# Impacto de las Redes Sociales en la Popularidad de las Consultas a Motores de Búsquedas

Santiago Ricci sricci.soft@gmail.com Docente: Prof. Gabriel H. Tolosa tolosoft@unlu.edu.ar

Universidad Nacional de Luján

Resumen Twitter es una red social que cuenta con millones de usuarios alrededor del mundo [7]. Entre sus características más notables se destacan los trending topics o tendencias. Estas son producto de un algoritmo que identifica los temas emergentes más populares <sup>1</sup>. De este modo, cabe preguntarse si el hecho de que cierto tema sea trending topic, influye de algún modo en el volumen de consultas que recibe un motor de búsqueda sobre dicho tema durante el período que es trending topic. Este artículo intenta analizar dicha cuestión, haciendo uso de los datos proporcionados por el servicio Google Trends<sup>2</sup>.

### 1. Introducción

Bajo el título de Tendencias (o trending topics, en inglés), Twitter ofrece un listado de frases, términos y hashtags relacionados con los temas más populares en la red social en un determinado momento. Por defecto, las tendencias se determinan de forma personalizada para cada usuario en base a los usuarios que sigue y a su localización. Sin embargo, un usuario puede optar por obtener trendings topics de forma no personalizada, en base a su región geográfica.

Algunos estudios como [7], sugieren que gran parte de los trendings topics estan relacionados con las noticias del momento. Por otro lado, se sabe que el buscador de Twitter es usado para monitorear cierto contenido, mientras que los motores de búsqueda Web son empleados para saber más acerca de dicho tema y que muchos usuarios ejecutan la misma consulta tanto en el motor de búsqueda de Twitter como en uno Web, con el fin de capturar ambos usos [9]. También, es conocido el uso de Twitter para expresar opiniones acerca de diferentes temas, lo cual se ha traducido en gran cantidad de artículos que plantean diferentes enfoques sobre cómo realizar Minado de Opinión sobre la red social, como por ejemplo, [12] o [13]. Estas premisas, junto a que según [7], gran parte de los usuarios de Twitter participan en trendings topics, en conjunto permiten plantear la siguiente hipótesis: "el hecho de que un tema sea trending topic en un determinado momento, se traducirá en el aumento de la popularidad

<sup>&</sup>lt;sup>1</sup> https://support.twitter.com/articles/101125-about-trending-topics

<sup>&</sup>lt;sup>2</sup> http://www.google.com/trends/

de consultas relacionadas con dicho tema en los motores de búsqueda Web". Este trabajo se centra en verificar la validez de dicha afirmación. Esta cuestión es importante debido a que los motores de búsqueda deben responder millones de consultas (queries) por día, lo cual implica eficiencia y efectividad para poder otorgar a los usuarios respuestas relevantes lo más rápido posible [10], entonces, si existe tal relación, puede sacarse provecho de la misma para mejorar el rendimiento.

# 2. Trabajos relacionados

No es de conocimiento del autor otros trabajos que traten la hipótesis planteada. Sin embargo, en [7] se realiza un estudio acerca de las características de los trendings topics y se comparan las búsquedas populares presentadas por Google Trends con los trendings topics presentando el grado de solapamiento entre ambos, el cual se encontró que es bajo. También se estudió la diferencia en la "frescura" (freshness) de los temas en ambos y se halló que en Twitter son más persistentes. Además, se concluye que los usuarios de Twitter tienden a hablar sobre noticias y que gran parte (31 %) de los trendings topics duran aproximadamente un día. En [2], se estudian los trendings topics y se afirma que aquellos con grandes duraciones están caracterizados por la naturaleza "resonante" del contenido de sus tweets asociados, el cual proviene, generalmente, de los medios de comunicación tradicionales. De este modo, Twitter se comporta como un amplificador selectivo del contenido generado por los medios tradicionales mediante cadenas de retweets. En [9] se compara la tarea de búsqueda de los usuarios en Twitter respecto a los motores de búsqueda y se concluye, en base al uso de los logs de la barra de Bing, que los usuarios utilizan Twitter para monitorear un tema y la Web para aprender más acerca del mismo. Además se compara los queries formulados en el motor de búsqueda de Twitter con los realizados sobre dicho motor de búsqueda Web. En [1], se utiliza a los trendings topics como base para predecir los temas que se volverán populares en el futuro cercano. En [4] se estudia al servicio desde el punto de vista estructural y del contenido. Una publicación posterior [6] amplía dicha caracterización.

# 3. Metodología

El enfoque general tomado para verificar la hipótesis fue capturar los trending topics durante una semana, derivar consultas a partir de los mismos y evaluar la evolución de la popularidad de dicha consulta mediante el uso de los datos de Google Trends. Para ello, se utilizaron una serie de métricas. En la sección siguiente, se detallan cada una de ellas.

### 3.1. Métricas

Se utilizaron tres métricas a tener en cuenta para evaluar los cambios en la popularidad de las consultas.

1. Variación del interés (Var): es la variación porcentual del interés de la consulta derivada para un  $trendings\ topic$ . Formalmente, se define del siguiente modo: sea  $I(n,q_{t_i})$  el interés en el query q asociado al  $trendings\ topic\ t_i$  en el día n (día en que el tema se convierte en  $trendings\ topic$ ), entonces la variación porcentual respecto al día anterior (n-1) se define como:

$$Var(n, q_{t_i}) = 100 \frac{I(n, q_{t_i}) - I(n - 1, q_{t_i})}{I(n - 1, q_{t_i})}$$
(1)

En el caso que  $I(n,q_{t_i})=0$  y  $I(n-1,q_{t_i})=0$ ,  $Var(n-1,q_{t_i})=0$ ; y para el caso que  $I(n,q_{t_i})\neq 0$  y  $I(n-1,q_{t_i})=0$ ,  $Var(n,q_{t_i})=100I(n-1,q_{t_i})$ . Esta métrica, intenta capturar el hecho de que si los  $trendings\ topics$  influyen en el volumen de consultas, entonces debe existir una gran diferencia en el interés en dicha consulta entre cuando el tema es  $trendings\ topic$  y cuando no.

2. Cambio de tendencia (B): es la cuantificación del cambio en la tendencia que experimenta cierta consulta derivada cuando el tema se convierte en trending topic, respecto a su tendencia en los siete días anteriores. Formalmente, si q es la consulta derivada para el trendings topic  $t_i$  en el día n,  $m_{q_{t_i}}(x,y)$  es la pendiente de la línea de tendencia para la consulta q entre los días x e y, entonces el cambio en la tendencia se define como:

$$T(q_{t_i}) = \frac{m_{q_{t_i}}(n-6,n) - m_{q_{t_i}}(n-7,n-1)}{|m_{q_{t_i}}(n-7,n-1)|}$$
(2)

En consecuencia, debe cumplirse que para una consulta que viene experimentando interés creciente,  $T(q_{t_i})$  sea lo suficientemente grande para poder afirmar que el cambio fue consecuencia de que el tema sea trendings topic.

3. **Detección de picos:** es la aplicación de un algoritmo de detección de picos (burst detection) presentado en [11]. El mismo plantea que, para descubrir regiones con valores 'pico' en una serie temporal, debe calcularse la media móvil (MA) y tomar como valores pico aquellos que superen x desvíos estándar sobre el valor medio de MA. Por lo tanto, aquí intenta plasmarse que el día que un tema es trending topic, debería registrarse un pico en el interés en ese tema en los motores de búsqueda.

# 3.2. Obtención de los trendings topics

Empleando la API REST de Twitter $^3$ , se elaboró una aplicación que capturó los  $trendings\ topics$  para Argentina (no personalizados) desde el 05/12/2012 16:00 hs. hasta el 12/12/2012 16:00 hs. a intervalos de 5 minutos aproximadamente. Dicha frecuencia no fue elegida caprichosamente, sino que es la mínima posible, ya que la información retornada es mantenida en caché por el servidor por 5 minutos. También es importante mencionar que la frecuencia es aproximada como consecuencia de los retardos de comunicación. Bajo dichas condiciones,

<sup>&</sup>lt;sup>3</sup> https://dev.twitter.com/docs/api

se obtuvieron 2002 muestras cada una de 10 trending topics. Eliminando los duplicados en todo el período de captura, se obtuvo un total de 573 trending topics y eliminando solo las repeticiones del mismo día se obtuvo un total de 727 trending topics.

#### 3.3. Proceso de derivación de consultas

Este aspecto es muy importante en lo que respecta a los objetivos del trabajo, pudiendo ser determinante en los resultados, dado que para un trendings topic se pueden derivar múltiples consultas. Además, el trending topic en sí, no siempre se corresponde con la consulta que el usuario ejecutaría sobre el motor de búsqueda y en otros casos, puede no quedar claro qué consulta se puede derivar a partir del trendings topic. También puede ser dfícil extraer los términos en el caso de las hashtags, o los trendings topics pueden derivarse en consultas ambiguas (podría emplearse sus tweets correspondientes para intentar desambiguar). También es importante tener en cuenta que, según [9], los queries realizados sobre el motor de búsuqeda de Twitter difieren en longitud, en la función y en la sintaxis (ambos son free-text queries, pero en Twitter, '#' y '@' tienen significados especiales), con lo cual parece no ser del todo adecuado emplear la consulta asociada a cada trending topic que devuelve la API de Twitter. Un dato importante también aportado en dicha publicación, es que gran parte (45.95%) del conjunto intersección entre los queries formulados en Twitter y en la barra de búsuqeda de Bing son informacionales y acerca de celebridades.

Teniendo en cuenta las cuestiones anteriores, se tomó el siguiente criterio para derivar las consultas: si el trendings topic es una frase o un término, entonces la consulta es dicha frase o término. En el caso que sea un hashtag, se elimina el # y se intenta separar los términos haciendo uso de las mayúsculas y los números. Si no se dispone de mayúsculas, se utiliza la primer sugerencia de Google que normalmente obtienen los usuarios a medida que van escribiendo su consulta. Si no existen sugerencias, se ejecuta el trending topic en Google y se toma la búsqueda sugerida por Google (la que figura bajo la denominación: "Se muestran resultados de"). En caso de no existir tal sugerencia, se toma como consulta el hashtag (sin el #).

#### 3.4. Obtención de datos de Google Trends

Para analizar los cambios en la popularidad de las consultas derivadas, se requiere alguna fuente de información que disponga de los datos correspondientes a los queries derivados de los trending topics. Dado el difícil acceso a log de consultas de un buscador Web, se optó por utilizar los datos del servicio Google Trends. Este servicio permite ejecutar una consulta y obtener la evolución del interés a lo largo del tiempo fijando una ventana temporal. Si bien la documentación<sup>4</sup> de la herramienta no especifica el método exacto para calcular los valores que devuelve, se sabe que los datos reflejan el número de búsquedas de

<sup>&</sup>lt;sup>4</sup> http://support.google.com/trends/?hl=es#topic=13762

un término en comparación al total de búsquedas realizadas en Google a lo largo del tiempo<sup>5</sup>. Además, los datos están normalizados y son presentados en una escala de 0 a 100, donde el valor 0 se corresponde a la falta de datos (el servicio solo muestra los datos de los términos que sobrepasen un cierto límite en los volúmenes de búsqueda).

Habiendo aclarado las características de la herramienta, se prosigue a describir el uso de la misma. Para cada query derivado de cada trending topic del conjunto de  $trendings\ topics$  de cada día, se descargó el archivo CSV correspondiente a la consulta, empleando una ventana temporal de 30 días y el filtro de datos para Argetina. Dado que la captura de dichos datos se realizó el 16/12/2012, el período abarcado por los datos comienza el 17/11/2012 y se extiende hasta el 15/12/2012. Del total de  $trendings\ topics$  (eliminando las repeticiones del mismo día) se pudieron obtener datos de Google Trends en el 43 % de los casos (el servicio devuelve 29 puntos, uno por día).

# 4. Experimentos & Resultados

Esta sección tiene por objetivo presentar los experimentos y los resultados obtenidos.

### 4.1. Variación del Interés

Para cada trendings topic se calculó la primer métrica, determinada por la Ecuación 1. Luego, se calculó la cantidad de consultas que tuvieron interés creciente, dececiente y constante. El Cuadro 1 resume los resultados obtenidos. De allí, es importante destacar, que aproximadamente el 72 % del total de los trending topics para los que se cuenta con 29 puntos obtuvo un interés creciente, por lo tanto se hace necesario cuantificar dicho aumento con el fin de verificar que sea significante.

Cuadro 1. Variación en el interés respecto al día anterior

Tipo de variación	Cantidad	% relativo (1)	
Positivo	226		71,97 %
Negativo	61		19,43 %
Nulo	27		8,60 %
(1) Respecto a la cantidad de trending topics con 29 puntos (314).			

El Cuadro 2 resume la cantidad de  $trendings\ topics$  que alcanzan un porcentaje de crecimiento respecto al día anterior mayor al indicado en la columna '% de crecimiento' y menor o igual al de la fila siguiente. La Figura 1, muestra cómo

<sup>&</sup>lt;sup>5</sup> http://support.google.com/trends/answer/87285?hl=es&ref\_topic=13975

Cuadro 2. Detalle del aumento del interés respecto al día anterior

% de crecimiento	Cantidad de Trending Topics	% relativo (1)	% colección (2)		
10	31	13,72 %	9,87 %		
20	23	10,18 %	7,32 %		
30	30	13,27 %	9,55 %		
40	15	6,64 %	4,78 %		
50	10	4,42 %	3,18 %		
60	13	5,75 %	4,14 %		
70	11	4,87 %	3,50 %		
80	13	5,75 %	4,14 %		
90	1	0,44 %	0,32 %		
100	13	5,75 %	4,14 %		
200	23	10,18 %	7,32 %		
300	10	4,42 %	3,18 %		
400	5	2,21 %	1,59 %		
500	8	3,54 %	2,55 %		
600	1	0,44 %	0,32 %		
700	3	1,33 %	0,96 %		
800	1	0,44 %	0,32 %		
900	1	0,44 %	0,32 %		
>900	14	6,19 %	4,46 %		
(1) Respecto a la cantidad de trending topics con interés respecto al día anterior creciente (226).					
(2) Respecto a la cantidad de trending topics con 29 puntos (314).					

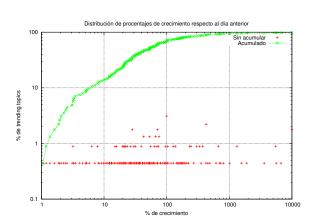


Figura 1. Distribución de variación en el interés respecto al día anterior

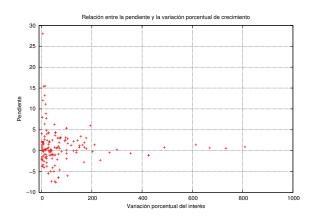
se distribuyen los aumentos. En la misma, puede observarse que alrededor del  $35\,\%$  del total de trending topics con interés respecto al día anterior creciente, obtuvo un aumento de más del  $100\,\%$ . Sin embargo, pese a que un buen porcentaje de los  $trendings\ topics$  obtiene un aumento significativo, hay que observar la tendencia de la consulta en una ventana temporal mayor, dado que si las consultas asociadas a los trending topics venían experimentando una tendencia creciente pronunciada, el hecho de que el tema sea tendencia no posee un gran efecto sobre la popularidad en la consulta.

#### 4.2. Cambio de tendencia de las consultas

Para cada consulta derivada se calculó la línea de tendencia de los 7 días anteriores a que el tema fuera trendings topic. En el Cuadro 3 puede observarse que alrededor del 61.46% de las consultas posee tendencia creciente respecto al total de trendings topics para los que se dispone 29 puntos. Esto plantea dos cuestiones: por un lado, pueden existir casos en los que la consulta derivada venía experimentando una tendencia decreciente, pero el interés aumentó al convertirse su tema asociado en trending topic; por el otro, al menos que el cambio en la popularidad sea lo "suficientemente grande" no puede considerarse como válida la hipótesis. Por consiguiente, se elaboró el Gráfico 2 que relaciona la métrica dada por la Ecuación 1 con la pendiente de la línea de tendencia. Nótese que los valores de la métrica dada por la Ecuación 1 mayores a 1000 fueron ignorados. Idealmente, la nube de puntos debería estar ubicada cerca del eje de las absisas y alejada del eje de las ordenadas. Obsérvese, que en la figura, la misma se encuentra centrada en torno a una variación del interés de aproximadamente un  $80\,\%$  y que la mayor densidad de puntos se observa en torno a las pendientes con valores entre 5 y -5. Esto muestra entonces, que existen casos donde se observa una tendencia decreciente en el interés y el día en que el tema se convirtió en trending topic, el interés creció abruptamente. Observando especificamente las pendientes positivas, se encuentran tanto, casos en los cuales la variación porcentual de interés dada por la primer métrica es grande en comparación al crecimiento que debería esperarse dada la pendiente de la línea de tendencia y visceversa. Es importante aclarar, que la tendencia que experimenta una consulta en los últimos 7 días no se corresponde necesariamente con la tendencia considerando una ventana temporal mayor.

Ahora, se pasará a mostrar los resultados obtenidos a partir del cálculo de la segunda métrica. El Cuadro 4 muestra un resumen de los mismos. A partir del mismo, puede deducirse que aproximadamente en el  $40\,\%$  de los casos, el crecimiento de la pendiente es de, al menos dos veces la pendiente de la recta que no considera el día en el que el tema fue trending topic. De nuevo, la ventana temporal empleada puede influír en los resultados.

Si bien, la figura no permite afirmar que la hipótesis se cumple para todos los casos, muestra que existen casos para los que pareciera que sí se cumple. Por otro lado, el análisis del cambio en la pendiente plasmado por la segunda métrica parece reforzar lo dicho previamente.



 ${\bf Figura~2.}$ Relación entre la pendiente de la línea de tendencia y la métrica determinada por la Ecuación 1

Cuadro 3. Tendencias de los 7 días previos a que un tema sea trending topic

Crecimiento	Cantidad de trendings topics	% relativo (1)		
Creciente	193	61,46 %		
Decreciente	109	34,71 %		
Estacionario	12	3,82 %		
(1) Respecto a la cantidad de trendings topics con 29 puntos (314).				

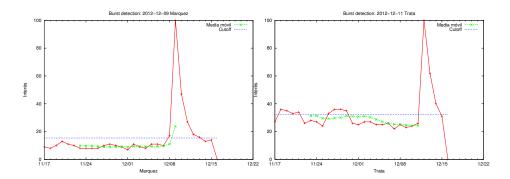
Cuadro 4. Relación entre pendientes

Valor de la relación	Cantidad de trendings topics	% relativo (1)	
< -0.5	12	5,31 %	
< 0	9	3,98 %	
< 0.5	26	11,50 %	
< 1	35	15,49 %	
< 1.5	29	12,83 %	
< 2	24	10,62%	
< 2.5	11	4,87 %	
< 3	9	3.98%	
< 5	21	9,29%	
< 10	22	9.73%	
>= 10	28	12,39 %	
(1) Respecto a la cantidad de trendings topicss cuya variación fue positiva (226).			

# 4.3. Detección de picos

Por último se ejecutó el algoritmo de detección de picos, utilizando la configuración recomendada en [11], es decir, una ventana de 7 días y x=1,5. Del total de trending topics para los que se contaba 29 puntos, 162 (51,6%) tuvieron un pico en el interés en el día que fueron  $trending\ topic$ . Esto parece reforzar la validez de la hipótesis. La Figura 3 muestra el caso del 'Marquez', del cual se derivó el query 'Marquez' el cual tuvo un pico en el interés el día que fue  $trending\ topic$ . La Figura 4 muestra un caso donde no se identificó un pico de

interés en la consulta. En ambas gráficas puede observarse la evolución de la media móvil hasta el día en que el tema fue trending topic y el umbral (cutoff) para considerar un pico en el interés.



**Figura 3.** Trending topic cuyo query obtu-**Figura 4.** Trending topic cuyo query no obvo un pico en el interés tuvo un pico en el interés

Es curioso observar que, en el caso de la Figura 4, el pico se produce un día después que el tema fue trending topic. En realidad, el tema apareció como trendings topic aproximadamente a las 22 hs. (22:09:38 es la hora exacta en la que aparece el primer registro), con lo cual se puede suponer que gran parte del tráfico va a estar contenido en el día siguiente. Esto sugiere que debe estudiarse la duración de los trendings topics. La Figura 5 muestra el porcentaje de trendings topics con una duración menor o igual a la especificada en el eje 'Duración' contando todas sus apariciones. Allí puede observarse que aproximadamente el 55% de los trendings topics (porcentaje calculado sobre el total de trending topics, sin contar las repeticiones) posee una duración de como máximo una hora y que solo unos pocos superan el día de duración. En [7] se ofrece un estudio similar, considerando una ventana temporal mayor. Viendo estos resultados, se hace necesario estudiar el aumento del tráfico en relación a la duración del trendings topic, cuestión que se deja abierta para trabajos futuros. Esto es importante, dado que debe verificarse que en la extensión de un trendings topic en el tiempo, se produzca suficiente tráfico como para alterar de forma abrupta la tendencia que venía experimentando la consulta en cuanto al interés. También debe estudiarse qué sucede luego de finalizado el período en el cuál el tema es trendings topic.

# 5. Conclusiones y trabajos futuros

El presente trabajo ha presentado indicios, mediante el uso de tres métricas, de que el hecho que un tema sea trendings topic en un determinado momento puede traducirse en el aumento de la popularidad de las consultas asociadas a dicho tema en dicho momento. Tanto la métrica 2, como la 3, han mostrado casos en los que la hipótesis parece cumplirse. Es muy probable que sea verdadera

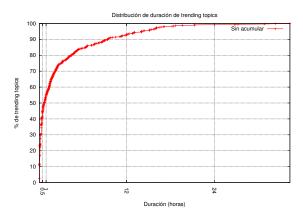


Figura 5. Duración de los trendings topics

para aquellos trendings topics referidos a noticias del momento. Esta hipótesis, se propone para trabajos futuros. También se ha planteado que el proceso de derivación de consultas a partir de los trending topics no es simple y que posee una estrecha relación con la validéz de la hipótesis discutida en el presente artículo. Lo mismo sucede con el período abarcado en el estudio de las tendencias. Además, se dejan abierta las siguientes cuestiones para trabajos futuros:

- Caracterizar el volumen de tráfico de un motor de búsqueda mediante métodos indirectos: como esta información es propietaria de los proveedores de los servicios de búsqueda (se los caracteriza como "ambientes no cooperativos" [8]), se requieren métodos "externos" que permitan obtener aproximaciones válidas. Este problema ha sido abordado en el área de Recuperación de Información Distribuida para obtener descripciones de los recursos objetivo (por ejemplo, Query Based Sampling [3]).
- 2. Refinar los métodos de derivación de consultas a partir de los *trendings topics* y verificar su impacto en los resultados.
- 3. Estudiar qué sucede con el interés luego de que el tema deja de ser trending topic.
- 4. Investigar el uso de los *trendings topics* y sus *tweets* asociados como información para desambiguar las consultas.
- 5. Investigar el uso de los *trendings topics* como un factor a tener en cuenta en las políticas de admisión de las cachés de los motores de búsqueda Web e el prefetching de resultados [5].

# Referencias

 Agarwal, S.; Agarwal, S. Social networks as Internet barometers for optimizing content delivery networks. In 3rd International Symposium on Advanced Networks and Telecommunication Systems (ANTS). 2009.

- Asur, S.; Huberman, B. A.; Szabo, G.; Wang, C. Trends in social media: Persistence and decay. In 5th International AAAI Conference on Weblogs and Social Media. 2011.
- 3. Callan, J.; Connel, M. Query-based sampling of text databases. In ACM Transactions on Information Systems, v. 19, n. 2, pp. 97-130. 2001.
- Java, A.; Song, X.; Finin, T.; Tseng B. Why we twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (WebKDD/SNA-KDD '07), pp. 56-65. 200
- 5. Jonassen, S.; Barla Cambazoglu B.; Silvestri F. Prefetching query results and its impact on search engines. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (SIGIR '12), pp. 631-640. 2012.
- Krishnamurthy, B.; Gill, P.; Arlitt M. A few chirps about twitter. In Proceedings of the first workshop on Online social networks (WOSN '08), pp. 19-24. 2008.
- Kwak, H.; Lee, C.; Park, H.; Moon, S. What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web (WWW '10), pp. 591-600. 2010.
- Luo S. Federated search of text search engines in uncooperative environments. PhD Thesis. Carnegie Mellon University. 2006.
- 9. Teevan, J.; Ramage, D.; Ringel Morris, M. #TwitterSearch: a comparison of microblog search and web search. In Proceedings of the fourth ACM international conference on Web search and data mining (WSDM '11), pp. 35-44. 2011.
- Tolosa G. H.; Feuerstein E. Mejoras algorítmicas y estructuras de datos para búsquedas altamente eficientes. En Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación (WICC 2012), p. 740-744. 978-950-766-082-5. 2012.
- Vlachos, M.; Meek, C.; Vagena, Z.; Gunopulos, D. Identifying similarities, periodicities and bursts for online search queries. In Proceedings of the 2004 ACM SIGMOD international conference on Management of data (SIGMOD '04), pp. 131-142. 2004.
- 12. Wang, X.; Wei, F.; Liu, X.; Zhou, M.; Zhang, M. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In Proceedings of the 20th ACM international conference on Information and knowledge management (CIKM '11) pp. 1031-1040. 2011.
- Xinfan M.; Furu W.; Xiaohua L.; Ming Z.; Sujian L.; Houfeng W. Entity-centric topic-oriented opinion summarization in twitter. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12), pp. 379-387. 2012.