

Cómo citar este artículo / Referencia normalizada

F López-Cantos (2015): “La investigación en comunicación con metodología BigData”. *Revista Latina de Comunicación Social*, 70, pp. 878 a 890.

<http://www.revistalatinacs.org/070/paper/1076/46es.html>

DOI: [10.4185/RLCS-2015-1075](https://doi.org/10.4185/RLCS-2015-1075)

La investigación en comunicación con metodología *BigData*

Communication research using *BigData* methodology

F López-Cantos [[CV](#)] [[ORCID](#)] [[GGS](#)] Profesor en Comunicación Audiovisual y Publicidad. Departamento de Ciencias de la Comunicación. Universitat Jaume I de Castellón (España)
flopez@uji.es

Abstract

[ES] Introducción: Las tecnologías digitales están facilitando nuevas herramientas y metodologías de análisis en el ámbito de la comunicación, al igual que en otras áreas de investigación. En este trabajo abordamos los aspectos relacionados con la investigación [1] en comunicación con la metodología *Big Data*, que está creando muchas expectativas y está dando buenos resultados en otros ámbitos científicos. **Metodología:** Tomamos como corpus de trabajo un conjunto de artículos y utilizamos sobre esa muestra las técnicas de análisis y representación de *Big Data* como metodología de análisis de contenido para valorar su pertinencia y efectividad. **Conclusiones:** Como resultado de nuestro estudio podemos concluir que esta metodología en nuestra área de conocimiento puede ser de utilidad pero tiene una efectividad limitada.

[EN] Introducción: Las tecnologías digitales están facilitando nuevas herramientas y metodologías de análisis en el ámbito de la comunicación, al igual que en otras áreas de investigación. En este trabajo abordamos los aspectos relacionados con la investigación en comunicación con la metodología *Big Data*, que está creando muchas expectativas y está dando buenos resultados en otros ámbitos científicos. **Metodología:** Tomamos como corpus de trabajo un conjunto de artículos y utilizamos sobre esa muestra las técnicas de análisis y representación de *Big Data* como metodología de análisis de contenido para valorar su pertinencia y efectividad. **Conclusiones:** Como resultado de nuestro estudio podemos concluir que esta metodología en nuestra área de conocimiento puede ser de utilidad pero tiene una efectividad limitada.

Keywords

[ES] Big Data; minería de datos; análisis de contenido; metodología; epistemología; representación científica; investigación en comunicación.

[EN] Big Data; minería de datos; análisis de contenido; metodología; epistemología; representación científica; investigación en comunicación.

Contents

[ES] 1. Introducción. 2. Metodología. 3. Resultados. 4. Conclusiones y discusión. 5. Bibliografía.

[EN] 1. Introduction. 2. Methodology. 3. Results. 4. Conclusions and discussion. 5. Bibliography.

Traducción de **Yuhanny Henares**

1. Introducción

Estamos sujetos a las cifras, a los datos, tanto en lo relativo a la adopción de estándares de obligado cumplimiento que faciliten la objetivación y valoración de nuestros logros académicos para determinar el curso de nuestra carrera profesional, como a la hora de emprender los propios trabajos de investigación y promocionar su publicación en revistas de impacto nacional e internacional que, como no, han de estar preferentemente basados en investigaciones empíricas y presentar resultados de investigación cuantificables. “Numbers, numbers, numbers...”, recitaba a modo de crítica a esta letanía contemporánea que nos invade Bruno Latour (2009), poniendo en cuestión esta obsesión actual por la cuantificación que está acabando por impregnar todos los ámbitos públicos y privados y la propia cultura contemporánea.

Y en este nuevo entorno digital contemporáneo de respuesta inmediata y global y altamente interconectado y mediado por algoritmos matemáticos que impregnan todos los rincones de nuestra existencia cotidiana se genera, como consecuencia lógica, una avalancha de datos que se acumulan con ritmo exponencial *ad infinitum*. Y en esta vorágine caótica sin sentido aparece, como suele ocurrir, una nueva criatura de la razón pura que, *mutatis mutandi*, promete apaciguar el caos existente, e incluso, aportar nuevas perspectivas de conocimiento pero que, sobre todo y en primer lugar, se ha convertido en un nuevo gran negocio propio de la era digital que cotiza con plenos derechos en el NASDAQ: las *nuevas* tecnologías de extracción, representación y análisis de *Big Data*.

Existe un creciente número de trabajos que se preguntan por la utilidad de las nuevas técnicas en diversas áreas científicas (Leonelli, 2014; Taylor, Schroeder y Meyer, 2014), tratan aspectos epistemológicos (Kitchin, 2014; Raghavan, 2014), metodológicos (Burrows y Savage, 2014; Murthy y Bowman, 2014) o éticos (Zwitter, 2014; Lyon, 2014; Gurevitch, 2014). Y algunos trabajos se ocupan específicamente de el impacto de su uso en el ámbito de la comunicación (Schroeder, 2014; Penney, 2014; D’Heer y Verdegem, 2014; Fischer, 2014), determinando la notoria significancia que está adquiriendo el tratamiento de grandes bases de datos con técnicas de análisis de contenido.

Una búsqueda simple del término *Big Data* en *Google* arroja más de 765 millones de entradas. Sin embargo, y a pesar de estas cifras, parece que esta nueva tendencia de investigación en Ciencias Humanas y Sociales tiene escaso impacto en nuestra área y al parecer nos está resultando del todo

ajena, paradójicamente al tiempo que sufrimos la progresiva exigencia de *cuantificación* de nuestro trabajo.

A priori, sin duda, el gran atractivo de poder trabajar con *Big Data* está potenciado con las nuevas posibilidades que ofrecen las actuales tecnologías digitales para su tratamiento y para la obtención de resultados provechosos de manera rápida y económica, y las mejoras que puedan derivarse del análisis de los fenómenos estudiados.

En cualquier caso, cada nuevo tiempo trae consigo nuevas mitologías y por ello hemos incluido en el título de este trabajo la posibilidad de que se abra una nueva época, la *Era Big Data*, ya que como bien sabemos y citando de nuevo a Latour (2009): “Change the instruments, and you will change the entire social theory that goes with them”.

No hemos encontrado ni un solo artículo en que aparezca siquiera tal término en las revistas de más impacto en nuestra área de conocimiento recogidas en el directorio IN-RECS 2011, situación que contrasta como decimos con el incremento de la presencia de esta área de conocimiento en las revistas académicas del ámbito de la comunicación internacionales.

Por ello nos parece pertinente analizar hasta dónde llegan esas mejoras que se nos promete en su aplicación en nuestra área de conocimiento, las Ciencias de la Comunicación, y determinar hasta dónde esta tendencia en investigación puede ser de utilidad como metodología de análisis en los

2. Metodología

Para abordar esta investigación el método de trabajo que utilicemos deberá de cumplir una serie de premisas para que nos resulte válido y, sin duda, nos alejará algo de la ortodoxia y las prácticas habituales.

Partimos de una hipótesis sencilla: hay una corriente de investigación que está emergiendo en Ciencias Sociales y Humanas representada en un conjunto de 117 textos que hemos recuperado con la búsqueda del término *Big Data* en *Sage Journals*.

Hemos llegado hasta ahí, obviamente, porque en algún momento en nuestras lecturas nos hemos encontrado con el término, e incluso hemos ojeado algún artículo sobre ello en un primer acercamiento a nuestro objeto de estudio, y queremos emprender una investigación más exhaustiva, para lo cual realizamos una búsqueda inicial en varias bases de datos, SageJournal en nuestro estudio. Podríamos haber utilizado otras bases de datos (JCR, Scopus, etc.) pero lo hemos hecho solo en *Sage* porque para nuestros fines resulta suficiente como muestra para testear el análisis de las técnicas de análisis de BigData y lo que pretendemos es sólo extraer conclusiones respecto a la validez de la aplicación de la metodología de minería de datos en nuestra área, no analizar el impacto del BigData en el conjunto de las investigaciones en comunicación, aspecto que para su adecuado análisis se debería de tratar de manera específica y no está en las pretensiones de este trabajo.

En nuestra búsqueda el resultado son un total de 117 referencias que conforman nuestro corpus de trabajo para testear la metodología BigData. Y ahora es cuando difiere el método de investigación que vamos a utilizar respecto a los habituales.

Advanced Search

Advanced searches of SAGE Journals use a signature fielded Boolean system. Use this award-winning search tool to construct a query specifying your terms and their logical relationships using the Boolean operators AND, OR, and NOT. [Learn more](#) about advanced searches on SAGE Journals.

(big data) and ()) All fields
and () and () All fields
+ Add Row Search Clear All Fields

Search Within
 SAGE Journals Available to Me All SAGE content My Favorite Journals
 Select from a list of disciplines Select from a complete list of journals

<input type="checkbox"/>	Health Sciences
<input type="checkbox"/>	Life & Biomedical Sciences
<input type="checkbox"/>	Materials Science & Engineering
<input type="checkbox"/>	Social Sciences & Humanities
<input type="checkbox"/>	Anthropology & Archaeology
<input checked="" type="checkbox"/>	Communication & Media Studies
<input type="checkbox"/>	Criminology & Criminal Justice

Ilustración 1 Búsqueda en Sage Journal

Si hiciéramos una investigación con la metodología convencional lo siguiente sería seleccionar de la lista de referencias las que nos pareciesen más adecuadas y hacer una primera lectura de las mismas. A partir de ahí, iríamos recopilando otra bibliografía desde la inserta en cada uno de los artículos que fuésemos leyendo y, de manera manual o con el uso de un procesador de textos, tomaríamos anotaciones para su posible uso, en una suerte de análisis de contenido que nos permitiera profundizar y conocer cada vez más en relación a nuestro objeto de estudio.

Después de varias lecturas y relecturas, algunos descartes y ampliaciones, y las consiguientes nuevas anotaciones y actualizaciones de conceptos, podríamos ir empezando a elaborar un discurso argumentativo en el que presentar nuestro particular análisis del corpus que conforma nuestro objeto de estudio. Tras varios borradores y reelaboraciones obtendríamos finalmente un texto adecuado para su divulgación en el que, de manera ordenada, plasmaríamos nuestro discurso de acuerdo a los cánones establecidos y siguiendo las indicaciones de estilo y composición de la revista elegida para, con suerte, poder ser estimada positivamente su publicación.

El método de trabajo que vamos a emplear es en ciertos aspectos similar pero, sin embargo, en otros es radicalmente distinto. Y es que no vamos a leer nada, no vamos a leer ni una sola línea de texto. Esa tarea la vamos a dejar en manos de estas tecnologías BigData que tanto nos promete. Y veremos si se cumplen las expectativas.

Vamos a utilizar un paquete de software que hemos seleccionado entre los accesibles gratuitamente, *QDAMiner*. Para trabajar con datos estructurados se puede utilizar, por ejemplo, *RapidMiner* o el mucho más sofisticado software *R*, con multitud de posibilidades de análisis y gráficas, pero nuestra apuesta es la de trabajar en minería de datos tal cual, es decir, con datos sin estructura alguna para ver las posibilidades reales de las tecnologías *Big Data*.

Trabajar con bases de datos ya elaboradas ofrece más posibilidades de análisis, pero en general son datos no estructurados los que nos encontramos cuando procedemos al análisis de contenido en cualquier trabajo de investigación en comunicación, por ejemplo para analizar la prensa escrita, y por ello nos resulta de mayor interés trabajar con datos no estructurados y por ello nuestra selección ha sido el software *QDAMiner*.

Vamos analizar la funcionalidad de este software, y a determinar si toda la ingeniería informática que hay detrás del término *Big Data* sirve de algo o no nos sirve de nada en la mayoría de trabajos de investigación que emprendemos: podríamos leer y analizar cien o doscientos textos pero no cien mil, que es lo que promete el software *Big Data*.

Las cuestiones a las que queremos responder son las siguientes:

En primer lugar, ¿la extracción de datos por la fuerza bruta sin previa selección y filtro por nuestra parte nos sirve de algo? Y en caso de que esa utilidad sea solo parcial ¿se pueden aplicar técnicas específicas que mejoren la calidad de los datos? y ¿en qué medida implican igualmente nuestra intervención?

En segundo lugar, ¿una vez obtenidos los datos con qué criterios los consideramos adecuados para nuestro estudio?

Y, finalmente y una vez validados ¿que forma de presentación resultará más adecuada para su análisis?

Y más todavía, ¿a partir de ese primer análisis se puede volver de nuevo al principio y recorrer de nuevo el camino para profundizar en la investigación y/o obtener nuevas perspectivas?

Para responder a estas cuestiones, vamos a evaluar los resultados obtenidos en las distintas fases del procedimiento metodológico de análisis con técnicas BigData, siguiendo las pautas que presentamos a continuación.

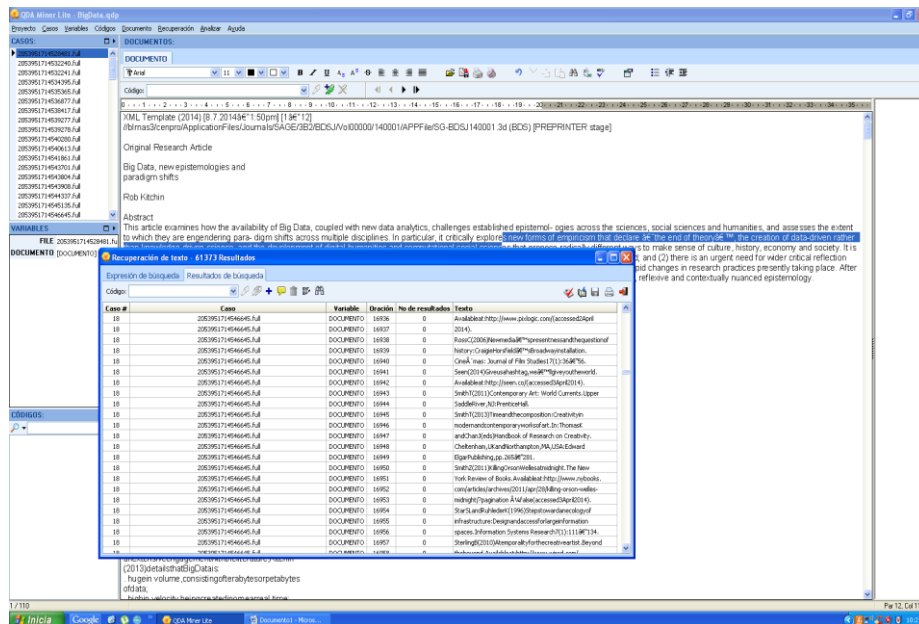
3. Resultados

En primer lugar, y una vez almacenados todos los artículos en formato *pdf* en nuestro ordenador, realizamos a importación de archivos a *QDAMiner*. Se almacenan un total de 110 archivos, los accesibles a nuestra suscripción en *Sage Journal*, y la importación hace por lotes y sin mayores complicaciones en menos de un minuto. Se genera un proyecto *QDAMiner* a partir de la carpeta en que están localizados en el que se incluyen todos los documentos originales convertidos a texto plano.

Posteriormente, y una vez en *QDAMiner*, procedemos a emprender el análisis de contenido. Esta fase sí resulta mas controvertida y nos empezamos a encontrar con problemas y empezamos a poner en cuestión la utilidad del software y las promesas de la *Era Big Data*. Si hacemos búsquedas de los datos brutos para segmentar el contenido podemos utilizar recuperar datos al nivel de lo que se denomina en este software *oración*. Por el contrario, si establecemos palabras clave de búsqueda que pudieran ser de interés nos enfrentamos a los problemas típicos de categorización semántica, y

además estaríamos entonces necesariamente leyendo el texto con el sesgo que inevitablemente introduciría nuestra particular lectura.

Procedemos a evaluar la eficacia de la primera opción, ya que hemos decidido evitar la lectura y categorización, y exploramos la herramienta de recuperación de texto. Obtenemos un total de 74.894 segmentos al nivel de unidad mínima, que es el de oración, para el total de 110 documentos (o *casos* como se les denomina). Desconocemos el algoritmo que utiliza el software para determinar estas unidades, a juzgar por los resultados la segmentación en bruto no parece que nos sirvan para nada.



Vamos a explorar, en consecuencia, búsquedas más específicas que tengan alguna utilidad. Pero entonces la pregunta que inmediatamente nos hacemos es, ¿y qué texto buscamos?, al fin y al cabo, sea el que sea va a constituir un *a priori kantiano* y no va a devolvernos más resultados que aquellos que ya sabemos previamente que pueden existir, y eso, sin duda, hará que perdamos el resto de la información que pudiéramos extraer y que, al menos hipotéticamente, pudiera ser también de algún de interés.

Pero no hay más remedio, es necesario estructurar la información de algún modo más eficiente. Entonces ¿qué hacer? ¿esto es todo lo que ofrece este software y promete la tecnología *Big Data*? Entendemos que no, sólo estamos en las primeras fases del análisis, y realmente hasta el momento no hemos dedicado mas que unos pocos minutos de nuestro tiempo a todo ello, debe haber algo más, pero, ¿qué más podemos hacer?

Estamos en la versión Lite, esto es gratuita, y hay muchas opciones del software que no son accesibles pero ¿realmente vamos a necesitar esas funcionalidades o es más de lo mismo? Revisamos el apartado de ayuda de la aplicación y, efectivamente, es más de lo mismo: mayor versatilidad a la hora de cruzar datos y explotar los ya existentes pero no vamos a poder obtener ninguna información útil a no ser que codifiquemos de forma manual, al menos mínimamente, los textos a partir de un

tesauro con su estructura jerárquica y correspondencias que, también, hemos de crear nosotros mismos.

Y la pregunta es, ¿de qué nos sirve el software y *Big Data* entonces? De bien poco, sólo para informatizar un proceso manual que luego podremos explotar pero que, al fin y al cabo, tendremos que realizar como siempre se ha hecho, por nosotros mismos. Y no parece que entonces haya más que humo tras la etiqueta *Big Data* pues nada se puede hacer si se ha de etiquetar manualmente cuando nos enfrentamos con grandes volúmenes de datos, ¿o sí?

Y, sí, algo más se puede hacer utilizando una técnica que combina el trabajo manual y el automático y que consiste en codificar uno o varios textos manualmente para, después y de manera automatizada, extender la codificación al resto del corpus de estudio. Es decir, hay que seleccionar uno o varios artículos representativos y analizarlos y codificarlos concienzudamente según nuestros intereses para, posteriormente, explorar el resto del corpus.

Esto nos supondrá un trabajo de lectura que queríamos ahorrar y la compleja y dura tarea de elaboración de un tesauro que mantenga una mínima coherencia semántica y, en cualquier caso, el inconveniente ya expuesto mas arriba: la selección de los artículos representativos, y su categorización, de manera irremediable introducirá un sesgo en nuestra investigación. Pero no hay más opción.

La selección de los documentos de muestra para su codificación la podemos resolver de dos maneras. Bien intentando determinar a partir de una lectura rápida de títulos de los artículos y resúmenes cuáles serían los textos representativos, para lo que tendríamos también que determinar con qué criterios lo serían. O bien, extrayendo una muestra estadísticamente significativa al azar. La primera opción no correspondería al trabajo propio con *Big Data* pues sólo sería válida para corpus de tamaño reducido y, en cualquier caso, estaría sesgada por el a priori kantiano que venimos señalando; y la segunda adolece del sesgo propio a la selección estadística de la muestra.

A partir de este momento, además, se abre un sinfín de posibilidades pues tenemos que plantearnos el marco teórico desde el que vamos a emprender la elaboración del tesauro para codificar los artículos seleccionados. Y cada vez nos alejamos más de nuestros datos en bruto, introducimos mayor sesgo en los datos que recuperamos y perdemos más información original por el camino.

Contando con ello, procedemos a realizar una operación de codificación sencilla a partir de una selección de quince artículos de muestra al azar, los rastreamos en una lectura rápida de título y resumen y de ellos seleccionamos cinco títulos que nos parecen más adecuados. En ellos analizamos los resúmenes y construimos un pequeño tesauro utilizando una combinación de técnica abajo-arriba / arriba-abajo, es decir, vamos a establecer categorías a priori según nuestros intereses investigadores y vamos a establecer otras a partir de la lectura de los resúmenes de esa muestra de artículos.

Los textos seleccionados de muestra para su codificación son los siguientes:

- *Complementary Social Science? Quali-quantitative experiments in Big Data World.*
- *Big Data and The Brave New World of Social Media Research.*
- *Big Data, new epistemologies and paradigm shifts.*

- *Emerging practices and perspectives on Big Data analysis in economics: Bigger and Better or more the same?*
- *Theses on the Philosophy of History: The Work of Research in the Age of Digital Searchability and Distributability.*

Y para proceder al análisis de sus resúmenes establecemos previamente las siguientes categorías en nuestro breve tesoro, que iremos construyendo *ad hoc*. Nos interesa encontrar todo lo relacionado con las *metodologías de trabajo con Big Data*, la *aplicación a la investigación en comunicación*, lo que puede afectar a las actuales *teorías de la comunicación*, aquellos textos que traten cuestiones relacionadas con la *epistemología* y, finalmente, aquellos que plantean *cuestiones éticas*.

A partir de ellos y del análisis de los textos de muestra generaremos códigos con los términos significativos encontrados y realizaremos búsquedas en el conjunto del corpus para extender nuestro análisis al total de los documentos que versan sobre distintos aspectos relativos a *Big Data*. Por ejemplo, para nuestros propósitos incluiremos todas las búsquedas de términos como *methodology*, *empirical*, *ethics* o *theory*, en inglés ya que en el corpus no hay una sola revista en otra lengua.

Realizamos la búsqueda de esos términos y codificamos los segmentos en que aparecen en sus correspondientes categorías. Con todo ello lo que obtenemos es la siguiente tabla de frecuencias de aparición que exportamos a un fichero Excel:

Categoría	Código	Descripción Cuenta	% Códigos	Casos	% CASOS
metodologías	methodology	23	3,70%	19	17,30%
teorías	theory	213	34,60%	64	58,20%
ética	ethics	78	12,70%	29	26,40%
aplicaciones	empirical	108	17,50%	42	38,20%
BigData	Big Data	194	31,50%	62	56,40%

Esta lista podría ser, obviamente, mucho más larga pero no es necesario para lo que nos ocupa y bien sirve de muestra.

Todo lo que hemos conseguido después de aplicar todas las posibilidades que nos ofrece el software de minería de datos ha sido una tabla de frecuencias de aparición. Para su elaboración podríamos establecer filtros, variables durante las búsquedas, cruzar datos e incluso generar tablas multidimensionales utilizando algoritmos sofisticados. Nuestra tabla podría ser todo lo compleja que quisiéramos pero, al fin y al cabo, no sería más que eso, una tabla de frecuencias de aparición.

Finalmente, y con los resultados obtenidos utilizando todas las herramientas metodológicas que nos ofrece BigData para el análisis de contenido, que son más bien pobres y muy sesgados, lo que podemos hacer es elaborar representaciones gráficas de los datos que permitan su análisis de manera sencilla por los investigadores o, también, para su presentación al público en general.

methodology **theory** ethics
 empirical **Big Data**

En un nivel mayor de concreción podríamos, incluso, desarrollar nuestro análisis de los textos y profundizar, por ejemplo, en las distintas teorías que aparecen en nuestro corpus de estudio:

Categoría	Código	Descripción	Cuenta	% Códigos	Casos	% CASOS
teorías	theory		213	25,40%	64	58,20%
teorías	semiotics		5	0,60%	4	3,60%
teorías	cultural studies		28	3,30%	17	15,50%
teorías	meme		191	22,70%	5	4,50%
BigData	Big Data		194	23,10%	62	56,40%

theory semiotics **meme**
 cultural studies **Big Data**

En cualquier caso, y como podemos ver, para obtener unos mínimos resultados, aunque gráficamente puedan sofisticarse mucho más y llegar a ser muy seductores, se requiere de nuestra intervención, que si bien resulta mas o menos eficiente con pequeñas colecciones de textos, a medida que aumenta el tamaño y complejidad del corpus de trabajo empieza a no serlo, y se impone un sesgo que hace inválidos los resultados obtenidos.

Aunque en este gráfico la frecuencia de aparición de la palabra meme es 191 esa cifra, obviamente, sólo indica eso, que aparece la palabra tal número de veces, pero no se analicen desde este marco epistemológico los aspectos teóricos asociados a *Big Data* en los textos de nuestro corpus, en los que no se emprende con profundidad tal análisis en apenas cinco de ellos, mientras que se analizan desde el punto de vista de la semiótica en cuatro, a pesar de que la aparición del término es significativamente menor. Pero para obtener estos resultados debemos leer todos los textos, lectura que resulta imposible cuando nos enfrentamos ante una colección de textos mayor a unas decenas.

4. Conclusiones y discusión

En este sentido, y en definitiva, las metodologías y tecnologías que nos ofrece esta supuestamente *nueva Era Big Data* no son nada novedosas y ya son habituales desde hace mucho tiempo en el ámbito de las ciencias documentales y la biblioteconomía desde que hay disponibles tecnologías informáticas para el tratamiento de textos y la conformación de bases de datos. Nada nuevo, sólo ordenadores más rápidos y computación distribuida, pero eso no es suficiente para un cambio de paradigma en la investigación.

Con el término *Big Data* lo que se está promoviendo en realidad son las tecnologías que aumentan la capacidad de almacenamiento, recuperación y visualización de datos y permiten enfrentarse a grandes cantidades de información con mayor rapidez y a costes razonables. Se promete gran rapidez en el análisis de datos y, especialmente, el aumento de la capacidad de síntesis con la utilización de herramientas gráficas que muestran los resultados de manera rápida y visualmente atractiva. Pero esto no es nada novedoso, sólo que históricamente la recopilación y análisis de datos y la presentación visual de los resultados de su análisis suponía una labor tediosa y que consumía grandes recursos y tiempo, y hoy se hace de manera mucho más eficiente y se obtienen resultados gráficos más espectaculares, pero en absoluto se está desarrollando un nuevo paradigma de investigación en sí mismo.

El problema que generan estas promesas de mejora es que se se continúan fundamentando en un neopositivismo impregnado de atributos trasnochados asociados a la objetividad. Y esta nueva mitología que se está construyendo alrededor del término *Big Data* y las supuestas mejoras que introduce en la generación y avance del conocimiento parece obviar las sólidas críticas que recibió el neopositivismo por parte de Khun (1962), con sus propuestas de relativismo de los paradigmas científicos, o el más radical Feyerabend (1975), con su anarquismo epistemológico, entre otros.

Con toda probabilidad el deslumbramiento que produce la supuesta *New Age* está propiciada por la potencia gráfica de las herramientas que se utilizan y que facilita la elaboración de espectaculares imágenes estáticas o dinámicas que multiplican en varios órdenes la capacidad de penetración de esta nueva mitología en todos los ámbitos. Las aplicaciones que ofrece la computación distribuida para el tratamiento rápido de datos estructurados en, por ejemplo, investigación médica y farmacológica, economía o análisis de redes, entre otros muchos campos de aplicación son notorias, pero no por ello han desaparecido los problemas inherentes a la representación del conocimiento.

Son bien conocidas las dificultades de acotación de los campos semánticos y la imposibilidad de evitar los sesgos propios de la selección de contenidos en tanto que inherentes al lenguaje verbal para la elaboración de bases de datos en los ámbitos de las Ciencias de la Documentación, véase por ejemplo López Yepes (1995).

En lo relativo al lenguaje audiovisual, las limitaciones de la representación de datos ya habían sido puestas en entredicho en el conocido trabajo de Anscombe (1973), pero el problema actual es que las nuevas promesas van más allá de la mera mejora en la eficacia en la representación y pretenden ser la culminación de un viejo sueño que ya empezaron a gestar aquellos navegantes-geógrafos que impulsaron la cartografía de los nuevos territorios desde que se empezaron a imprimir mapas (Ford, 1992: 131). Y de un siglo a esta parte, se ha ido transitando de la ilustración de lo terrenal a la representación de lo inmaterial, con los problemas inherentes a ello (Lynch, 2006; ref. propia 2013),

en lo que Heidegger vino a denominar hace unas décadas *The Era of the World Picture*, en la que el fin último de la modernidad acababa con la conquista del mundo en tanto que imagen (cfr. Gross, 2006), y cuyas implicaciones y limitaciones se han explorado con profundidad y análisis crítico durante las últimas décadas, como por ejemplo en Manovich (2002), entre otros.

De manera que lo cierto, y ya para concluir, es que la promesa de una mejora en la representación del conocimiento, o incluso de un cambio de paradigma a esta supuesta *Nueva Era Big Data*, no es más que otra versión de la ya conocida mitología tecno-neopositivista contemporánea que se ha venido gestando en las últimas décadas alrededor de Silicon Valley (M. Gregg, 2015), hoy promovida con el inestimable impulso que ha supuesto la representación espectacular de datos que permiten las capacidades de computación actuales.

En todo caso, los intereses tecno-económicos que promueven la supuesta nueva era *Big Data* se suma a las cuestionables propuestas que nos trae la cultura digital contemporánea, con la cuestionable imposición de la cuantificación en todos los ámbitos con nuevas falsas viejas promesas que sólo alimentan ancestrales mitologías.

En lo relativo al ámbito de las Ciencias de la Comunicación, entendemos que un software de minería de datos como el evaluado en este trabajo o similar, puede ser de cierta utilidad para facilitar el análisis de contenido en revisiones de carácter bibliográfico como la presente y, en general, y en investigaciones empíricas en la que se precise el análisis de corpus de datos textuales de cierto tamaño. Pero no es ninguna panacea y se ha de tener en cuenta que los inevitables sesgos que se van introduciendo a lo largo de todo el proceso van siendo mayores en relación al tamaño del corpus analizado, y por mucho cuidado que se tenga en las cuestiones metodológicas descritas supra necesariamente los resultados que se obtienen son “multifaceted and multitrue” (B. Foucault & I. Meirelles, 2015).

5. Nota

1 Para una recopilación de las distintas técnicas y metodologías de análisis en comunicación se puede consultar L. Vilches (Coord.) (2011).

6. Bibliografía

FJ Anscombe (1973): "Graphs in Statistical Analysis," en *American Statistician*, 27, pp. 17-21.

R Burrows and M. Savage (2014): "After the crisis? Big Data and the methodological challenges of empirical sociology", en *Big Data & Society*, April–June, pp. 1–6.

E D’heer and P. Verdegem (2014): "Conversations about the elections on Twitter: Towards a structural understanding of Twitter’s relation with the political and the media field", en *European Journal of Communication*, Vol. 29(6), pp. 720–734.

- PK Feyerabend (1975): *Tratado contra el método. Esquema de una teoría anarquista del conocimiento*. Madrid: Tecnos, ed. 1981.
- E Fisher (2015): “‘You Media’: audiencing as marketing in social media”, en *Media, Culture & Society*, Vol. 37(1) 50–67.
- B Foucault & I Meirelles. (2015): “Visualizing Computational Social Science: The Multiple Lives of a Complex Image”, en *Science Communication*, Vol. 37(1), pp. 34-58.
- BJ Ford (1992): *Images of Science: A History of Scientific Illustration*. London: British Library.
- Gregg, M. (2015): “Inside the Data Spectacle”, en *Television & New Media*, Vol. 16(1), pp. 37–51.
- A Gross (2006): “The Verbal and the Visual in Science: A Heideggerian Perspective”, en *Science in Context*, 19 (4), pp. 443-474.
- L Gurevitch (2014): “Google Warming: Google Earth as eco-machinima”, en *Convergence*, Vol. 20(1), pp. 85–107
- B Kitchin (2014): “Big Data, new epistemologies and paradigm shifts”, en *Big Data & Society*, April–June, pp. 1–12.
- TS Kuhn (1962): *The Structure of Scientific Revolutions*. Chicago: Chicago University Press
- B Latour (2009): “Tarde’s idea of quantification”, en *The Social after Gabriel Tarde: Debates and Assessments*. London: ed. M. Candea, Routledge, pp. 145–162.
- S Leonelli (2014): “What difference does quantity make? On the epistemology of Big Data in biology”, en *Big Data & Society*, April–June pp. 1–11.
- M Lynch (2006): “The Production of Scientific Images. Vision and Re-Vision in the History, Philosophy, and Sociology of Science”, *Visual Cultures of Science: rethinking representational practices in knowledge building and science communication*. Hanover, N.H.: Dartmouth College Press., p. 26 y ss.
- D Lyon (2014): “Surveillance, Snowden, and Big Data: Capacities, consequences, critique”, en *Big Data & Society*, July–December, pp. 1–13.
- L Manovich (2002): “The Anti-Sublime Ideal in Data Art”, disponible en <http://manovich.net/index.php/projects/data-visualisation-as-new-abstraction-and-anti-sublime>.
- D Murthy y S Bowman (2014): “Big Data solutions on a small scale: Evaluating accessible high-performance computing for social research”, en *Big Data & Society*, July–December, pp. 1–12.
- J Penney (2014): “Motivations for participating in ‘viral politics’: A qualitative case study of Twitter users and the 2012 US presidential election”, en *Convergence*,
- P Raghavan (2014): “It’s time to scale the science in the social sciences”, en *Big Data & Society*, April–June, pp 1–4.

R Schroede. (2014): “Big Data and the brave new world of social media research”, en *Big Data & Society*, July–December, pp. 1–11.

L Taylor; R Schroeder y E Meyer (2014): “Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same?”, en *Big Data & Society*, July–December, pp. 1–10.

L Vilches (coord.) (2011): *La investigación en comunicación. Métodos y técnicas en la era digital*. Barcelona: Gedisa.

A Zwitter (2014): “Big Data ethics”, en *Big Data & Society*, July–December, pp. 1–6.

Cómo citar este artículo / Referencia normalizada

F López-Cantos (2015): “La investigación en comunicación con metodología BigData”. *Revista Latina de Comunicación Social*, 70, pp. 878 a 890.

<http://www.revistalatinacs.org/070/paper/1076/46es.html>

DOI: [10.4185/RLCS-2015-1075](https://doi.org/10.4185/RLCS-2015-1075)

- En el interior de un texto:

... F López-Cantos (2015: 878-890)...

... F López-Cantos (2015: 878-890)...

Artículo recibido el 20 de octubre de 2015. Aceptado el 27 de noviembre.

Publicado el 21 de diciembre de 2015.