

# *Big data, medios digitales y sociedad: retos para la investigación en comunicación*

## *Big data, digital media and society: challenges for communication research*

---

Roberto Alejandro López Novelo\*

Universidad Anáhuac México

Av. Universidad Anáhuac núm. 46, Col. Lomas Anáhuac,  
Huixquilucan, Estado de México, C.P. 052786, México

roberto.a.lopez@anahuac.mx

<https://orcid.org/0000-0003-0978-9784>

Editor: Rogelio del Prado Flores

Fecha de recepción: 24 de junio de 2021

Fecha de aceptación: 25 de agosto de 2021

<https://doi.org/10.36105/stx.2021edesp50aniv.07>

### RESUMEN

Actualmente la investigación en comunicación se enfrenta a un nivel de complejidad nunca antes visto para la disciplina, esto se debe entre otros fenómenos, a la presencia de Internet y a una serie de entornos digitales a través de los cuales los usuarios generan millones de datos sobre temas políticos, sociales, económicos, culturales, educativos, publicitarios, mercadológicos, etc., en formatos diversos que están a disposición para ser analizados y aprovechados por investigadores de diversas áreas. El reto para los investigadores es en principio visualizar a Internet como un espacio social en el que personas, instituciones, organizaciones e incluso sistemas informáticos generan y comporten datos; además, conocer los métodos de investigación novedosos e innovadores que existen en la actualidad para realizar el análisis formal de dichos datos. Uno de los métodos que más se utilizan en la actualidad es el *big data*, que permite al investigador en comunicación realizar investigaciones novedosas.

**Palabras clave:** big data, medios digitales, internet, investigación, comunicación.

\* Coordinador del área académica de investigación de la Facultad Comunicación de la Universidad Anáhuac México. Doctor en Ciencias Políticas y Sociales con Orientación en Comunicación por la Universidad Nacional Autónoma de México (UNAM), Maestro en Comunicación (UNAM) y Licenciado en Ciencias de la Comunicación (UNAM). Estancia posdoctoral en Big Data y Comunicación Política en el Observatorio de los Contenidos Audiovisuales (OCA), Universidad de Salamanca, España.

## ABSTRACT

Currently, communication research faces a level of complexity never seen before for the discipline, this is due, among other phenomena, to the presence of the Internet and a series of digital environments through which users generate millions of data on topics political, social, economic, cultural, educational, advertising, marketing, etc., in various formats that are available to be analyzed and used by researchers. It is in principle to visualize the Internet as a social space in which people, institutions, organizations and even computer systems generate and carry data; and, to know the new and innovative research methods that currently exist to carry out the formal analysis of data. One of the most widely used methods today is Big Data, which allows communication researchers to carry out innovative research.

**Keywords:** big data, digital media, internet, research, communication.

## INTRODUCCIÓN

En el contexto actual, la investigación en comunicación se enfrenta a un nivel de complejidad nunca antes visto para la disciplina, esto se debe entre otros fenómenos, a la presencia de las Tecnologías de Información y Comunicación (TIC) en casi todas las actividades del quehacer humano y a la aparición de un nuevo espacio sociotécnico y sociocultural (Internet) al que acuden cotidianamente millones de personas en todo el mundo para realizar diversas actividades como: búsqueda de información, participación en movimientos sociales, actividad política, procesos educativos, tan solo por mencionar algunos ejemplos. Internet ofrece una serie de entornos digitales que permiten a los usuarios no solo consumir información, sino que ahora les da la posibilidad de crear contenidos de diversa índole y generar redes complejas y densas, a través de las cuales expresan sus intereses en diversas áreas de interés.

La acumulación de información en redes sociales representa millones de datos que están a nuestra disposición y pueden ser utilizados para fines de investigación, siempre y cuando el investigador conozca los nuevos métodos de innovación e investigación disponibles para hacer frente a los retos que plantea el contexto digital. El reto es entender que las TIC y particularmente las redes sociodigitales que se generan y crecen exponencialmente todos los días en Internet, facilitan la comunicación entre diferentes elementos capaces de procesar información proveniente de personas y máquinas, y la obtención de datos a través de los procesos de informatización de la sociedad, a partir del uso de plataformas diversas, flexibles y exponenciales. Este fenómeno trae como consecuencia que aumente el nivel de

complejidad de la interacción social y su tratamiento sea cada vez más complejo y difícil con las metodologías de investigación tradicionales.

Actualmente existen metodologías que permiten hacer un análisis formal y profundo de estos millones de datos que son generados por los actores sociales en Internet, entre ellas, el *big data*. El *big data* permite al investigador en comunicación identificar patrones y obtener consecuencias significativas en el estudio de fenómenos de carácter comunicativo. No debemos dejar pasar el hecho de que hoy las redes sociodigitales se presentan como laboratorios sociales en los que cotidianamente los usuarios opinan sobre varios temas.

## INTERNET Y WEB 2.0: NOCIONES GENERALES

Hay un viejo cuento que se enmarca en la tradición budista y que narra cómo, en una ocasión, pasó un rey que montaba un elefante por un pueblo de ciegos. Estos, emocionados por la visita de tan inesperado animal, le solicitaron al monarca permiso para tocarlo...uno se acercó a una pata, otro a una oreja, uno más le cogió la trompa y, por último, uno se montó en el animal. Tras haber partido el rey y su animal, los ciegos hablaban de lo que era un elefante: el primero lo describió como un tronco rugoso; el segundo lo contradijo diciendo que era más parecido a un tapiz grueso que se contoneaba cuando se le tocaba; el tercero mencionó que, en realidad, el elefante era una especie de serpiente gigante y, por último el que se había trepado al animal les dijo, contundente, que estaban todos herrados, pues el elefante era una pequeña montaña que se movía. (Gómez, 2007, pp. 11-12)

Hoy en día para muchos investigadores entender ¿cómo funciona Internet? y ¿cuál es su potencial para la investigación? es similar a lo que sucede con la metáfora del pueblo de ciegos. En el contexto actual, Internet se presenta como un espacio tecnosocial al que acuden cotidianamente millones de personas en el mundo a realizar actividades vinculadas a la búsqueda de información, al uso de las redes sociales, al trabajo, al ocio y entretenimiento, a la educación, etc. Situación que convierte a esta red en un laboratorio social vivo atractivo para la investigación ya que “los usuarios actuales de la web proponen servicios, intercambian información, hacen comentarios, se implican, participan” (Piotet, 2009, p. 14). La evolución de los sistemas o servicios que ofrece Internet ha traído no solo la aparición de nuevas formas de comunicación, interacción y relación social para los usuarios de la red, también ha generado una nueva lógica para entender el potencial de los entornos digitales de Internet.

Estamos frente a una visión nueva de la red que deja atrás la idea de una web en donde los datos y la información tenían mayor relevancia, para dar paso a una web en la que las personas tienen mayor presencia frente a la propia información. Esto adquiere sentido si se piensa

que la “web, reducida a su definición más sencilla, de hecho, no es más que una herramienta para establecer relaciones entre personas, entre datos o documentos y, en la mayoría de los casos, es un híbrido entre estos tres elementos” (Piotet, 2009, p. 51).

Así, “el concepto web 2.0, debe su origen a una tormenta de ideas entre los equipos de *O’Reilly Media* y *Media Live International* a mediados de 2004, fortalecido por la primera *Web 2.0 Conference* en octubre de ese mismo año” (Cobo, 2007, p. 27). En un primer momento, podemos entender a la web 2.0 como todas las aplicaciones, utilidades y servicios de Internet, que se sustentan en una base de datos, que puede ser modificada constantemente por los usuarios, ya sea en el contenido (modificando la información y los datos presentados) o bien en la forma de presentarlos. La web 2.0 hace referencia a entornos, aplicaciones, tecnología y usos, ahí radica su complejidad. La web 2.0 se sustenta en siete principios fundamentales:

*la world wide web* como plataforma de trabajo, el fortalecimiento de la inteligencia colectiva, la gestión de bases de datos como competencia básica, el fin del ciclo de las actualizaciones de versiones del *software*, los modelos de programación ligeros en busca de la simplicidad, el *software* no limitado a un solo dispositivo y las experiencias enriquecedoras de los usuarios. (Cobo, 2007, p. 15)

Estos principios permiten que la web 2.0 sea visto como un desarrollo tecnosocial, a través del cual los usuarios de la red pueden realizar actividades en beneficio propio o común con otros usuarios, haciendo uso de los nuevos sistemas para la comunicación. Es necesario mencionar que las aplicaciones web 2.0 están desarrolladas pensando en la dinámica social de los usuarios de la red. Y es a través de estas herramientas que los usuarios se comunican, relacionan, interactúan, colaboran, participan en el ciberespacio y como consecuencia se crean todos los días laboratorios sociales sumamente complejos y dinámicos.

#### ***BIG DATA: CONCEPTOS GENERALES PARA EL INVESTIGADOR***

“En el entorno digital contemporáneo de respuesta inmediata, global y altamente interconectado y mediado por algoritmos matemáticos que impregnan todos los rincones de nuestra existencia, se generan grandes cantidades de datos que requieren ser analizados” (López-Cantos, 2015, p. 879), estos datos provienen de diversas y distintas fuentes como redes sociodigitales, páginas web, sensores, sistemas de *streaming*, transacciones bancarias, mercados financieros, GPS, etc. *Big data* es un término cercano a la Ciencia de datos, dis-

ciplina que combina matemáticas, programación, ciencias de la computación, inteligencia artificial, etc. El término *big data* se refiere fundamentalmente a “volúmenes masivos y complejos de información tanto estructurada como no estructurada que es recogida durante cierto periodo de tiempo y que requiere de métodos computacionales para extraer conocimiento” (Arcila, Barbosa, y Cabezuelo, 2016, p. 624).

Otros autores establecen que *big data* es un *data ecosystem*, que tiene como principal característica la alta conectividad entre los datos, sin importar su procedencia o estructura, y que permiten reconocer patrones de diversa índole (Suárez y Guerrero, 2016) y que el término se refiere a los volúmenes de información que no pueden ser procesados y analizados mediante procesos tradicionales (Camargo y Aguilar, 2015). Según Dana Boyd (2012), la era del *big data* está en marcha, ya que científicos de la computación, físicos, economistas, matemáticos, politólogos, entre otros profesionales, claman por el acceso a cantidades de información producidas por personas, cosas y sus interacciones y se convierte en un sistema de conocimientos que está cambiando los objetos de conocimiento y permite a las Ciencias Sociales la implementación de innovadores métodos cuantitativos de investigación (Boyd y Crawford, 2012).

Es importante señalar que el término *big data* se relaciona en la actualidad con cuatro dimensiones que son fundamentales para su comprensión:

- Volumen: cantidad de datos. Cantidades masivas de datos que las organizaciones intentan aprovechar para mejorar la toma de decisiones.
- Variedad: diferentes tipos y fuentes de datos. Tiene que ver con gestionar la complejidad de múltiples tipos de datos, incluidos los estructurados, semiestructurados y no estructurados (texto, datos web, tuits, datos de sensores, audio, video, clics, archivos, etcétera).
- Velocidad: datos en movimiento. La velocidad a la que se crean, procesan y analizan los datos continúa aumentando. Los sistemas tradicionales de almacenamiento no son eficientes.
- Veracidad: incertidumbre de los datos. Se refiere al nivel de fiabilidad asociada a ciertos tipos de datos (Schroedeck, Shockley, Smart, Romero y Turano, 2012).

## *MACHINE LEARNING*

*Machine learning* es un concepto que se deriva de la minería de datos, que hace referencia al diseño de programas o algoritmos que pueden aprender reglas a partir de datos, adaptarse

a cambios y paulatinamente mejorar el rendimiento con la experiencia. Es un campo multidisciplinar donde convergen la estadística y la computación. Es un proceso automático que extrae patrones de grandes cantidades de datos (Kelleher, Mac Namee y D'arcy, 2015). El *machine learning* se puede dividir en:

- Aprendizaje supervisado o predictivo en donde la máquina aprende no solo de los propios datos finales (*inputs*), sino que es posible darle modelos o datos adicionales ya categorizados (*outputs*) para que el aprendizaje sea mucho más fiable.
- Aprendizaje no supervisado o descriptivo en el que solo se dan los *inputs* para que encuentren patrones interesantes a partir de los datos (Arcila, Ortega, Jiménez y Trullenque, 2017, pp. 628-629).

Para que el *machine learning* funcione es necesario elaborar relaciones lógicas de variables y conceptos, debemos recordar que la base de la exploración de datos es la estadística.

#### ANÁLISIS DE SENTIMIENTOS

El análisis de sentimientos es un campo activo dentro de la investigación en el que se busca extraer sentimientos, opiniones, ideas y comportamientos de los usuarios a partir de un texto, usando el procesamiento del lenguaje natural (NLP) como método (Öztürk y Ayvaz, p. 137). Para Hutto y Gilbert (2014), “el análisis de sentimiento es utilizado para estudiar un amplio rango de problemas que están vinculados con la interacción computadora-humano, fenómeno de interés para investigadores de diversas áreas como la sociología, el marketing y la ciencia política” (p. 216).

El *sentiment analysis* sirve para analizar la opinión, sentimientos, evaluaciones, actitudes y emociones sobre productos, servicios, organizaciones, individuos, eventos y tópicos (Preethi, Uma y Kumar, 2015), también es una de las principales técnicas de estudio de datos textuales a gran escala (*big data*) que se emplean en la investigación en el área de las Ciencias sociales y en la Comunicación política. Su objetivo es reconocer y evaluar el valor emocional existente detrás de los textos analizados a través de su estructura, clasificándolos en positivos, negativos o neutros (Arcila, Ortega, Jiménez y Trullenque, 2017; Dattu y Gore, 2015).

## ANÁLISIS DE SENTIMIENTOS EN TWITTER

En los últimos años el número de tecnologías y personas que expresan sus puntos de vista y opiniones se ha incrementado en la web (Madhappan, Vasudem y Varma, 2012), particularmente en las redes sociales sirven para informar, expresar opiniones y sentimientos diversos sobre diferentes aspectos de la sociedad, productos, servicios y demás. Por eso empresas y colectivos en general han mostrado su interés en las opiniones y sentimientos que sus clientes y potenciales usuarios tienen sobre sus actividades (Gómez, Jaimes, Hidalgo y Lujan, 2018). En la actualidad, la red social Twitter se ha convertido en el canal de comunicación predilecto de los usuarios de Internet, esto se debe no solo a la facilidad de operación sino también de distribución de contenidos que ofrece esta red. Aunado a lo anterior, Twitter es el servicio más popular de *microblogging* a través del cual los usuarios expresan opiniones sobre diversos temas (Vinodhin y Chandrasekaran, 2012).

## NUEVOS ECOSISTEMAS PARA LA INVESTIGACIÓN: REDES SOCIALES

Según Castells (2006), una *red social* es una estructura formada por individuos y/u organizaciones que se conectan para establecer diversos tipos de relaciones. Se puede establecer que la principal característica de las redes sociales es la posibilidad que ofrecen para crear comunidades, esto es posible porque estos sitios permiten que los datos y las aplicaciones *web* alojadas en su dominio se convierten en elementos independientes y compartidos por diversos usuarios. Independientemente de los tipos de redes existentes en Internet (ocio, relaciones sociales, profesionales, etc.), es común que los usuarios utilicen estas redes para establecer dinámicas de interacción y relación social, y vínculos de comunicación con otros usuarios. Si se observa el universo de las redes sociales digitales se podrá encontrar que también millones de usuarios utilizan cotidianamente Internet para fines diferentes como: socialización, creación de comunidades con intereses particulares, aplicación de procesos creativos, investigación, educación, solución de problemas, intercambio de información, etc. Todo esto es posible gracias al desarrollo de herramientas y aplicaciones que hoy están destinadas primordialmente a la colaboración, la participación, la interacción y la relación social.

Asimismo, las herramientas de intercambio conectadas en redes permiten crear vínculos, tejer relaciones entre datos, entre personas o entre personas y datos.

La dimensión relacional de la web se ha acelerado debido al fuerte aumento del número de usuarios y de herramientas a su disposición. Cuanto más web actores hay más relaciones se establecen entre estos, más rico es el sistema y mejor funciona. (Pisani, 2008, p. 21)

Se puede inferir que lo realmente atractivo para los internautas con respecto a las redes sociales radica, en un primer momento, en el potencial que tienen estas redes para las dinámicas relacionales, debido a que en ellas se cuenta con altos niveles de libertad de expresión y ofrecen aplicaciones que permiten poner en contacto a usuarios con otros usuarios, no importando si estos son cercanos o no; y segundo, que estas redes ofrecen interactividad y la posibilidad de generar inteligencia colectiva y capital social que puede aprovecharse en procesos de investigación.

Actualmente existe un interés científico por la explotación de grandes cantidades de datos textuales disponibles en Internet gracias al uso masivo de los denominados medios sociales y de otras fuentes textuales de información. Esto se debe fundamentalmente a que en la actualidad Internet y particularmente las redes sociales se presentan como espacios donde las personas generan, publican y comparten contenidos de diversa índole de manera cotidiana. Twitter es una de las redes sociales donde más se generan, concentran y comparten datos, de esta manera, una vez que se recogen grandes cantidades de datos textuales (estructurados, semiestructurados o sin estructura) mediante procedimientos que van desde la recolección manual hasta procesos automatizados; y construida una base de datos, es fundamental el uso de métodos computacionales para realizar los análisis de datos y obtener el conocimiento que está inmerso en ellos y es que a medida que evolucionan los métodos de análisis científico asociados a la minería de datos y se perfeccionan las técnicas de análisis de sentimiento enfocadas a las redes sociales y *microblogs*, el interés por su aplicación en las ciencias de la comunicación y en particular en las ciencias políticas viene incrementándose en la última década de manera exponencial (Arcila, Ortega, Jiménez y Trullenque, 2017).

#### *BIG DATA PARA LA INVESTIGACIÓN EN COMUNICACIÓN*

En la actualidad el investigador en Comunicación se enfrenta a un ecosistema mediático digital, en el que Internet se erige como un laboratorio social en el que se llevan a cabo diversos fenómenos de comunicación, y en el que todos los días se acumula información que es valiosa para investigaciones de diversa índole, siempre y cuando el investigador conozca y utilice los recursos tecnológicos necesarios para obtener, procesar y analizar

dicha acumulación de datos. Constantemente aparecen nuevas tecnologías que permiten mejorar la eficiencia tanto de organizaciones como de investigadores en la búsqueda de conocer los gustos y necesidades de los consumidores, realizar predicciones y generar patrones de comportamiento. En el Cuadro 1 se presentan algunos recursos para el manejo de datos.

CUADRO 1. HERRAMIENTAS *BIG DATA* PARA EL ANÁLISIS DE DATOS

HERRAMIENTA <i>BIG DATA</i>	DESCRIPCIÓN	URL
Apache Hadoop	Es un framework que permite procesar grandes volúmenes de datos usando modelos de programación simples.	<a href="https://hadoop.apache.org">https://hadoop.apache.org</a>
Elasticsearch	Permite el procesamiento de grandes cantidades de datos y ver la evolución de estos en tiempo real. Además proporciona gráficos.	<a href="https://www.elastic.co">https://www.elastic.co</a>
Apache Storm	Es una herramienta open source que puede ser usada en cualquier lenguaje de programación. Crea tipologías de los macrodatos.	<a href="https://storm.apache.org">https://storm.apache.org</a>
Mongo DB	Es una base de datos NoSQL utilizada para trabajar en grupos de datos que varían con frecuencia, o que son semiestructurados o inestructurados. Se emplea para almacenar datos de aplicaciones móviles.	<a href="https://www.mongodb.com">https://www.mongodb.com</a>
Apache Spark	Es una herramienta más rápida que Hadoop: analiza datos por lotes y transmisión en tiempo real, permite la creación de aplicaciones en diferentes lenguajes (Java, Python, R y Scala).	<a href="https://www.spark.apache.org">https://www.spark.apache.org</a>
Python	Ofrece la ventaja de que se requieren conocimientos mínimos de informática para utilizarlo. Permite a los usuarios la creación de librerías para analizar datos.	<a href="https://www.python.org">https://www.python.org</a>
Apache Cassandra	Es una base de datos NoSQL, es la mejor opción para escalabilidad.	<a href="https://cassandra.apache.org">https://cassandra.apache.org</a>

HERRAMIENTA <i>BIG DATA</i>	DESCRIPCIÓN	URL
Lenguaje R	Es un entorno y lenguaje de programación enfocado al análisis estadístico.	<a href="https://www.r-project.org">https://www.r-project.org</a>
Apache Drill	Es un framework open source que permite trabajar en análisis interactivos de grupos de datos a gran escala.	<a href="https://drill.apache.org">https://drill.apache.org</a>
Apache Oozie	Es un sistema en flujo que permite definir un amplio rango de trabajos escritos o programados en diferentes lenguajes.	<a href="https://oozie.apache.org">https://oozie.apache.org</a>

FUENTE: ELABORACIÓN PROPIA CON INFORMACIÓN DE LAS PÁGINAS DE INTERNET.

La comprensión y uso de algunas de estas herramientas son importantes porque el volumen y la velocidad a la que se genera la información en el contexto actual es de tal dimensión que el análisis no puede realizarse con las metodologías tradicionales. No debemos olvidar que el potencial de los grandes datos no radica en el volumen, sino en su contextualización, en su estudio como flujo de información, y no como datos individuales; así, un tuit por sí solo no aporta mucha información, pero un estudio masivo de tuits publicados durante un periodo de tiempo determinado le permitirá al investigador en comunicación analizar tendencias e identificar patrones en los usuarios de la red (González-Benito, 2016).

De acuerdo con Bustamante (2017), el análisis de datos actualmente es vital para organizaciones, empresas e instituciones, y también para usuarios, profesores, directivos, investigadores, etc., ya que estos sirven para, entre otras cosas, pronosticar diversos aspectos de consumo, marketing, política, salud, educación, investigación. En el Cuadro 2 se presentan algunas utilidades del uso del *big data*.

El Cuadro 2 permite visualizar de forma clara algunos de los usos reales que tiene la investigación en *big data*, que de una u otra manera pueden ser explotados por los comunicólogos interesados en la implementación de las técnicas *big data* en sus investigaciones.

CUADRO 2. ÁREAS DE APLICACIÓN DE LA INVESTIGACIÓN EN *BIG DATA*

ÁREA DE APLICACIÓN	DESCRIPCIÓN
Segmentación de clientes	Los datos se utilizan para comprender mejor a los clientes, sus comportamientos y preferencias. El objetivo principal es crear modelos predictivos.
Optimización de procesos	El <i>big data</i> se está utilizando cada vez más para optimizar los procesos de negocio en las empresas. Ejemplo de ello es la cadena de suministro y las rutas de reparto.
Rendimiento personal	Los datos que se generan en los dispositivos como los smartwatch se utilizan para encontrar elementos que benefician a las personas.
Salud pública	Los datos masivos se utilizan para descodificar cadenas enteras de ADN y eso permite encontrar nuevos tratamientos y conocer mejor las enfermedades, sus desencadenantes y sus patrones de propagación.
Rendimiento deportivo	En el deporte de alto rendimiento se utilizan técnicas de análisis de datos para determinar patrones y estilos de jugadores ganadores. Ejemplos: tenis, béisbol, automovilismo, etcétera.
Ciencia e investigación	La ciencia se beneficia por las posibilidades que ofrece el <i>big data</i> , en el análisis de datos provenientes de diversas fuentes y formatos.
Optimización de máquinas y dispositivos	El análisis <i>big data</i> ayuda a mejorar los dispositivos, para que sean más inteligentes y autónomos.
Seguridad	El <i>big data</i> se utiliza en la mejora de la seguridad.
Optimización de ciudades	Los grandes volúmenes de datos que provienen de diversas aplicaciones ayudan a que las ciudades mejoren en algunos aspectos: tráfico, clima, transporte, etcétera.
Mercados financieros	El <i>big data</i> es importante en la toma de decisiones en los mercados financieros a nivel global.
Entretenimiento	El análisis de datos masivos ayuda en el éxito de diversos productos de entretenimiento. Ejemplo: Netflix.
Procesos electorales	En los procesos electorales el análisis proveniente sobre todo de las redes sociales permite determinar los sentimientos positivos o negativos con respecto a un candidato o partido político.
Predicción	<i>Big data</i> ofrece la posibilidad de predecir acciones y comportamientos partir de datos históricos que muestran las tendencias seguidas en cierto periodo.

FUENTE: ELABORACIÓN PROPIA CON INFORMACIÓN TOMADA DE LOS SITIOS [www.deustoinformacion.com](http://www.deustoinformacion.com), [www.baoss.es](http://www.baoss.es), [www.digitalbizmagazine.com](http://www.digitalbizmagazine.com) y [www.merca20.com](http://www.merca20.com)

## CONCLUSIÓN

El investigador en comunicación se enfrenta hoy a un contexto digital en el que Internet, a través de sus entornos digitales y la web 2.0, se erige hoy como un nuevo espacio social, en el que se generan millones de datos de diversa índole, fuente y formatos que están a disposición de los usuarios de la red para ser utilizados en diversas actividades, entre ellas la investigación, siempre que se cuenten con los conocimientos mínimos indispensables de herramientas que permitan el tratamiento y análisis de dichos datos. Así, el *big data* es una herramienta necesaria ya que le permite al investigador en comunicación, identificar patrones y obtener consecuencias significativas en el estudio de fenómenos sociales, que de otra manera no se podrían comprender en sus dimensiones correctas.

## REFERENCIAS

- Arcila, C., Ortega, F., Jiménez, J., y Trullenque, S. (2017). Análisis supervisado de sentimientos políticos en español: Clasificación en tiempo real de tweets basado en aprendizaje automático. *El Profesional de la Información*, 26(5). <https://doi.org/10.3145/epi.2017.sep.18>
- Arcila, C., Barbosa, E., y Cabezuelo, F. (2016). Técnicas Big Data: Análisis de textos a gran escala para la investigación científica y periodística. *El Profesional de la Información*, 25(4).
- Baviera, T. (2017). Técnicas para el análisis de sentimiento en Twitter: Aprendizaje automático supervisado y sentistrength. *Digitos. Revista de Comunicación Digital*, 33-50. <http://dx.doi.org/10.7203/rd.v1i3.74>
- Blakiwal, A., Alora, P., Madhapan, S., y Varma, V. (2012). Mining sentiments from Tweets. Proceedings of the 3er Workshop on computational Approaches to subjectivity and sentiment analysis, pp. 11-18. <https://www.aclweb.org/anthology/W12-3704.pdf>
- Boyd, D., y Crawford, K. (2012). Critical questions for Big data. *Information, Communication and Society*, 15(5), 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- Bustamante, A., Nicoletta, B., Guillen, A., y Thais, S. (2017). Un acercamiento mal Big Data y su utilización en Comunicación. *Mediaciones Sociales*. Ediciones Complutense. <https://doi.org/10.5209/MESO.58112>
- Camargo, J., Camargo, F., y Joynes, L. (2015). Conociendo Big Data. *Revista de la Facultad de Ingeniería*, 24(38), 63-77.
- Campos, E. (2017). Twitter y la Comunicación Política. *El Profesional de la Información*, 26(5), 785-793. <https://doi.org/10.3145/epi.2017.sep.01>
- Castells, M. (2010). *Comunicación y Poder*. Alianza Editorial.
- Cobo, C., y Kuklinski, H. (2007). *Planeta Web 2.0. Inteligencia colectiva o Medios fast food*. FLACSO.

- Dattu, B., y Gore, D. (2015). A survey on sentiment analysis on Twitter Data using different techniques. *International Journal of Computer Science and Information Technologies*, 6, 5358-5362.
- Gómez, E. (2007). *Las metáforas de Internet*. UOC.
- Gómez, E., Jaimes, R., Hidalgo, O. y Lujan, S. (2018). Influencia de redes sociales en el análisis de sentimiento aplicado a la situación política de Ecuador. *Enfoque UTE*, 9(1), 67-78. <https://doi.org/10.29019/enfoqueute.v9.n1.235>
- González, G. (2016). *Internet, comunicación y Sociedad red. Algoritmos para un periodismo multiconectado* [Tesis doctoral]. Universidad Carlos III de Madrid.
- Hutto, C., y Gilbert, E. (2014). VADER: Anparsimonius Ruled-based Model for Sentiment Analysis of Social Media Text. *Association for the Advancement of Artificial Intelligence*, pp. 216-225. <https://ojs.aaai.org/index.php/IWCSM/article/view/14550>
- Kelleher, J., Mac Namee, B., y D'arcy, O. (2015). *Machine learning for predictive data analytics*. MIT.
- López-Cantos, F. (2015). La investigación en comunicación con metodología *big data*. *Revista Latina de Comunicación Social*, 70, 878-890.
- Öztürk, N., y Serkan, A. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, 35, 136-147. <https://doi.org/10.1016/j.tele.2017.10.006>
- Piotet, D. y Pisani, F. (2009). *La alquimia de las multitudes. Cómo la web está cambiando el mundo*. Paidós.
- Preethi, P., Uma, V., y Kumar, A. (2015). Temporal sentiment analysis and causal rules extraction from Tweets for event prediction. *Procedia computer science*, 48, 84-89. <https://doi.org/10.1016/j.procs.2015.04.154>
- Schroek, M., Shockley, R., Smart, J., Romero, N., y Turano, P. (2012). *Analytics: El uso de big data en el mundo real*. Oxford University.
- Sitaram, A., y Huberman, B. (2010). Predicting de future with social media. *International Conference on Web Intelligent Agent Technology*. <https://doi.org/10.1109/WI-IAT.2010.63>
- Suárez, S., y Guerrero, F. (2016). La conversación sobre Big Data en Twitter. *Comunicació: Revista de recerca I d'Anàlisi, Societat nCatalana de Comunicació*, 33, 113-131. <https://doi.org/10.2436/20.3008.01.151>
- Wang, H., Can, D., Kazemzadeh, A., Bak, F., y Narayanan, S. (2012). A system for real time Twitter sentiment analysis of 2012 U.S. Presidential Election Cycle. *50<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, p. 115-120. <https://www.aclweb.org/anthology/P12-3020.pdf>
- Vinodhini, G., y Chandrasekaran, R. (2012). Sentiment Analysis and opinion mining: A survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2, 282-292.

Esta obra está bajo Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional.

