

Ensayo / Essay**Prensa y "Big Data": El desafío de la acumulación y análisis de datos***Press and "Big Data": The challenge of data accumulation and analysis*

Fecha de publicación: 17/06/2013

**Raymond COLLE***

Profesor jubilado Facultades de Comunicación Pontificia Universidad Católica de Chile y Universidad Diego Portales, Santiago de Chile

1. Introducción

Como señaló hace poco el diario español ABC, "Uno de los sectores que más interés está suscitando en las empresas es el estudio y procesamiento de grandes e ingentes cantidades de datos. El uso del *Big Data* en las compañías será cada vez más importante en los próximos años y solo en España crecerá un 300% en dos años." (23 de mayo 2013). El tema del *Big Data*, es decir, de los datos masivos, no es ajeno a los medios de comunicación: la acumulación de noticias en sus bases de datos presenta, desde hace tiempo, este tipo de problemática. Pero con el auge de las redes

sociales y las versiones digitales de los periódicos se agrega otro tipo de *Big Data*: la recolección de datos sobre los lectores, tal como ocurre para las otras empresas presentes en la web. Tener en cuenta lo que ocurre en la “red de redes”, es cada vez más necesario: de ello puede depender ahora la supervivencia del medio de comunicación.

Abordamos aquí esta problemática en su doble aspecto: la información noticiosa y la extraíble acerca de los lectores, precisando sus componentes centrales y cómo pueden ser enfrentadas.

2. La masa noticiosa

Sistemas poderosos para recolectar datos y administrarlos en grandes bases de datos están presentes en toda compañía grande y media, y con más razón en los medios de prensa, cuyo centro de documentación ha debido recurrir a este tipo de sistema para acceder tanto a los contenidos publicados como a la enorme cantidad de datos de referencia siempre necesarios. Sin embargo, cuando se pretende transformar estos datos en factor de éxito surge la dificultad de extraer de ellos un conocimiento verdaderamente útil tanto para los periodistas como, a través de ellos, para los lectores.

Si bien en Estados Unidos la informatización de los centros de documentación de los medios de prensa se inició en los años 70, en Europa empezó a mediados de los 80. Y el costo de infraestructura (especialmente para conservar grandes cantidades de datos en medios magnéticos) llevó a muchos medios a constituir sistemas que estuvieran al servicio de conjuntos de empresas, anticipándose al concepto y a la realidad de las “nubes”. Gracias a la posterior evolución de los sistemas de compresión de datos y el abaratamiento de soportes digitales cada vez más potentes, casi todos cuentan hoy con bases de datos propias donde conservan información, al menos, sobre toda publicación propia.

Ante este panorama, hace algo más de diez años se nos ocurrió una pregunta: ¿Puede el periodismo beneficiarse de las nuevas tecnologías de manejo avanzado de estas bases de datos? Consideramos que, como gran “consumidor” –y difusor– de información, no debía quedarse al margen de los aportes que estas técnicas podían significar para sus analistas y comentaristas. Al contrario, estimamos que no se podía ignorar la importancia de éstas para su labor en el nuevo siglo. Nadie mejor que un medio de comunicación puede “auscultar” la vida de la sociedad para descubrir los hilos que unen los hechos y explican los éxitos y fracasos en



múltiples campos. Fue el origen de nuestro libro “Explotar la información noticiosa: *Data Mining* aplicado en el periodismo” (Colle, 2002).

La “minería de datos” (*Data Mining*) consiste esencialmente en aplicar técnicas que permiten descubrir y explicitar la información oculta en los megabytes de las bases de datos. La principal ventaja de los sistemas más avanzados es que “aprenden” automáticamente a partir de la historia del contenido investigado, formulando y probando hipótesis acerca de las reglas, reiteraciones o todo tipo de relaciones –muchas no perceptibles a primera vista– que aparecen entre diferentes estados, momentos o casos (como las diversas noticias, en el caso de la prensa). Cuando se ha descubierto un conocimiento valioso y conciso acerca del conjunto de interés, se puede incorporar en algún proceso de toma de decisión, por ejemplo de selección del tema de un reportaje, y luego en su desarrollo.

Hace diez años, el sistema y la metodología propuestos parecieron sin duda complejos y costosos, y son escasos los medios periodísticos que tomaron en cuenta la propuesta. Hoy día, sin embargo, con la explosión de las redes sociales, recurrir a este tipo de proceso se hace cada vez más urgente. *Big Data* no tiene que ver únicamente con el volumen, tamaño o como su nombre lo indica “Grandes Datos”, también hay que considerar factores como la velocidad de generación de la información, así como el número y la variedad de fuentes que producen la misma y, finalmente, la habilidad de extraer valor de dicha información (Diario TI, 25 de abril 2013).

De hecho están naciendo medios que fundan declaradamente su contenido en el análisis de datos. Es el caso, por ejemplo, de *Fact Tank*, que se muestra como una plataforma en la que se analizarán las noticias del día con el objetivo de dar contexto y proporcionar más información a los usuarios desde una perspectiva del periodismo de datos. Las cifras y los números serán la principal tarea a abordar de este sitio (Web “Clases de Periodismo”, 28 de mayo 2013).

Es más, está apareciendo una nueva especialización para los periodistas: el “periodismo de datos” que, según algunos autores, podría ser “la salvación del periodismo” (Gustavo Gorriti, en el Seminario Internacional de Periodismo de Investigación celebrado en mayo de 2013 en Perú). Gorriti señala que se requieren grandes inversiones –sin duda en software y hardware, así como en personal calificado– para que el periodismo de investigación pueda lograr importantes frutos, como ocurre, por ejemplo, con Propublica, una agencia norteamericana especializada en periodismo de



investigación. También es necesario el desarrollo de contenido digital mediante el uso de Internet. Las investigaciones pueden proveer de información trascendental si se filtran y analizan bien los datos. Así, se podría lograr que las suscripciones ayuden a financiar estas mismas investigaciones y proyectos, e incluso generar ganancias (Web "Clases de Periodismo", 29 de mayo 2013).

3. Lectores y redes sociales

De momento, en España, según asegura International Data Corporation (IDC) en un informe presentado hace escasos meses, solo un 4,8% de las empresas utiliza ya la tecnología de *Big Data* y un 14,6% considera utilizarla en un futuro inmediato (TicBeat, 28 de mayo de 2013). Aunque la proporción sea baja, y más baja aún, sin duda, en los medios de prensa, la necesidad de acumular y procesar datos es cada vez más ineludible, especialmente para los medios que publican en Internet. Solo las empresas que realmente procesan grandes volúmenes de datos podrán mejorar continuamente sus servicios, descubrir oportunidades y mitigar –en "tiempo real" – los riesgos que significa cualquier cambio en su formato, su selección de contenidos o su forma de comercialización, como ha sido demostrado por los cambios adoptados por grandes periódicos como *The New York Times*, *Newsweek* y *Daily Mail*, o revistas especializadas como *Wired*, mientras en España el diario *El Mundo* está probando *Orbyt*, su apuesta en lo que respecta a prensa online de pago.

El diario británico *Daily Mail* ha creado "centros de contenidos" asociados a las prioridades de su línea editorial. Su equipo está revisando constantemente los datos y los analiza para optimizar la experiencia del lector. Busca medir el éxito de los artículos incluso antes de su publicación, para saber qué elementos en la historia misma se pueden mejorar (@cdperiodismo, 28-04-2013).

No solo sobrevivir, sino ganar, implica necesariamente mirar la gran Red y aprender de la conducta y de los intereses de los lectores. Para los que están teniendo éxito,

"El Santo Grial ha sido mirar a la gente en la vida real, mirarla fuera del lugar de trabajo, ver cuando está por su cuenta, ver la comunicación como parte de su curso de vida en general, ¿cómo le ha cambiado la revolución electrónica y en qué?"

dice Kalev Leetaru, un miembro de la Facultad de Biblioteconomía y Documentación de la Universidad de Illinois. "Eso ha sido algo difícil de ver en el pasado debido a que no ha habido gran cantidad de datos." Pero ahora hay



más datos que la mayoría de las computadoras pueden procesar, procedentes de plataformas públicas de redes sociales como Twitter (Badger, 2013).

4. Las dificultades

El análisis de los datos masivos, como hemos visto, es cada vez más importante para el éxito de los medios, tanto en lo propiamente periodístico (periodismo de investigación o “de datos”) como en la gestión (adecuación a los lectores, según el modo de recepción –*smartphone*, tableta o PC–). Buenos mecanismos (aplicaciones) y personal capacitado son clave para ello.

Pero es necesario considerar hoy también la diversidad y movilidad de las fuentes tanto como la movilidad de las costumbres de los destinatarios. Las fuentes ya no pueden ser solamente el reporte directo o las agencias noticiosas. Hoy pasan muchas cosas en el mundo que se difunden por otros canales, principalmente las redes sociales. Se requiere no solo “estar al acecho” y recoger informaciones de interés, sino también estar preparado para verificar estas informaciones antes de retransmitirlas y poder hacerlo en muy corto tiempo. Una muestra de lo inconveniente de la precipitación y falta de verificación es lo que ocurrió con diversos medios locales en el caso del atentado en la maratón de Boston, en abril de 2013: hasta CNN y la Association Press (AP) cayeron en la trampa de la precipitación, quizás por seguir demasiado cerca lo publicado en Twitter (G.Lichfield).

Los antiguos lectores de diarios permanecían fieles a un medio determinado por mucho tiempo y sus preferencias por algunas secciones se mantenían también. Hoy puede haber una preferencia por los “*trending topics*” –que varían con mucha rapidez– y una fidelidad a ciertos temas más que a un medio de difusión (lo que explica el éxito de “agregadores” en equipos móviles). Captar y conservar público es mucho más complejo y la observación de las conductas se está transformando de hecho en una pesadilla para las grandes empresas si no cuentan con los recursos necesarios para un procesamiento ágil.

No hay, además, una manera fácil de coordinar los “datos en reposo” (en bases de datos, en este caso los contenidos que se publican) con los datos “en movimiento” (en este caso tanto las noticias que particulares publican en las redes sociales como las datos acerca de la conducta de los lectores) señala Kevin Jordan, a pesar de que el éxito de una compañía –particularmente un medio de comunicación–



dependa de ello. “No es poca cosa integrar grandes cantidades de datos de una variedad de fuentes. Es aún más difícil averiguar exactamente cómo traducir esa información en más visitas –en tiempo real–, clientes por cliente” reconoce Paul Dunay.

Es indispensable estar constantemente al día, en “tiempo real”, con lo que acontece tanto en las fuentes como en los destinos. En ambas dimensiones, la empresa requiere contar con aplicaciones (*software*) especializadas y permanentemente actualizadas. La “buena noticia” es que las hay. La “mala” es que el ajuste puede implicar un cambio importante en la forma de operar y en la “cultura interna” de la empresa. Hay que aprender a redefinirse a cada instante, principalmente sobre la base de los cambios de conducta de los destinatarios, que hay que aprender a observar e interpretar. Y, ante todo, hay que tener conciencia de que estos *Big Data* tienen una importancia fundamental.

Medios como *The New York Times* han aprendido de una primera –mala– experiencia y han asumido esta perspectiva. El *Wall Street Journal* se está adecuando a las redes sociales con una función específica para que sus lectores creen un perfil personal al estilo de *LinkedIn*. El diario digital *Mediapart*, que empezó en 2011, hoy es rentable gracias a una fórmula de destapes y debates, con acceso de pago, que ha dado excelentes resultados. El nuevo sitio *Fact Tank* podría destacarse por el hecho de dedicarse al análisis de las noticias del día con el objetivo de dar contexto y proporcionar más información a los usuarios. La web del diario británico *Daily Mail*, *MailOnline*, es la tercera mayor fuente de noticias en los Estados Unidos, justo detrás del *New York Times* y del *WashingtonPost*, según informó el sitio *10,000 Words* en abril pasado. Sus claves son la utilización del análisis de datos y la creación de videos sobre las historias que se vuelven populares ¡En todos ellos, la base es el análisis de datos!

Habrà que observar que ocurrirá con el *Chicago Sun-Times*, que despidió a sus 28 fotógrafos, ordenando a sus periodistas tomar fotos con sus iPhone. “Los consumidores de noticias esperan alta calidad cuando pagan por el periodismo profesional de una marca como el Chicago, que construyó su reputación estelar con la fotografía de noticias”, lamentó el fotoperiodista Robb Montgomery, en un artículo para la CNN. ¿Fue esta decisión basada en un estudio de los receptores o en un análisis de costos (que podría resultar erróneo a la larga)?

¿Cómo enfrentar estas dificultades? ¿Cómo avanzar? “Empezar a lo “grande” es empezar la casa por el tejado” nos dice Matti Keltanen, experto en diseño de servicios



digitales. Hay que saber primero cuales son los datos importantes a acumular, investigar cuales son las herramientas adecuadas para su análisis, juntar una primera cantidad – suficiente para un análisis– y sacar las lecciones de la experiencia. Lo más probable es que ya se disponga de alguna herramienta adecuada (p.ej. para analizar los datos noticiosos). “El pequeño secreto del *Big Data* es que ningún algoritmo puede decirte lo que es relevante o desvelarte su significado. Los datos se convierten entonces en otro problema que resolver. El enfoque del *Lean Data* –usar pocos datos– supone empezar con preguntas relevantes para el negocio y encontrar maneras de contestarlas a través de datos, en lugar de dedicarnos a cribar incontables conjuntos de datos” agrega Keltanen (TicBeat, 19 de mayo 2013).

También se puede “externalizar” la vertiente relativa al conocimiento de los destinatarios, al menos lo que no sea su comportamiento en línea cuando visitan los servicios del medio en Internet. Existen centros de estudio en las universidades o en empresas de análisis de mercado que pueden prestar el servicio, posiblemente en forma más económica y eficiente. Y podrían eventualmente asesorar también la puesta en marcha de procedimientos de seguimiento de los lectores en “tiempo real”.

5. Conclusión

La lección nos parece clara: hay que contar con aplicaciones adecuadas y con un personal bien entrenado para responder con rapidez a las dos componentes del “negocio” de la información hoy: agilidad en recoger y transmitir la noticias, y agilidad en detectar y adaptarse a las cambiantes conductas de los destinatarios.

“*Big Data* es, según apuntan las grandes firmas de análisis del mercado de TI, la próxima gran ola de la industria tecnológica. Una tendencia que (...) cambiará la forma de analizar el activo más importante que existe –la información– de empresas de todos los sectores económicos” (E. Macías, TicBeat, 28 de mayo 2013).

6. Referencias

[1] @cdperiodismo (2013): El análisis de datos y el video, claves para que un medio alcance la popularidad, *Clases de Periodismo* 28-04-2013, Recuperado el mismo día. <http://www.clasesdeperiodismo.com/2013/04/28/el-analisis-de-datos-y-el-video-claves-para-que-un-medio-alcance-la-popularidad/>



[2] Badger, E. (2013) How Twitter Is Changing the Geography of Communication, *The Atlantic Cities* 17-05-2013. Recuperado el mismo día. <http://www.theatlanticcities.com/technology/2013/05/how-twitter-changing-geography-communication/5601/>

[3] Colle, R. (2002): *Explotar la Información Noticiosa - Data Mining aplicado a la Documentación Periodística*, Ed. Depto de Biblioteconomía y Documentación, Universidad Complutense de Madrid, 266p.

[4] ---- (2012): Odisea digital 2001-2011: Cómo ha cambiado el entorno digital en los últimos años, *Communication Papers*, N° 1, pp. 88-97. http://communicationpapers.files.wordpress.com/2013/01/7c_olle_raymond.pdf

[5] Dunay, P. (2013): 6 Tips for Turning Big Data into Great Customer Experiences, *Socialmedia Today* 22 de mayo 2013, Recuperado el mismo día. <http://socialmediatoday.com/pauldunay/1484226/6-tips-turning-big-data-great-customer-experiences>

[6] Jiménez Cano, R. (2013): Las redes sociales sirven para hacer un mejor periodismo, *El País* 22 de mayo 2013. Recuperado el mismo día. http://tecnologia.elpais.com/tecnologia/2013/03/15/actualidad/1363378627_134732.html

[7] Jordan, K. (2013): Why Big Data Has Become a Nightmare, *The Tibco Blog* 18 de abril 2013, Recuperado el mismo día. <http://www.thetibcoblog.com/2013/04/18/why-big-data-has-become-a-nightmare/>

[8] Keltanen, M. (2013): Pongamos el big data a dieta, *TicBeat* 19 de mayo 2013, Recuperado el mismo día. <http://bigdata.ticbeat.com/pongamos-el-big-data-dieta/>

[9] Lichfield, G. (2013): Four ways the media failed in covering the Boston bombings, and one reason why, *Quartz* 20 de abril 2013, Recuperado el mismo día. <http://qz.com/76668/boston-marathon-and-the-media/>

[10] Macías, E. (2013): Big data propicia el auge de los científicos de datos en las organizaciones, *TicBeat* 28 de mayo 2013, Recuperado el mismo día. <http://bigdata.ticbeat.com/big-data-propicia-auge-cientificos-datos-organizaciones/>

[11] Montgomery, R. (2013): The photographers have left the building, *CNN* 6 de junio 2013, Recuperado el mismo día. <http://edition.cnn.com/2013/06/06/opinion/montgomery-sun-times-photogs/>

