2 Panorama general

Este capítulo presenta un panorama general de los temas que se conjuntan en este documento. Para ello, y teniendo en cuenta que el presente trabajo aúna conceptos de diversos ámbitos, se ha estructurado en cuatro subsecciones. La primera presenta una panorámica de lo que se conoce como minería de textos, partiendo desde la revisión del concepto, su taxonomía y su importancia en áreas de especialidad. La segunda sección está dedicada a introducir el concepto de minería web. La tercera sección de este capítulo muestra una visión global del ámbito de las redes sociales, su uso y aplicaciones importantes generadas alrededor de ellas. Finalmente en la cuarta subsección se hace una introducción a los conceptos básicos Twitter y sus ventajas sobre otras redes sociales.

2.1 Minería de Textos

La minería de textos es un proceso en donde se interactúa con una colección de documentos a través del tiempo mediante el uso de herramientas de análisis. La minería de textos al igual que la minería de datos busca extraer información útil identificando y explorando patrones interesantes. En el caso de la minería de texto la fuente de los datos son documentos y los patrones de interés no se encuentran entre los registros de una base de datos formal, si no en datos no estructurados en el conjunto de documentos.

Ciertamente la minería de textos es inspirada en gran parte por la minería de datos, por lo tanto no es de sorprender que se empleen las mismas técnicas y algoritmos para el descubrimiento de patrones. Debido a que la minería de datos supone que los datos han sido almacenados en un formato estructurado, gran parte del proceso previo cae en dos tareas fundamentales: depuración y normalización de los datos. En contraste, en la minería de textos el procesamiento se centra en la identificación y extracción de características representativas de documentos de lenguaje natural, operaciones que intentan transformar los datos no estructurados de los documentos en un formato explicito para poder aplicar técnicas de minería de datos.

2.2 Minería web

La minería web en términos generales se puede ver como la aplicación de métodos de minería de datos adaptados a la web para descubrir y extraer información de documentos y servicios Web, analizando el contenido de documentos web, las páginas web que están vinculadas a través de hipervínculos y estadísticas de uso para ayudar a los usuarios a satisfacer necesidades de información. La minería web se puede categorizar en tres áreas: minería de estructura web, minería de contenido web y minería de uso web [26].

La **minería de estructura web** explota la información que proporcionan los hipervínculos respecto a la estructura de la página que pudiera ser importante [26]. Por ejemplo, se pueden descubrir páginas importantes, también se pueden descubrir comunidades de usuarios con intereses en común, tareas que en la minería de datos no es posible debido a que usualmente no hay una estructura de enlaces en una tabla relacional.

En la **minería de contenido web** se extrae información o conocimiento del contenido de la página web [26]. Por ejemplo, se puede clasificar y agrupar las páginas web de acuerdo con ciertos temas. Tareas que son similares a las de minería de datos. Sin embargo, en la minería de contenido web es posible descubrir datos útiles tales como descripciones de productos, fotografías publicadas, etc., para muchos propósitos. Además, es posible extraer opiniones de clientes y descubrir los sentimientos de los consumidores. Tareas que no son tradicionales de la minería de datos.

Por último la **minería de uso web** estudia la relación que existe entre documentos en la web que son identificados por búsquedas anteriores registrando búsquedas y accesos de los usuarios a dichos contenidos [26].

2.3 Redes sociales

A lo largo de su historia, el ser humano ha buscado el apoyo de las personas que le rodean ya que éstas le pueden resultar de ayuda para satisfacer tanto sus necesidades básicas, como de seguridad, protección y afecto. Esto se debe a que el ser humano es social y vive inmerso en un entramado de vínculos interpersonales que afectan su vida. Así, el comportamiento de cada individuo afecta y a su vez se ve afectado por las interacciones sociales en las que participa.

El problema de las redes sociales en el mundo no virtual es que la mayoría de las conexiones entre las personas están ocultas, la red puede tener un gran potencial pero sólo es tan valiosa como las conexiones y las personas que una persona puede ver. El valor de las conexiones es explotado por las redes sociales en internet que se apoyan en el *principio de los seis grados de separación*[3], una teoría que intenta probar la idea de que cualquier persona en la tierra puede estar conectada a otra persona a través de una cadena de conocidos que no tiene más de cinco intermediarios. De esta manera se conectan entre sí a ambas personas con sólo seis enlaces y esto da como resultado que el conjunto de conocidos de cada persona forme la población mundial[4].

[3] Teoría inicialmente propuesta en 1930 por el escritor húngaro Frigyes Karinthy en un cuento llamado *Chains*. Wikipedia 2010-09-11, "Seis grados de separación". http://es.wikipedia.org/wiki/Seis_grados_de_separaci%C3%B3n.

[4]Semioticon 2010-08-11, "The Networked Individual: A Profile of Barry Wellman". http://www.semioticon.com/semiotix/semiotix14/sem-14-05.html.

Las redes sociales en internet son sitios web donde las personas se dan de alta, creando así un perfil (una página web personal) para posteriormente agregar el perfil de sus amigos, personas a quienes invita u otros usuarios que ya pertenecen a la red social. De esta manera permite así la interacción entre personas que no necesariamente se conocen pero que comparten intereses, preocupaciones o necesidades. Esto es porque las redes sociales no sólo sirven para mostrar fotografías o documentos.

Hay varios tipos de redes sociales para cada interés o necesidad, dividiendo así a las redes sociales en tres tipos: generales, profesionales y temáticas. Ellas permiten al usuario dejarse ver en internet, encontrar personas, mantener relaciones distantes, conocer gente nueva, compartir conocimientos, aprender, compartir contenidos, participar en grupos de interés comunes, debatir, divertirse, realizar negocios, encontrar y ofrecer trabajo e incluso realizar trabajo colaborativo.

2.4 Acerca de Twitter

Twitter es un servicio de red social en internet basado en lo que se conoce como microblogging o nanoblogging. Una mezcla entre la acción de postear[5] o escribir en los *blogs*[6] (*blogging*) con red social y mensajería instantánea[7] que permite a los usuarios enviar y publicar mensajes breves (en promedio 140 caracteres) generalmente sólo de texto [27].

Los usuarios de Twitter (tweeteros) escriben actualizaciones cortas, o *tweets* (gorjeos: el sonido emitido por las aves), de 140 caracteres o menos. Acto llamado por algunos usuarios como twittear (o tweetear). Estos mensajes son publicados en la página principal del usuario y enviados a sus *followers* (seguidores), usuarios de Twitter que reciben los gorjeos y descubren las noticias ("¿qué está pasando?") relacionadas con los temas que le interesan.

Los mensajes de los usuarios que se siguen son mostrados en una cronología, o *timeline* (*TL*), un término utilizado para describir gorjeos que son ordenados cronológicamente en Twitter. Es como recibir un periódico en el que siempre se encuentran titulares interesantes, pudiendo descubrir noticias mien-

---

[5] El acto de crear temas o escribir mensajes en algún medio informático.

[6] Un *blog*, o en español también una bitácora, es un sitio web periódicamente actualizado que recopila cronológicamente textos o artículos de uno o varios autores, apareciendo primero el más reciente, donde el autor conserva siempre la libertad de dejar publicado lo que crea pertinente. Habitualmente, en cada artículo de un blog, los lectores pueden escribir sus comentarios y el autor darles respuesta, de forma que es posible establecer un diálogo. Wikipedia 2011-04-16 “Blog”, http://es.wikipedia.org/wiki/Blog.

[7] La mensajería instantánea es una forma de comunicación en tiempo real entre dos o más personas basada en texto. El texto es enviado a través de dispositivos conectados a una red como Internet. . Wikipedia 2011-04-16 “Mensajería Instantánea”, http://es.wikipedia.org/wiki/Mensajer%C3%ADa_instant%C3%A1nea.

tras están sucediendo, aprender más sobre los temas que son importantes para el usuario, y obtener las novedades en el momento. Información que puede ser utilizada por los usuarios para encontrar su propia voz y mostrar a otros lo que les interesa. Los gorjeos pueden tener un interés periodístico, informativo, o simplemente divertido y los usuarios pueden retwittear, o retweetear los gorjeos (hacer un retweet) para compartir información rápidamente a sus seguidores, o responder con su reacción a un gorjeo haciendo un retweet y agregando un comentario. Los retweets pueden ser de dos tipos, uno donde se publica directamente el gorjeo original en la línea del tiempo del usuario y otro donde se inicia con el termino *RT* seguido de la "mención" del usuario original utilizando su nombre en Twitter precedido por el signo @. Estos retweet pueden ser vistos por sus creadores en el apartado de sus gorjeos retwitteados (o tweets retweetteados). Las menciones no sólo sirven para identificar de qué usuario se retwitteo (o retweeteo) un gorjeo, también sirve como su nombre lo indica para mencionar a otros usuarios y atraerlos así al mensaje o incluso iniciar una conversación a partir de la respuesta a un gorjeo interesante. Los usuarios pueden ver sus menciones en el apartado de "menciones".

Además de las actualizaciones públicas @respuestas, es posible enviar gorjeos privados a los seguidores, también llamados mensajes directos (MD o DM). Al igual, la gente a la que se sigue puede enviar un mensaje privado. Estos mensajes se escriben iniciando con la letra "d" y el nombre de usuario del seguidor.

Twitter identifica los Temas del Momento *(o trending topics).* Temas que se están popularizando(o que son promovidos) y que son representados por sus palabras clave[8]. Pero, aunque los usuarios pudieran estar hablando de un tema en común, en Twitter no existía una forma para que los usuarios pudieran darle seguimiento a un tema por ellos mismos, por lo que los usuarios crearon un método de categorización [25]. Actualmente los usuarios clasifican sus gorjeos con lo que se conoce como etiquetas o hashtags, las cuales se componen del símbolo de almohadilla o símbolo de gato (#) antecediendo a las palabra que identifica la categoría[9], como por ejemplo: #México, #música, #peliculaX. Y son estas etiquetas precisamente las que aparecen más frecuentemente como temas del momento en comparación con las palabras que pudieran tener en común muchos gorjeos en un momento dado.

Una etiqueta muy utilizada es la de #FF que significa "*Follow Friday*." algo así como "Viernes para Seguir" y la utilizan los usuarios de Twitter para sugerir a otros a quiénes seguir los Viernes.

A diferencia de otras redes sociales Twitter no tiene soporte nativo para compartir videos o archivos; no están contenidos dentro del gorjeo, en vez de esto el mensaje tiene que contener la dirección URL

[8] Palabras o términos más representativos de un contenido.

[9] A parte de no utilizar las etiquetas con propósitos de enviar spam, no hay reglas formales para su utilización, por lo que cualquier palabra clave con una # delante puede ser considerada una etiqueta.

de lo que se desea compartir. Esto lo deja a servicios de terceros como son por mencionar algunos: YouTube, Dailymotion, Vimeo, Tu.tv, justin.tv y Megavideo para videos y Megaupload, Rapidshare, Mediafire y Hotfile para compartir archivos. Recientemente se agregó la opción de incluir imágenes dentro del gorjeo de forma nativa, pero en sus inicios Twitter lo dejaba a servicios como Twitpic, Yfrog, Instagram, Lockerz y Flickr.

Claro que incrustar una dirección URL podría consumir los 140 caracteres. Para hacer frente a este problema Twitter ofrece un servicio de enlaces donde los enlaces compartidos en Twitter.com serán acortados a un enlace http://t.co. de 19 caracteres de longitud (Muy parecido al servicio de bit.ly).

Todos los gorjeos guardan la fecha y la hora en la que se publicaron, el identificador del usuario y su nombre de usuario de Twitter de quien lo publica y a quien va dirigido (de ser el caso). También se incluye la dirección URL del gorjeo, la imagen del usuario, el mensaje del gorjeo, el idioma que específico el usuario en su perfil de Twitter y finalmente la ubicación geográfica.

La función de Tweets con Ubicación permite añadir selectivamente información de ubicación en los gorjeos. Esta función está desactivada al inicio, y es necesario activarla para poder usarla. Una vez activada la opción es posible añadir la ubicación individualmente a cada gorjeo en Twitter.com y vía otras aplicaciones o dispositivos móviles que den soporte a esta opción. La información que se comparte públicamente puede ser ya sea la ubicación exacta (coordenadas) o un lugar (barrio o cuidad), añadiendo así contexto a las actualizaciones de los usuarios y puede ayudar a unir los gorjeos en una conversación local.

Twitter ofrece ciertas ventajas sobre otros microbloggingins como lo son Tumblr (https://www.tumblr.com/), Meme de Yahoo (http://meme.yahoo.com/), Picotea (http://picotea.com/es/), Google buzz (http://www.google.com/buzz), Friendfeed (http://friendfeed.com/) o Identi.ca (http://identi.ca/). En algunos de ellos es posible incrustar videos, imágenes o archivos en sus mensajes, pero pocos de ellos cuentan con geolocalización. Por defecto los mensajes de Twitter son públicos, algo que no es muy común en los ya mencionados. Por otra parte, la API de Twitter es accesible y fácil de implementar, algo que en otras redes puede no estar disponible para los desarrolladores externos. Finalmente Twitter tiene muchos más usuarios en términos generales, especialmente en México siendo de particular interés para el presente trabajo.

Lo anterior se puede extender a otras redes sociales de propósito general como Facebook (http://www.facebook.com/), Google+ (https://plus.google.com/), redes mucho muy utilizadas donde también es posible publicar el "¿qué está pasando?" pero que por su característica de compartir información personal como nombre, dirección, correo electrónico, fotografías personales y actividades que se realizan en el día, se han convertido en redes sociales privadas. Es posible hacer uso de las actualiza-

ciones de los usuarios, siempre y cuando dichos usuarios acepten el vínculo entre su perfil y una aplicación de terceros, algo que resulta particularmente complicado. Por otro lado las publicaciones de los usuarios van dirigidas a su círculo de amigos: que hace en su trabajo, como le fue hoy, etc. Publicaciones que difieren a las que se encuentran en Twitter, donde se publica generalmente para sus seguidores; gente desconocida que quiere saber no sólo de lo que hace otro usuario, sino también qué piensa o qué ocurre donde él está presente.