

# En un solo tweet hay más metadatos que caracteres

ALICIA HERNÁNDEZ MARTÍN

Twitter se ha convertido en una de las redes sociales más utilizadas y una excelente herramienta de comunicación a través de la cual millones de personas distribuyen información mediante los llamados tweets.

Pero un tweet es mucho más que un mensaje instantáneo o la publicación del estado de un usuario, es también una potente fuente de información para las empresas dedicadas a la minería de datos gracias a los metadatos que lleva asociados. Cada mensaje publicado lleva incorporado un paquete de información que viaja por la Red pasando totalmente desapercibido para la mayoría de los usuarios de este servicio de microblogging.

Cada tweet está compuesto por tres bloques fundamentales: el primero de ellos es un identificador único de 10 dígitos que reconocerá el mensaje durante su recorrido por el ciberespacio, el siguiente lo constituyen los 140 caracteres que forman el mensaje publicado y en el último se encuentran los metadatos. Este último bloque lo componen todos aquellos datos adicionales que proporcionan información como el nombre o alias del autor, su dirección en la red, su ubicación, la fecha en que se creó su cuenta de Twitter, el idioma que utiliza, la zona horaria del lugar desde donde se envía el mensaje, el número

de personas a las que sigue, el número de contribuciones que ha realizado en la red social, el identificador de todas las respuestas que recibe, cuáles son sus favoritos, los retweets que realiza... Y así se podrían seguir enumerando hasta 150 tipos de referencias distintas.

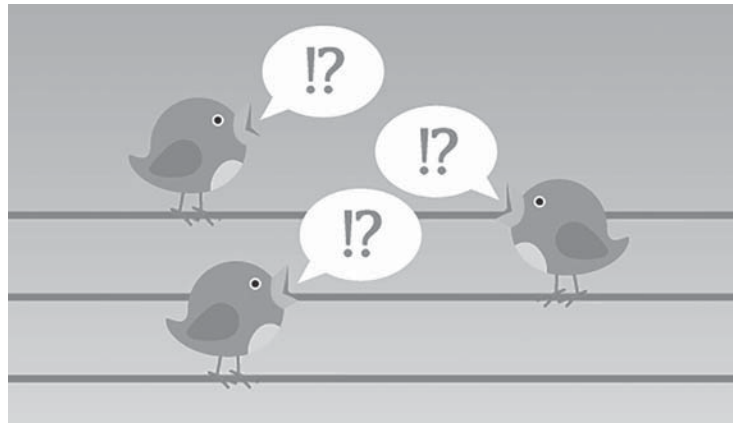
Todos estos datos, sumados a los aportados en los propios mensajes de los usuarios, permiten extraer gran canti-

dad de información que están aprovechando las empresas dedicadas a la minería de datos para diferentes fines. Una adecuada combinación de metadatos puede servir para realizar estudios demográficos agrupando los datos por segmentos de población, para establecer tendencias de voto, para realizar estadísticas acerca de cuáles son las películas o programas de TV más vistos por los usuarios, qué marcas



recomiendan y usan la mayoría de ellos, cuáles son las que descartan, etc.

Los resultados del análisis de toda esta información son tremendamente valiosos para las agencias de publicidad y mercadotecnia porque les permite encontrar patrones y puntos de vista entre sus clientes potenciales, además de mantenerse al tanto sobre la evolución de las técnicas de



sus competidores y las tendencias de consumo.

En un documento presentado ante la Comisión del Mercado de Valores en Estados Unidos en octubre del año pasado, Twitter hacía públicos sus datos y reconocía tener 218 millones de usuarios activos al mes, los cuales envían 500 millones de mensajes al día. Esto nos da una idea de la potencia que tiene este servicio como fuente de información.

Las oportunidades que ofrece son inmensas pero, a su vez, un tanto intranquilizadoras, debido a que la información que proporcionan los metadatos asociados a los tweets supone una amenaza para la protección de los datos personales de los usuarios que los envían. La posibilidad de analizar estos datos se convirtió en un tema candente cuando los documentos filtrados por Edward Snowden revelaron que la Agencia de Seguridad Nacional de Estados Unidos había revisado correos electrónicos, llamadas telefónicas e intervenciones en las redes sociales de los estadounidenses. Varios grupos en defensa de las libertades civiles luchan desde entonces para que se limite el acceso masivo tanto al contenido como a los metadatos de los mensajes transmitidos a través de la red, ya que de ellos se puede conseguir gran cantidad de información de carácter personal de los usuarios. ■

The diagram illustrates a JSON object representing a Twitter tweet, with various fields annotated by callouts explaining their meaning and usage. The JSON object is as follows:

```

{
  "id": 12296272736,
  "text": "An early look at Annotations: http://groups.google.com/group/twitter-api-announce/browse_thread/thread/fa5da2608865453",
  "created_at": "Fri Apr 16 17:55:46 +0000 2010",
  "in_reply_to_user_id": null,
  "in_reply_to_screen_name": null,
  "in_reply_to_status_id": null,
  "favorited": false,
  "truncated": false,
  "user": {
    "id": 6253282,
    "screen_name": "twitterapi",
    "name": "Twitter API",
    "description": "The Real Twitter API. I tweet about API changes, service issues and happily answer questions about Twitter and our API. Don't get an answer? It's on my website.",
    "url": "http://apiwiki.twitter.com",
    "location": "San Francisco, CA",
    "profile_background_color": "e1dfee",
    "profile_background_image_url": "http://a3.twimg.com/profile_background_images/59931895/twitterapi-background-new.png",
    "profile_background_title": false,
    "profile_image_url": "http://a3.twimg.com/profile_images/689684365/api_normal.png",
    "profile_link_color": "0000ff",
    "profile_sidebar_border_color": "87bc44",
    "profile_sidebar_fill_color": "e0ff92",
    "profile_text_color": "000000",
    "created_at": "Wed May 23 06:01:13 +0000 2007",
    "contributors_enabled": true,
    "favourites_count": 1,
    "statuses_count": 1628,
    "friends_count": 13,
    "time_zone": "Pacific Time (US & Canada)",
    "utc_offset": -28800,
    "lang": "en",
    "protected": false,
    "followers_count": 100581,
    "geo_enabled": true,
    "notifications": false,
    "following": true,
    "verified": true,
    "contributors": [3191321],
    "geo": null,
    "coordinates": null,
    "place": {
      "id": "2b6ff8c22edd9576",
      "url": "http://api.twitter.com/1/geo/id/2b6ff8c22edd9576.json",
      "name": "SoMa",
      "full_name": "SoMa, San Francisco",
      "place_type": "neighborhood",
      "country_code": "US",
      "country": "The United States of America",
      "bounding_box": {
        "coordinates": [
          [
            [-122.42284884, 37.76893497],
            [-122.3964, 37.76893497],
            [-122.3964, 37.78752897],
            [-122.42284884, 37.78752897]
          ]
        ],
        "type": "Polygon"
      },
      "source": "web"
    }
  }
}

```

The callouts provide the following explanations:

- DEPRECATED**: Points to the `id` field.
- The tweet's unique ID. These IDs are roughly sorted & developers should treat them as opaque (http://bit.ly/dCkppc).**: Points to the `id` field.
- Text of the tweet. Consecutive duplicate tweets are rejected. 140 character max (http://bit.ly/4ud3he).**: Points to the `text` field.
- Tweet's creation date.**: Points to the `created_at` field.
- The ID of an existing tweet that this tweet is in reply to. Won't be set unless the author of the referenced tweet is mentioned.**: Points to the `in_reply_to_status_id` field.
- The screen name & user ID of replied to tweet author.**: Points to the `in_reply_to_user_id` field.
- Truncated to 140 characters. Only possible from SMS.**: Points to the `truncated` field.
- The author's user name.**: Points to the `user.name` field.
- The author's biography.**: Points to the `user.description` field.
- The author's screen name.**: Points to the `user.screen_name` field.
- The author's URL.**: Points to the `user.url` field.
- The author's "location". This is a free-form text field, and there are no guarantees on whether it can be geocoded.**: Points to the `user.location` field.
- Rendering information for the author. Colors are encoded in hex values (RGB).**: Points to the `user.profile_background_color` field.
- The creation date for this account.**: Points to the `user.created_at` field.
- Whether this account has contributors enabled (http://bit.ly/50npuu).**: Points to the `user.contributors_enabled` field.
- Number of favorites this user has.**: Points to the `user.favourites_count` field.
- Number of tweets this user has.**: Points to the `user.statuses_count` field.
- Number of users this user is following.**: Points to the `user.friends_count` field.
- The timezone and offset (in seconds) for this user.**: Points to the `user.time_zone` field.
- The user's selected language.**: Points to the `user.lang` field.
- Whether this user is protected or not. If the user is protected, then this tweet is not visible except to "friends".**: Points to the `user.protected` field.
- Number of followers for this user.**: Points to the `user.followers_count` field.
- DEPRECATED in this context**: Points to the `user.notifications` field.
- Whether this user has a verified badge.**: Points to the `user.verified` field.
- DEPRECATED**: Points to the `user.contributors` field.
- The place ID**: Points to the `place.id` field.
- The contributors' (if any) user IDs (http://bit.ly/50npuu).**: Points to the `place.contributors` field.
- The printable names of this place**: Points to the `place.name` field.
- The URL to fetch a detailed polygon for this place**: Points to the `place.url` field.
- The type of this place - can be a "neighborhood" or "city"**: Points to the `place.place_type` field.
- The place associated with this Tweet (http://bit.ly/b8L1Cp).**: Points to the `place` object.
- The country this place is in**: Points to the `place.country` field.
- The application that sent this tweet**: Points to the `place.source` field.
- The bounding box for this place**: Points to the `place.bounding_box` field.
- Whether this user has geo enabled (http://bit.ly/4pFY77).**: Points to the `user.geo_enabled` field.
- Whether this user has geo enabled (http://bit.ly/b8L1Cp).**: Points to the `place.geo` field.

Map of a Twitter Status Object  
 Raffi Krikorian <raffi@twitter.com>  
 18 April 2010

Mapa de un tweet compuesto por Raffi Krikorian, ingeniero de Twitter, en abril de 2010. Aunque haya pasado algo de tiempo desde su publicación no ha cambiado demasiado respecto al esquema técnico de un tweet de hoy, y se pueden ver señalados en la imagen algunos de los metadatos que componen un tweet acompañados de una breve explicación de para qué sirven.