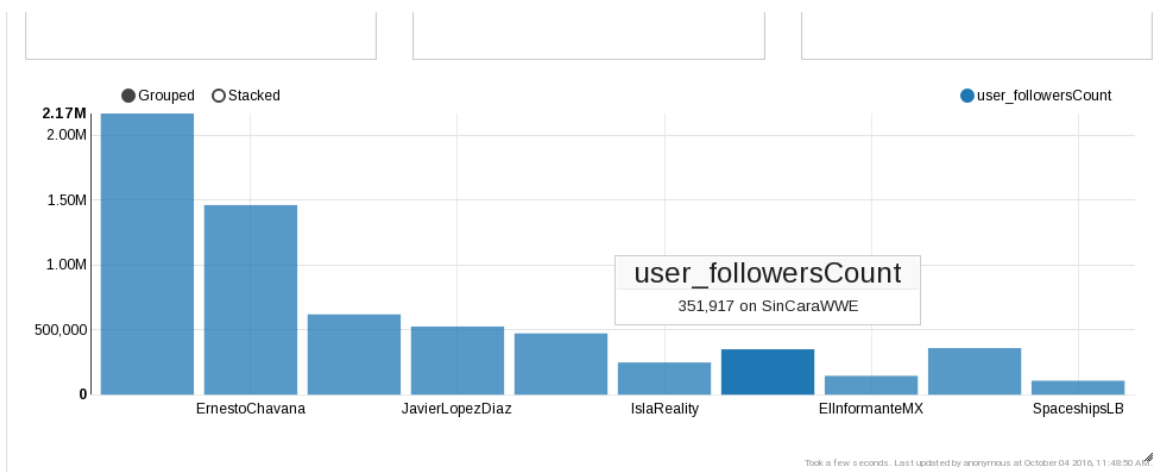
demostración incluye datos recabados durante cinco días a partir del 19 de septiembre por intervalos de tiempo de 1s.

En primer lugar, del total de twits relacionados con Lucha Libre mexicana se hace una relación del número de twits publicados por un solo seguidor. De entre los usuarios de la red que más publicaron sobre el tema se recaba la información relevante.

Se crean gráficas dinámicas sobre los resultados que el sistema va arrojando, es decir, gráficas con información que va cambiando conforme se va actualizando en tiempo real.

La siguiente gráfica contiene en el eje de las abscisas los nombres de cuentas que más twitearon. En el eje de las ordenadas se coloca el número de seguidores de dicha cuenta. Si se coloca el cursor sobre una barra en específico se despliega una etiqueta con el nombre de la cuenta y el dato exacto de seguidores que tiene.

TweetsLuchas+Datos



El siguiente recuadro muestra información que permite conocer más sobre el origen de los twits. Así, se observa que existe la etiqueta “null”, asignada 228 veces, fue dada a aquellos usuarios que al twitear no incluyeron información de su localización. 21 cuentas sólo mostraban que provenían de México, sin especificar zona. 18 usuarios generaron twits desde Tonalá Jalisco, 13 del Distrito Federal, 8 del Distrito Federal pero además especificaron la zona (Cuauhtémoc), 6 desde Leipzig, etc.



## TweetsLuchas+Datos

```
%sql
```

```
select user_location, COUNT(user_location) AS conteo from
tweets4 GROUP BY user_location ORDER BY conteo DESC
```



user_location	conteo
null	228
México	21
TONALÁ, JALISCO	18
Distrito Federal, México	13
México Df	9
Mexico	8
Cuauhtémoc, Distrito Federal	8
Leipzig	6
Ciudad de México	6

El recuadro de abajo también incluye un conteo, pero esta vez en lugar de considerar los usuarios que twitearon en determinada zona, ahora se consideró el número de twits que se generaron en cada área registrada.

Los resultados arrojaron muchas zonas de México desde donde se twiteó pero poca cantidad de twits.

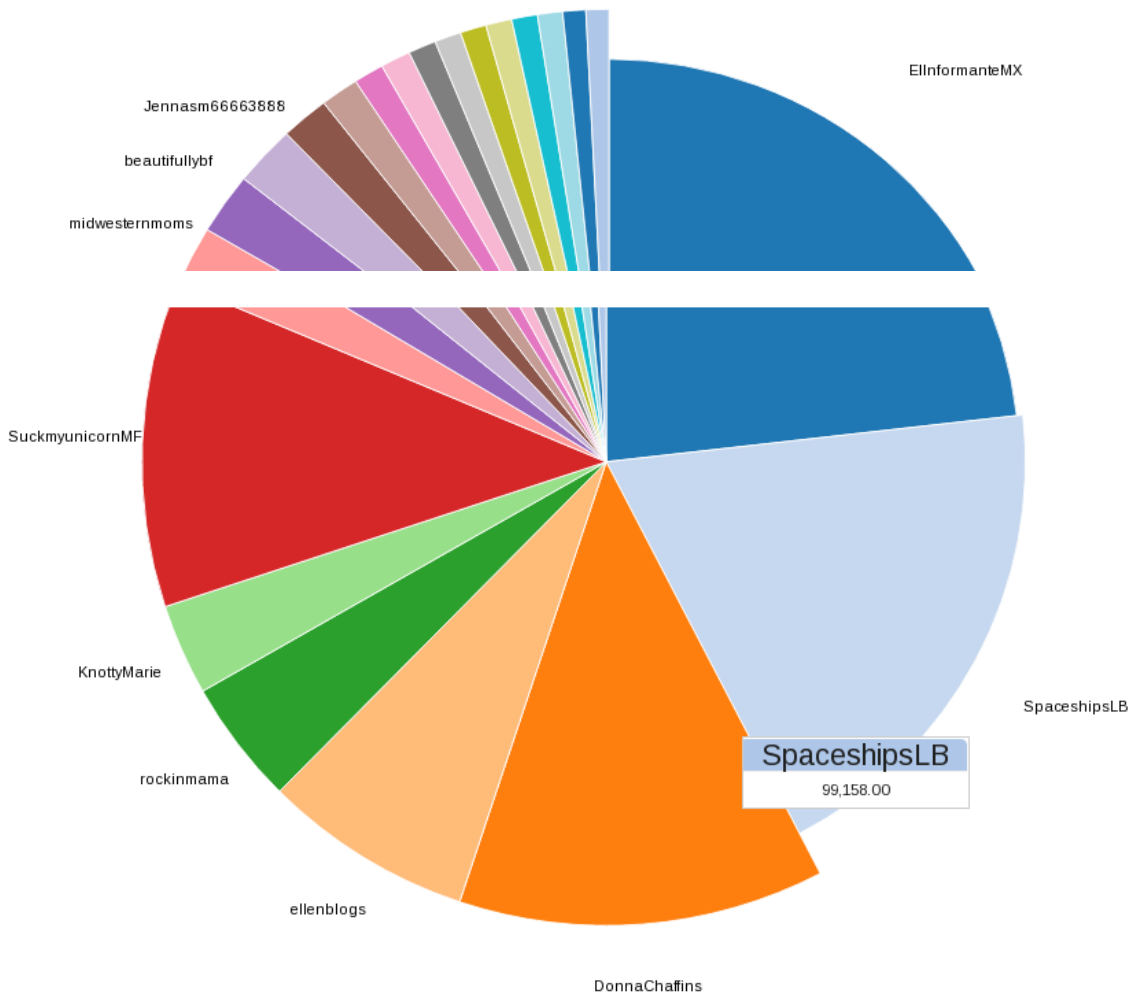


user_location	conteo
Mexico	8
Mexico, CDMX	2
Todo México & Todo Por Mexico	2
Cuernavaca/Morelos- Mexico	2
Chihuahua, Chihuahua Mexico	1
Mexico, Peru	1
Puebla, Mexico	1
Mexico city	1
Monterrey Nuevo León,Mexico	1
Mexico , USA	1
San Diego, CA & Mexico City	1
Mexico City	1
Yucatan, Mexico	1
Monterrey, Mexico	1



El siguiente gráfico muestra cada una de las cuentas que twitearon asuntos relacionados con el tema acotado. El área asignada dentro del gráfico representa los seguidores o followers que cada cuenta tiene. Al colocar el cursor sobre un color muestra la etiqueta con el nombre del usuario y su número de seguidores.

- EInformanteMX
- SpaceshipsLB
- DonnaChaffins
- ellenblogs
- rockinmama
- KnottyMarie
- SuckmyunicornMF
- midwesternmoms
- beautifullybf
- Jennasm66663888
- EmilyReviewsCom
- shellthings
- healingtomato
- womanmanyrole
- JosueMazito
- NFStrangers
- g\_kohls
- MamaMommyMom
- mamasclubhouse
- WritesRomances
- cuzin\_logic
- Giveaways4Mom



En concreto, de un usuario o cuenta de interés se puede obtener información como: Número de follows y número de followers, número de twits que ha publicado, fecha en que se abrió la cuenta, lugar desde el que tuitea (si tiene activada la opción de geolocalización) e incluso análisis de sentimientos, esto es, el tono emocional de sus publicaciones, para ello es necesario hacer un adecuado análisis de las palabras que se asignarán como parámetros para calificar cada



sentimiento, porque una asignación a la ligera puede provocar resultados poco confiables.

Después de conocer un poco más sobre cómo funciona el Social Media Mining, es fácil ver la gran cantidad de aplicaciones que tiene en todos los sectores: Político, empresarial público y privado, comercial, educativo, entre otros.

Conocer la información que arroja este tipo de análisis permite que los productores creen mejores estrategias de venta para sus consumidores, por ejemplo, si el parámetro inicial que se le da al sistema es: Comida Rápida, los inversionistas pueden detectar qué tipo de comida se menciona más, en qué zonas, lo que les permitirá conocer la clase de establecimiento que conviene más colocar en cada zona.

En el caso del Sector Político, con el parámetro adecuado se puede conocer qué lugares sienten cierta preferencia por algún político, partido o política pública y basados en eso crear mejores estrategias de campaña y planes de desarrollo (se sabe que una de las mayores herramientas de campaña de Barack Obama en las elecciones presidenciales fue el adecuado manejo de información en redes sociales).

En el área de salud y educación, se pueden detectar las zonas marginadas o que requieren de mayor atención e incluso conocer el tipo de atención que cada área requiere con más frecuencia. En este sentido existen programas que se encargan de monitorear el movimiento en redes sociales para detectar anticipadamente posibles epidemias

Para mayor información sobre nuestro servicio de Social Media Mining favor de enviar un correo: [info@saxsa.com.mx](mailto:info@saxsa.com.mx)