

Inteligencia colectiva: de la comunicación entre bacterias a la difusión de *cibermemes*

Víctor Márquez Gil
*Escuela Politécnica Superior,
Universidad Autónoma de Madrid.*
victor.marquez@estudiante.uam.es
Tutora: Ruth Cobos Pérez

Resumen. En este trabajo se presenta la idea de inteligencia colectiva y la inspiración biológica de la misma, así como su redefinición con la llegada de las aplicaciones sociales y la Web 2.0. Se documenta el estado del arte de los trabajos relacionados con la realización de un caso de estudio del activismo social y ecológico en la plataforma Twitter y como estos movimientos surgidos de la red tienen un impacto en las sociedades y en el comportamiento humano. Finalmente se presenta un prototipo de aplicación para acceder a las tendencias de pensamiento de dicha inteligencia colectiva que surge por una colaboración implícita. Este prototipo podría tener aplicación en diferentes áreas de estudio, en el trabajo se presenta un ejemplo relacionado con el activismo en Twitter.

Índice

1. Introducción	3
1.1. Motivación	3
1.2. Objetivos	3
1.3. Campos relacionados	3
1.4. Inspiración biológica de la idea de inteligencia colectiva	4
1.4.1. Quorum Sensing en bacterias	4
1.4.2. Selección de semillas al germinar	4
1.4.3. Memética: los mensajes en sistemas de microblogging como <i>cibermemes</i>	4
2. Estado del arte	5
2.1. Inteligencia colectiva	5
2.1.1. Twitter	5
2.1.2. Earth Hour y activismo en redes sociales en línea	8
2.2. Expansión de consultas	9
3. Materiales y métodos	10
4. Consulta y resultados	11
4.1. Consulta	11
4.2. Resultados	12
4.3. Discusión	12
5. Conclusiones	12
6. Trabajo futuro	13
7. Referencias bibliográficas	14

1 Introducción

1.1 Motivación

En los orígenes de la World Wide Web, la publicación de páginas no era algo trivial, requería de conocimientos de HTML, un servidor Web y un mecanismo para cargar los ficheros en este como FTP, sin hablar de páginas dinámicas y los lenguajes de scripting del lado del servidor. Algo más tarde surgieron los editores WYSIWYG¹ que facilitaron el proceso ya que no se requerían conocimientos de HTML. Con el auge de la Web 2.0, la publicación de contenidos en Internet se ha convertido en algo sumamente fácil, los editores WYSIWYG están embebidos en las herramientas sociales y ya el proceso de carga de información al servidor se hace totalmente transparente para el usuario final mediante el uso de un botón con una etiqueta similar a “Publicar”.

Esta evolución ha favorecido que ingentes masas de usuarios se den de alta en redes sociales en línea (o virtuales), sitios Web de revisión de productos y servicios o creen su blog y publiquen diariamente sus pensamientos, sus opiniones, sus sentimientos, noticias, enlaces, etc. Este contenido es rico en expresión de opinión y sentimiento. Conocer la opinión de los demás es importante para la toma de decisiones, pero el volumen de información es imposible procesarlo manualmente. Se hace por tanto indispensable el tratamiento automático para solucionar el problema de la sobrecarga. Pero si echamos un vistazo global no todo son problemas y parece que hay propiedades emergentes de esta conexión de personas y ordenadores que tienen impacto en las sociedades y el comportamiento humano.

1.2 Objetivos

En primera instancia se pretende dar a conocer cómo podría tratarse el problema de dicha sobrecarga. También se intenta arrojar algo de luz sobre cuáles son las tendencias de pensamiento colectivo, para este fin se ha acotado el problema a la red social Twitter ([Twitter Inc., 2010](#)) y a la lengua inglesa. Por último, se documentará el impacto de los medios sociales virtuales en el mundo real y sus sociedades, de lo que se desprende el concepto de inteligencia colectiva.

1.3 Campos relacionados

Todos son objetivos con varias facetas y para acercarnos a su solución del problema entran en juego diversos campos de investigación. Se han identificado los siguientes:

Sistemas colaborativos: estudio de las herramientas y técnicas que pueden ayudar a los usuarios a realizar tareas o actividades conjuntamente en grupo. En este trabajo se estudian los patrones de pensamiento colectivo en la red social Twitter ([Twitter Inc., 2010](#)) estableciendo la relación con la asignatura.

Recuperación de información: ciencia que estudia la búsqueda de documentos, de información dentro de los documentos y de los metadatos de los mismos. En nuestro caso la aplicación de algoritmos automáticos de búsqueda se aplica a la extracción de los mensajes de Twitter relacionados con algún tema en concreto.

Procesamiento del lenguaje natural: rama de la inteligencia artificial que hace referencia a

1 WYSIWYG: De las siglas en inglés “What You See Is What You Get”, es decir, lo que ves es lo que obtienes. Son editores de texto cuyo resultado final es el que el usuario ve en pantalla al editar, ya que se genera el código HTML automáticamente atendiendo a los comandos que el usuario invoca en una interfaz gráfica similar a la de los procesadores de texto.

las técnicas de tratamiento del lenguaje y su aplicación en diversas áreas por medio de métodos computacionales. En el trabajo se hace uso de la red semántica WordNet para la expansión de las consultas al corpus de Twitter.

1.4 Inspiración biológica de la idea de inteligencia colectiva

1.4.1 *Quorum Sensing* en bacterias

Hasta hace relativamente poco tiempo era un dilema si las bacterias eran o no organismos sociales. Dicho dilema se resolvió al descubrir el *Quorum Sensing*. Se trata de un proceso de comunicación bacteriana que consiste en la liberación en el medio de moléculas autoinductoras, cuando la densidad de dichas moléculas alcanza un umbral (o *Quorum*), se desencadena un evento dependiente de cada especie de bacterias (Prescott *et al.*, 2002). Estos eventos pueden ir desde la bioluminiscencia a la regulación del tamaño poblacional.

1.4.2 Selección de semillas al germinar

En algunas especies de plantas, las semillas liberan en el medio una sustancia. Con altas concentraciones de dicha sustancia actúa como una toxina e inhibe la germinación de parte de los individuos. Este proceso se denomina inhibición interespecífica de la germinación (Robinson *et al.*, 2003). De esta manera las semillas se comunican de cierta manera para que no se consuma energía en que geminen y comiencen a crecer plantas que están condenadas a morir por falta de luz ya que las vecinas de la misma especie se harían sombra entre sí. En este caso, el evento que se desencadena es la inhibición de la germinación cuando la concentración de la sustancia en el medio alcanza un determinado umbral.

1.4.3 Memética: los mensajes en sistemas de microblogging como *cibermemes*

La memética es el estudio de la evolución de la transmisión y evolución de información cultural basado en el concepto de meme. El término *meme* fue acuñado por Richard Dawkins en 1976 en su libro *El gen egoísta* (Dawkins, 1976). Se define el meme como una unidad de evolución cultural estableciendo el paralelismo con los genes como unidades de evolución genética. Los memes pueden ser cualquier pieza de información que se puede transmitir de persona a persona y se anclan en la memoria de la humanidad –o cultura–, como por ejemplo: noticias, canciones y sus estribillos, refranes, hábitos, gestos, aforismos, creencias religiosas, chistes, leyendas urbanas o historias. Lejos de caer en el olvido, este acercamiento ha sido refinado y formalizado con el transcurso del tiempo.

Ahora cobra especial importancia con la explosión de contenido de la Web 2.0, y surgen nuevos tipos de memes que se transmiten por internet con forma de enlaces a sitios web, animaciones Flash, publicaciones en blogs, o aplicaciones, por poner algunos ejemplos. En este trabajo nos centraremos en el uso de sistema de microblogging (en concreto Twitter) como medio de transmisión de estos *cibermemes*.

Para enlazar con la inspiración biológica, además de establecer el paralelismos entre memes y genes, se pueden ver los *cibermemes* como la sustancia que se libera en el medio (las redes sociales y en la Web en general). Más adelante se discutirá cuáles pueden ser los eventos que se desencadenan en el mundo real cuando existe gran densidad de estos *cibermemes* en Internet.

2 Estado del arte

2.1 Inteligencia colectiva

“¿Cómo pueden conectarse las personas y las computadoras de manera que, colectivamente, actúen de manera más inteligente que los individuos, los grupos o las computadoras lo hayan hecho antes?” Esta es la pregunta de base que se formulan en el Centro de Inteligencia Colectiva del Instituto de Tecnología de Massachusetts².

La idea de inteligencia colectiva extiende el concepto de sinergia de grupo. La sinergia surge cuando el resultado de una acción grupal es superior al obtenido de la suma de los resultados de los individuos del grupo, en este caso el objetivo del grupo es común y el resultado suele obtenerse por colaboración. La inteligencia colectiva no requiere necesariamente de colaboración, aunque la sinergia es un caso particular de inteligencia colectiva. Como ejemplo de inteligencia colectiva, en el caso de tener que localizar un submarino perdido, se obtienen resultados más precisos promediando las estimaciones de expertos de diferentes disciplinas trabajando de manera independiente que tomando cualquiera de las estimaciones individuales (Leimeister, 2010).

La inteligencia colectiva no es un fenómeno nuevo y ha sido objeto de estudio científico desde hace años en biología, ciencias sociales, ciencias de la computación, etc. En (Surowiecky, 2004) el autor postula que las condiciones necesarias para que se de la inteligencia colectiva son la diversidad de opiniones, la independencia y la descentralización de los participantes.

La inteligencia colectiva ha recibido un nuevo significado con la emergencia de las aplicaciones Web sociales y el contenido generado por sus usuarios. La participación colectiva tiene efectos tanto positivos como la donación de fondos para desastres naturales (huracán Katrina, terremoto de Haití), el seguimiento de iniciativas de activismo medioambiental (el fenómeno Earth Hour³ de WWF⁴ del que hablaremos más adelante) o el desarrollo de la Wikipedia (actualmente la Enciclopedia con mayor volumen de datos del mundo); como negativas o de dudoso beneficio para la sociedad como pueden ser las decisiones de la masa al elegir a un personaje como el Chikilicuatre para representar a España en el concurso de Eurovisión en 2008, o el apoyo a personajes como Belén Esteban que sin logros en la vida han conseguido fama, dinero y poder para influir en la opinión de sus seguidores. También es interesante comentar cómo la opinión se desarrolló en la Web durante la campaña electoral de Barack Obama en los Estados Unidos de América (Hughes & Palen, 2009).

2.1.1 Twitter

Twitter (Twitter Inc., 2010) es un sitio Web de *microblogging* en el que los usuarios expresan sus pensamientos y emociones en mensajes públicos (o tweets) de no más de 140 caracteres. Nació en una compañía de San Francisco como un proyecto para un sistema de comunicación (O'Reilly & Milstein, 2009). Los mensajes son leídos por los seguidores del que los publica. Cada usuario del sistema puede seguir a cualquier otro usuario, pudiendo así leer sus mensajes. Los tweets se almacenan en una base de datos de la que se pueden filtrar los mensajes gracias al motor de búsqueda Twitter Search⁵. Twitter es un servicio relativamente reciente, por lo que no son muchos los artículos de investigación a nivel académico de la herramienta.

2 Center for Collective Intelligence, Massachusetts Institute of Technology (MIT): <http://cci.mit.edu/>

3 Earth Hour: <http://www.earthhour.org>

4 World Wildlife Foundation (WWF): <http://www.wwf.org>

5 Twitter Search: <http://search.twitter.com/>

Twitter se ha convertido en un fenómeno en el que el crecimiento del número de usuarios es exponencial (O'Reilly & Milstein, 2009). En el artículo (Krishnamurthy *et al.*, 2008), sus autores investigaron las tasas de crecimiento, la difusión geográfica y otras propiedades estadísticas de Twitter. En (Java *et al.*, 2007) y (Huberman *et al.*, 2008) se realiza un trabajo similar al anterior pero con más énfasis en la perspectiva de red social de Twitter. En (Java *et al.*, 2007) también se recogen una serie de temas populares discutidos en Twitter. También se ha estudiado el uso de Twitter para respuesta a emergencias y convergencia de masas (durante huracanes y elecciones presidenciales en los Estados Unidos de América) en (Hughes & Palen, 2009).

Los artículos relacionados más cercanos a la temática de este trabajo son los de (Cheong & Lee, 2009) y (Cheong, 2009). En ellos se exploran las tendencias de Twitter mediante la herramienta Twitter Trends disponible en la página de Twitter (Twitter Inc., 2010). En (Cheong, 2009) también se mencionan otros trabajos de seguimiento de hashtags⁶ en Twitter relacionados con campañas de activismo. En su trabajo (Cheong, 2009) estudian durante 10 días los 10 primeros temas de moda cada 5 minutos a través de consultas a través de la API de Twitter⁷ a la herramienta de Twitter Trends. Una vez recolectadas las tendencias, las agrupan en las siguientes categorías:

Categoría	La tendencia se refiere a...
activismo	Ejemplo: usar Twitter para concienciar acerca del calentamiento global
conferencia	Los usuarios conversan a tiempo real acerca de una conferencia que se está desarrollando
cultura	Cultura popular
entretenimiento	Ejemplos: música, televisión, películas, famosos
general	Frases comunes y nombres propios, pero sin suficiente contexto como para enmarcar la tendencia
meme	Cibermemes y memes relacionados con Twitter
meme+entretenimiento	Memes originados en los medios relacionados con el entretenimiento, por ejemplo, uno originado por un famoso
noticias	Sucesos y noticias locales y globales (incluyendo situaciones de crisis)
ciencia	Noticias científicas (excluyendo las de TI y temas relacionados con la tecnología)
spam	Spam, intentos de phishing, actividad maliciosa
deporte	Noticias y eventos deportivos
tecnología	Temas relacionados con la tecnología o las TI como juegos, gadgets, software, ...
Twitter	Cambios oficiales al servicio de Twitter anunciados por Twitter Inc.
marketing viral	Uso de Twitter para promocionar productos o servicios (sin intención maliciosa)

Los resultados se muestran en (Fig. 1). Como se puede observar, el entretenimiento copa la clasificación. Las noticias y los memes relacionados con entretenimiento abarcan el 35% de las tendencias de Twitter. Si a esto le añadimos el deporte y creamos una categoría artificial de pasatiempos, esta reuniría el 55% de las tendencias. La ciencia y la cultura en esta red social en línea no abarcan más que el 3% de las tendencias si las juntamos en la misma categoría. Parece que la tecnología despierta especial interés en la comunidad Twitter con un 12% del total de las tendencias, pero los autores han incluido los juegos en esta categoría por lo que no podemos saber qué porcentaje se atribuye a los pasatiempos.

6 Hashtags: Cuando una palabra es precedida por un # en Twitter se le denomina *hashtag* y tienen como objetivo agrupar los Tweets acerca de un determinado tema para facilitar la búsqueda de información.

7 API de Twitter: <http://apiwiki.twitter.com/>

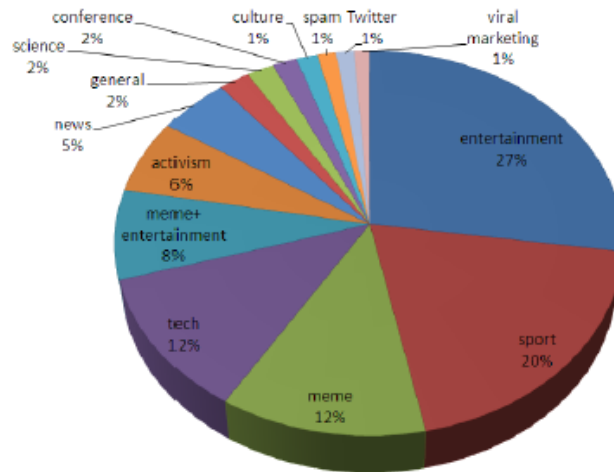


Figura 1: Temas de moda por categoría (Cheong, 2009)

En otro trabajo no académico (Pear Analytics, 2009) se clasifican los mensajes obtenidos durante dos semanas de lunes a viernes desde las 11:00 a las 17:00 en las siguientes categorías:

Categoría	Tipo de mensaje
noticias	Noticias de medios locales o globales no relacionados con tecnología y medios sociales provenientes de Mashable o TechCrunch
spam	Mensajes como “Mira cómo consigo 1000 seguidores en un día”
auto-promoción	Promoción de productos o servicios
sin interés general	Mensajes como “Me estoy comiendo un sandwich”
conversacionales	Mensajes que circulan entre amigos de manera similar a la mensajería instantánea
ReTweets	Mensajes que comienzan por RT y son retransmisiones de otros tweets

Los resultados se muestran en (Fig. 2).

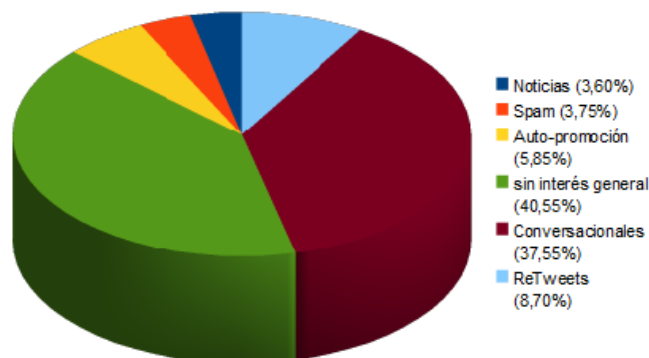


Figura 2: Mensajes por categoría (Pear Analytics, 2009)

Estos resultados son mucho más desalentadores. En este caso ya no se examinan únicamente las tendencias emergentes sino que se toma una muestra de 2000 mensajes. Aunque con un método de clasificación desconocido, se etiqueta al 41% de los resultados como “balbuceos sin sentido” que hemos traducido en la gráfica como mensajes sin interés general. Este estudio también revela que se usa Twitter como herramienta de mensajería instantánea dando lugar a un 38% de los mensajes del corpus extraído. Esto nos lleva a un 79% de mensajes que podrían clasificarse como charlatanería. Tampoco es muy alentador que del 21% restante, un 4% sea spam y un 6% sea publicidad de productos y servicios.

2.1.2 Earth Hour y activismo en redes sociales en línea

Earth Hour, o la La Hora del Planeta como se conoce en España, es una campaña de sensibilización y concienciación ante el calentamiento global emprendida por WWF. Comenzó en Sídney en 2007 y se pedía a los ciudadanos que apagasen la luz durante una hora un día de marzo. Ante el éxito del evento, se decidió realizar a escala global a partir del año 2008, y se sigue haciendo año tras año.

El 12 de febrero de 2009 se creó un grupo en la plataforma Facebook⁸ para la difusión de la iniciativa y se registró cada día el número de usuarios hasta el 28 de marzo de 2009, día en el que había que apagar la luz. El nombre del grupo es “Earth Hour 2009 (La Hora del Planeta)”⁹ y aún puede encontrarse en Facebook. La evolución del número de usuarios se muestra en (Fig. 3).

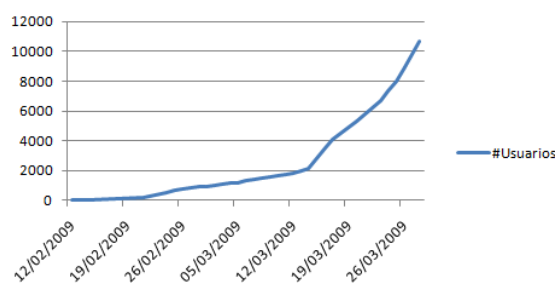


Figura 3: Evolución número de usuarios del grupo

La campaña de 2009 se difundió también en Twitter y se analizó a nivel local en Australia. En (Cheong & Lee, 2010) se observó una correlación entre el número de mensajes en Twitter acerca de la iniciativa y el porcentaje de energía ahorrada durante la Hora del Planeta. Los resultados se muestran en (Fig. 4).

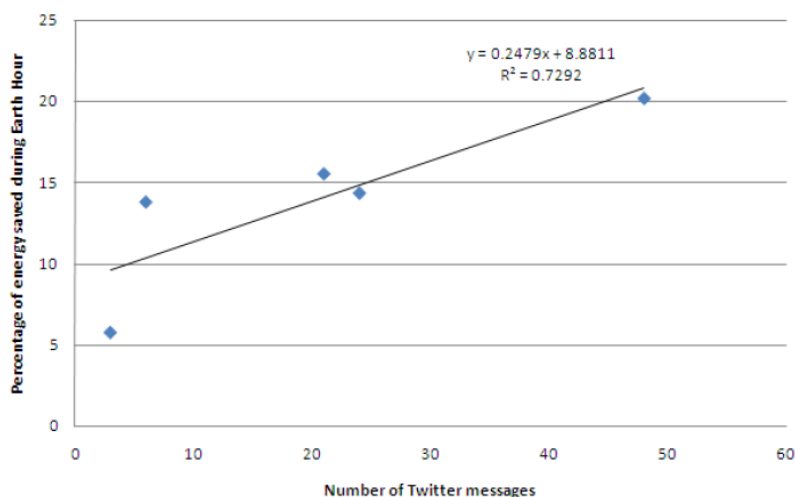


Figura 4: Porcentaje de energía ahorrado durante la iniciativa en función del número de mensajes en Twitter sobre Earth Hour (Cheong & Lee, 2010)

En España se apreció una divergencia entre la previsión de demanda energética la Red Eléctrica Española para el día 28 de marzo de 2009 y la demanda real de ese día como se muestra en (Fig. 5). En este caso se carece de datos para establecer correlación alguna, pero es un claro reflejo de cómo la difusión de un meme en medios sociales puede tener consecuencias en el mundo

8 Facebook: <http://www.facebook.com>

9 Grupo Earth Hour 2009 (La Hora del Planeta): <http://www.facebook.com/group.php?gid=63593843942>

real y en el comportamiento humano.

Como se ha visto anteriormente (Fig. 1), el 6% de las tendencias son de campañas de activismo político, social y ecológico. En (Cheong, 2010) se citan ejemplos de estudios relacionados con el ciberactivismo como las elecciones de Irán de 2009 seguidas con el hashtag #IranElection, la revolución moldava en Twitter (hashtag #Moldova), o iniciativas feministas en Facebook.

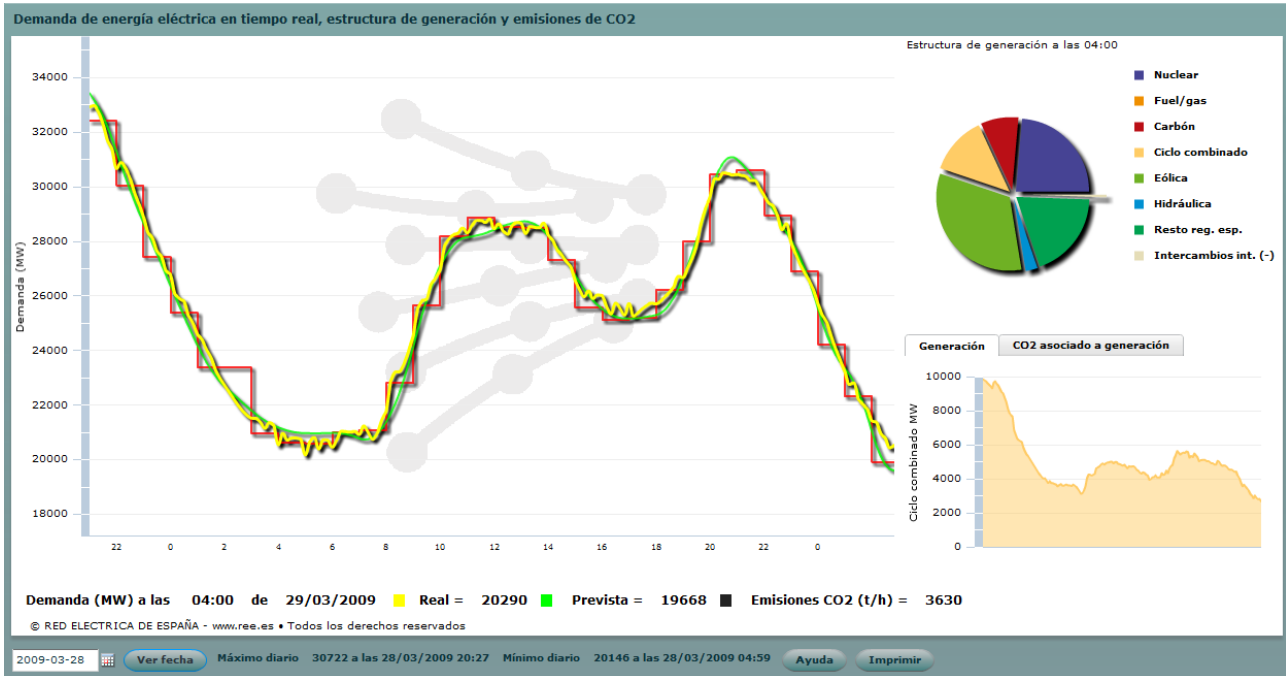


Figura 5: Demanda energética del día 28 de marzo de 2009¹⁰

En la misma línea de activismo ecológico, (Mankoff *et al.*, 2007) consideran la posibilidad de aprovechar el potencial de las redes sociales virtuales para motivar a sus usuarios para que reduzcan su huella ecológica. La huella ecológica es un indicador que mide el impacto que ejerce una cierta comunidad humana (país, región o ciudad) sobre su entorno, considerando tanto los recursos necesarios como los residuos generados para el mantenimiento del modelo de producción y consumo de la comunidad. Se mide en hectáreas per cápita y representa la superficie necesaria para producir los recursos consumidos por un ciudadano medio de una determinada comunidad humana, así como la necesaria para absorber los residuos que genera, independientemente de la localización de estas áreas.

Lamentablemente, no se han encontrado resultados sobre el impacto en la reducción de la huella ecológica o de la emisión de dióxido de carbono derivados del trabajo anterior. Pero en (Mankoff *et al.*, 2010) se demuestra gran aceptación de un sistema, integrado en redes sociales como MySpace¹¹ y Facebook, de recomendación de acciones que pueden adoptarse en la vida del día a día y pueden reducir la huella ecológica individual de cada usuario.

2.2 Expansión de consultas

Muy frecuentemente los usuarios de los motores de búsqueda y otros sistemas de recuperación de información introducen consultas incompletas sin ser conscientes de la efectividad de las consultas precisas. Para resolver este problema se estudian técnicas de expansión para

10 Obtenido de http://www.ree.es/operacion/curvas_demanda.asp

11 MySpace: <http://www.myspace.com/>

construir consultas robustas y completas. De esta manera se pretende devolver al usuario resultados que con su consulta original no podrían haber sido recuperados y que pueden ser relevantes.

La forma más básica de expansión de consulta es extraer el lexema cada término de la consulta para recuperar también las formas derivadas de la raíz de las palabras. De esta manera se incluyen en la consulta diferentes formas en género y número para adjetivos o nombres, y formas conjugadas para verbos. También se pueden detectar faltas de ortografía y proponer al usuario una nueva consulta corregida. Estas técnicas de expansión se pueden percibir fácilmente en el motor de búsqueda Google¹².

En (Zhang *et al.*, 2009) se describen diversos métodos más elaborados de expansión de consultas: mediante métodos estadísticos buscando co-ocurrencias de palabras de las consultas en corpus, utilizando ayudas léxicas como tesauros o libros de palabras. Aunque argumentan que estas técnicas no ayudan a mejorar la efectividad de la recuperación de información.

Métodos avanzados utilizan redes de conceptos creados de manera manual. De esta manera la efectividad de la recuperación de documentos depende de la calidad de la red conceptual. En (Zhang *et al.*, 2009) se propone un método de expansión de consultas que utiliza la red semántica WordNet para construir la red conceptual. Sus resultados indican que mediante esta técnica se mejora en un 7% la efectividad de la recuperación.

WordNet¹³ es una base de datos léxica del idioma inglés, con conexiones semánticas, construida manualmente y organizada en forma de tesoro. Nombres, verbos, adjetivos y adverbios se agrupan en conjuntos de sinónimos llamados Synsets. Cada Synset representa un concepto léxico que se caracteriza por su definición y se conecta con los demás Synsets por medio de relaciones semánticas. Para los nombres, las relaciones semánticas son la sinonimia, antonimia, hiponimia, hiperonimia, meronimia, holonimia y troponimia (Zhang *et al.*, 2009). Para los verbos son la sinonimia, antonimia, hiponimia, hiperonimia y troponimia.

3 Materiales y métodos

Para acceder a las opiniones y pensamientos de los usuarios de Twitter se ha ideado un prototipo de aplicación que, aprovechando los recursos existentes, permita sondear la inteligencia colectiva potencial de Twitter.

Para el tratamiento del lenguaje natural, se ha explorado un método de procesamiento superficial que encaja de cierta manera con la sintaxis del inglés. Como queremos extraer los pensamientos, se buscaran mensajes en el corpus de Twitter que contengan la forma “I think” o alguno de sus sinónimos proporcionados por un programa realizado en Java que accede a la base de datos de WordNet mediante el uso de la API JAWS¹⁴. Una vez obtenidos estos mensajes se hace una segunda criba para centrarnos en alguna temática en concreto buscando términos en los complementos del sintagma verbal. Para este proceso se utiliza la herramienta Searchtrix¹⁵ que se describe a continuación.

Searchtrix es un motor de búsqueda de mensajes en redes sociales en línea. Concretamente sus creadores afirman buscar mensajes en Twitter, Facebook y Topsy. Nosotros utilizaremos solo la opción de buscar en Twitter ya que en ninguna de las consultas realizadas se han obtenido mensajes de Facebook, y Topsy es un buscador social de noticias. Searchtrix cuenta con dos campos de

12 Google: <http://www.google.com/>

13 WordNet: <http://wordnet.princeton.edu/>

14 API Java JAWS para WordNet: <http://lyle.smu.edu/~tspell/jaws/index.html>

15 Searchtrix: <http://www.searchtrix.com/>

búsqueda. En ambos campos se pueden introducir cadenas separadas por comas. Para cada cadena del primer campo, se extraen todos los mensajes que la contengan y se realiza una consulta sobre los mensajes extraídos con cada cadena del segundo campo. Searchtrix devuelve los mensajes de las consultas realizadas detallando cuantos mensajes se obtuvieron de cada consulta.

Para ilustrar el funcionamiento supongamos el ejemplo en el que en el primer campo buscamos las cadenas “I think” y “I believe” y en el segundo “science”, “art” y “technology”. La consulta quedaría como se muestra en la siguiente tabla:

Primer campo	Segundo campo
i think, i believe	science, art, technology

La consulta nos devolvería mensajes que contengan “i think” seguido de “science”, “i think” seguido de “art”, “i think” seguido de “technology”, “i believe” seguido de “science”, “i believe” seguido de “art”, y “i believe” seguido de “technology”. También nos devolvería el número de mensajes para cada una de estas 6 consultas. La herramienta permite exportar estos datos a una hoja de cálculo dónde se pueden tratar para obtener los resultados que se exponen en el siguiente apartado.

Se ha explorado la relación de sinonimia que proporciona WordNet para crear redes conceptuales y usarlas en el segundo campo de búsqueda de Searchtrix, pero al no ser satisfactorios los resultados obtenidos, se ha decidido las redes conceptuales del segundo término de búsqueda de manera manual.

4 Consulta y resultados

4.1 Consulta

Dado que el *ciberactivismo* tiene cierto volumen de actividad en Twitter y este puede tener un impacto social y en el comportamiento humano –y de hecho tiene como objetivo tenerlo– en el mundo real, se estudiarán las tendencias de pensamiento colectivo alrededor de este tema. Puesto que la recuperación de información no es el objeto de la asignatura, se planteará una consulta de ejemplo para sondear la opinión colectiva acerca del activismo.

En el primer campo introduciremos la expansión de la consulta con el verbo “think” con el programa descrito anteriormente. Este nos devuelve los sinónimos que proporciona WordNet, y permite añadir a la forma verbal el pronombre personal “I” de la primera persona del singular del inglés. La aplicación nos devuelve el siguiente resultado:

Verbo	Expansión para el verbo
think	I think, I believe, I consider, I conceive, I opine, I suppose, I imagine, I reckon, I guess, I cogitate, I cerebrante, I retrieve, I recall, I call back, I call up, I recollect, I mean

Como algunas de las formas verbales expandidas son polisémicas las eliminaremos para intentar no desvirtuar los resultados. La siguiente tabla muestra la consulta completa realizada en Searchtrix, con la red conceptual del activismo creada manualmente.

Primer campo de Searchtrix	Segundo campo de Searchtrix
I think, I believe, I consider, I conceive, I opine, I suppose, I imagine, I reckon, I guess, I cogitate, I cerebrante, I mean	feminism, global warming, peace, poverty

4.2 Resultados

El número total de mensajes extraídos ha sido de 112. El resumen de los resultados se muestra en (Fig. 6). Como se puede observar, la mayoría de los mensajes relacionados con el activismo tienen que ver con movimientos por la paz con un 44,54% de los mensajes. Los siguientes activistas más dinámicos son los defensores del planeta que luchan contra el calentamiento global con un 36,61% de los mensajes. A estos les siguen los activistas contra la pobreza con un 10,71% de los mensajes recuperados y finalmente por los activistas a favor del feminismo con un 7,14% de tweets.

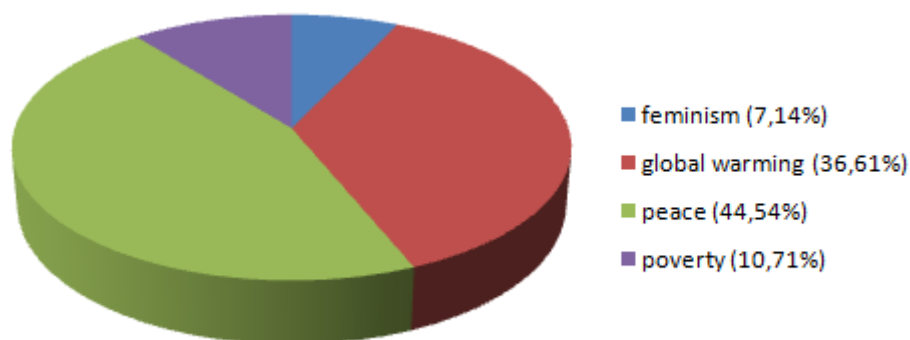


Figura 6: Mensajes por tipo de activismo

4.3 Discusión

Aunque los estudios comentados en el estado del arte (Cheong, 2009; Pear Analytics, 2009) sugieren que la mayoría de los mensajes de Twitter están relacionados con el entretenimiento o se clasifican como tweets sin interés, se ha desarrollado un prototipo de extracción de mensajes que permite estudiar otros fenómenos con más relevancia como puede ser el del ciberactivismo. Según los resultados obtenidos, que se pueden considerar como preliminares, ya que no se han revisado los mensajes manualmente ni se ha extendido la red conceptual del activismo por completo, se puede vislumbrar que el activismo más extendido en Twitter es por la paz. Sería conveniente contrastar estos resultados con una herramienta desarrollada al efecto, ya que se desconoce el mecanismo interno de funcionamiento de Searchtrix al no ser una herramienta de código abierto.

5 Conclusiones

Las redes sociales en línea sirven tanto como medio para influenciar a las masas como para reflejar eventos de la vida real, fuera de la Web. Aunque el activismo tiene cierta importancia como se puede desprender de los trabajos mencionados y puede ser muestra de una inteligencia colectiva a través de una colaboración implícita, parece que Twitter no es la herramienta ideal para la detección de propiedades emergentes de interés en el campo cultural o científico por el uso que le dan mayoritariamente sus usuarios. Parece que el potencial de la plataforma no es el que se esperaba, ya que 140 caracteres es un límite ideal para compartir ideas sintéticas y aforismos. Sin embargo se utiliza principalmente como herramienta para el entretenimiento y la charlatanería.

Aún así parece que con el método empleado para recuperar mensajes en Twitter sirve como

filtro preliminar para detectar ideas de interés. En este caso solo se ha explorado el activismo, pero la automatización total del proceso daría lugar a una herramienta útil en disciplinas como el marketing o la sociología.

6 Trabajo futuro

Como trabajo futuro, en el marco de la misma línea de investigación, se propone la automatización de los procesos del prototipo. También se podrían agregar funcionalidades de procesamiento del lenguaje natural más elaborado como análisis multilingüe, desambiguación de consultas, detección de ironía y sarcasmo, o detección de la polaridad de las opiniones. Esta última encaja en el marco de la asignatura para poder obtener una valoración cuantitativa de los comentarios en lenguaje natural del sistema KnowCat.

Siguiendo con los sistemas colaborativos, una vez extraídos los mensajes del mismo tema, mediante el uso de la API de Twitter se podría considerar un sistema automático de recomendación de usuarios para la creación de comunidades temáticas online. Esto sería especialmente útil para los activistas que hacen campañas en línea para aunar sus fuerzas en la misma lucha. En este caso, sería conveniente revisar y anotar manualmente los mensajes previamente para filtrar aquellos mensajes que formen parte de esos balbuceos sin interés que saturan estos sistemas.

En el dominio de la inteligencia colectiva se podría seguir estudiando cómo fenómenos de los sistemas sociales de la Web se pueden mapear en el comportamiento humano y en las sociedades en general.

7 Referencias bibliográficas

- Cheong, M. & Lee, V. (2009). "Integrating web-based intelligence retrieval and decision-making from the Twitter Trends knowledge base", *Proceedings of the Conference on Information and Knowledge Management 2009 Co-Located Workshops: Social Web Search and Mining 2009*, pp. 1-8.
- Cheong, M. (2009). "What are you Tweeting about?: A survey of Trending Topics within Twitter", Technical Report 2009/251, Clayton School of Information Technology, Monash University.
- Cheong, M. & Lee, V. (2010). "Twittering for Earth: A Study on the Impact of Microblogging Activism on Earth Hour 2009 in Australia", *Lecture Notes in Computer Sciences*, vol. 5991, pp. 114-123.
- Dawkins, R. (1976). "The Selfish Gene", New York: Oxford University Press.
- Huberman, B.A.; Romero, D.M. & Wu, F. (2008). "Social networks that matter: Twitter under the microscope", disponible en <http://ssrn.com/abstract=1313405>.
- Hughes, A.L. & Palen, L. (2009). "Twitter adoption and use in mass convergence and emergency events", *Proceedings of the 6th International Information Systems for Crisis Response And Management Conference*.
- Java, A.; Song, X.; Finin, T. & Tsen, B. (2007). "Why we Twitter: An Analysis of a Microblogging Community", *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, pp. 118-138.
- Krishnamurthy, B.; Gill, P. & Arlitt, M. (2008). "A Few Chirps About Twitter", *Proceedings of the Workshop on Online Social Networks 2008*, pp. 19-24.
- Leimeister, J. M. (2010). "Collective Intelligence", *Business & Information Systems Engineering*, vol. 2, no. 4, pp. 245-248.
- Mankoff, J.; Matthews, D.; Fussell, S. & Johnson, M. (2007). "Leveraging social networks to motivate individuals to reduce their ecological footprints", *Proceedings of the 40 Hawaii International Conference on System Sciences*, pp. 1-10.
- O'Reilly, T. & Milstein, S. (2009). "The Twitter Book", Sebastopol CA: O'Reilly Media Inc.
- Pear Analytics (2009). "Twitter Study – August 2009", disponible en <http://www.pearanalytics.com/blog/wp-content/uploads/2010/05/Twitter-Study-August-2009.pdf> (Último acceso: 01/09/2010)
- Prescott, L. M.; Harley, J. P. & Klein, D. A. (2002). "Microbiología", 5ª Ed. McGraw-Hill Interamericana de España, SL, Capítulo 6, pp. 141-142.
- Robinson, D.E.; O'Donovan, J.T.; Sharma, M.P.; Doohan, D.J. & Figueroa, R. (2003). "The Biology of Canadian Weeds. 123. *Senecio vulgaris* L.", *Canadian Journal of Plant Science*, vol. 83, pp. 629-644.
- Surowiecki, J. (2004). "The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations", New York: 1st Doubleday Books.

Twitter Inc. (2010). "Twitter", <http://www.twitter.com>

Zhang, J.; Deng, B. & Li, X. (2009). "Concept Based Query Expansion Using WordNet", *Proceedings of the 2009 International e-Conference on Advanced Science and Technology*, pp. 52-55.