

XIII Jornadas de Sociología. Facultad de Ciencias Sociales, Universidad de Buenos Aires, Buenos Aires, 2019.

La Infoesfera y el Proyecto GDELT.

David Taraborrelli.

Cita:

David Taraborrelli (2019). *La Infoesfera y el Proyecto GDELT. XIII Jornadas de Sociología. Facultad de Ciencias Sociales, Universidad de Buenos Aires, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/000-023/42>

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

Título: La Infoesfera y el Proyecto GDELT.

Autor: David Taraborrelli

Eje Temático 1: Filosofía, Teoría, Epistemología, Metodología

Mesa 9: Big data y ciencias sociales

Institución de pertenencia: FSOC/UBA

Email: boedo28@hotmail.com

Resumen:

La expansión vertiginosa de los medios masivos de comunicación y las tecnologías de información en las últimas décadas ha dado forma a nuevas configuraciones culturales. Una de las expresiones más actuales de estos procesos es el BigData. Este fenómeno ha transformando muchos de los procesos económicos, sociales y políticos con los que ha interactuado. En el ámbito de las ciencias sociales, emergieron nuevos objetos de estudio como las Redes Sociales o la Web, al tiempo que, se incorporaron nuevas técnicas de análisis de datos. Considerando este contexto, el presente trabajo se propone relacionar el concepto de Infoesfera, considerada por Franco Berardi como el espacio circular en el que transitan las “señales portadoras de intención cultural”, y el desarrollo del subcampo de las Ciencias Sociales conocido como Culturomics 2.0. Con el fin de visualizar algunas de las tensiones conceptuales se realizará un estudio de caso sobre el Proyecto GDELT. Que tiene por función monitorear los mensajes que circulan en los Medios Masivos de Comunicación Digitales y en las Redes Sociales; con el fin de crear una plataforma abierta en tiempo real que permita investigar los eventos, sueños, miedos y conflictos que se producen alrededor del mundo actual.

Palabras Claves: Culturomics, Big Data, Infoesfera, GDELT, Medios de comunicación

Introducción.

En la década del sesenta surge un neologismo que cohesiona los términos información y esfera. Su origen está asociado a las transformaciones impulsadas por la Cibernética, la Teoría de la Información, la Ecología y la pregnancia de la figura de la Aldea Global. Su utilización refería principalmente al conjunto de emisores y mensajes que de manera constante y creciente poblaban el éter, distribuyendo la cultura en las sociedades contemporáneas.

La expansión vertiginosa de las tecnologías de información que se produce en los años 90' modifica fuertemente a los medios masivos de comunicación y los conduce por el tamiz de la convergencia digital, dando forma a nuevas configuraciones culturales. Una de las expresiones más actuales de estos procesos es el BigData. Este fenómeno ha transformado muchos de los procesos económicos, sociales, culturales y políticos con los que ha interactuado. Por ejemplo, en el ámbito de las ciencias sociales, emergieron nuevos objetos de estudio como las Redes Sociales o la Web, al tiempo que, se incorporaron nuevas técnicas de análisis de datos que permiten construir novedosos objetos de estudio y preguntarse por problemas cuya escala desborda las posibilidades de las metodologías tradicionales.

Considerando lo expuesto, este trabajo¹ se propone explorar, relevando fuentes secundarias, el desarrollo del concepto de infoesfera a partir de la serie de estudios denominados culturomics. Estos últimos comprenden el uso de grandes bases de datos textuales y la aplicación de técnicas computacionales para identificar procesos de cambios culturales. Con el fin de visualizar algunas de las tensiones conceptuales se focalizará en el caso del Proyecto GDELT² y su objetivo de mapear una gran proporción de la infoesfera.

Recorrido conceptual de la Infoesfera.

La trayectoria del concepto de Infoesfera se desarrolla en los últimos 50 años. Sus orígenes están asociados a preguntas por la comunicación y las transformaciones que se comenzaban a manifestar en los ámbitos culturales. En la década del setenta la idea de información en conjunción con la de esfera encontró asidero en las propuestas de: a) la Cibernética (Wiener, 1958), el hombre y las máquinas comparten la finalidad de ser “islas, locales y temporales, de entropía decreciente, en un medio en el cual la entropía tiende a aumentar” (p. 35); b) la Teoría de la Información (Shannon, 1948), la información puede determinarse cuantitativamente; c) en la figura de Aldea Global propuesta exitosamente por Marshall McLuhan en sus libros “The Gutenberg Galaxy: The Making

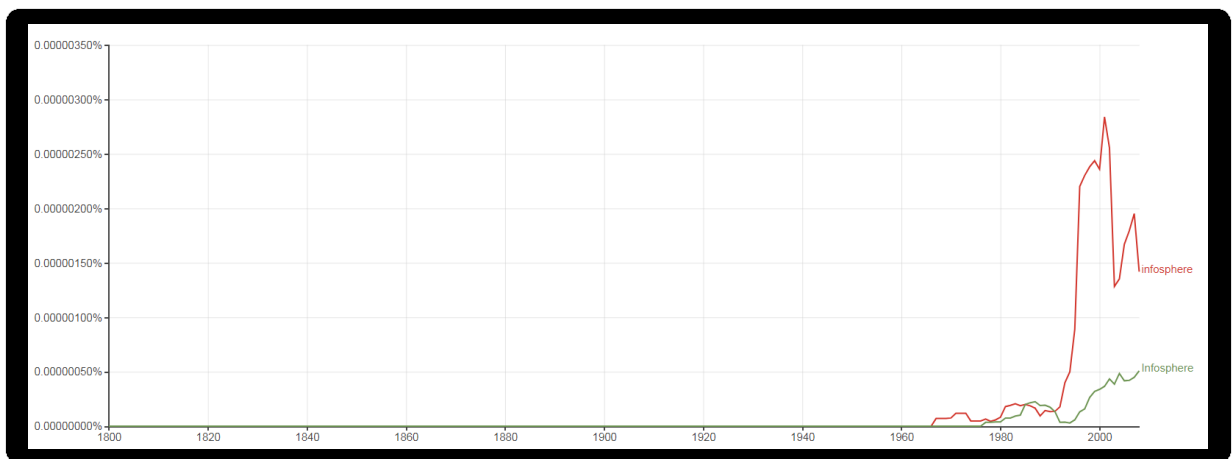
¹ Este trabajo forma parte de los avances de mi investigación de maestría en Cultura y Comunicación (UBA), en la que estudio la relación entre medios de comunicación y agendas políticas en la era de la convergencia digital.

² <https://www.gdelproject.org/> [último acceso: 9/07/2019]

of *Typographic Man*” (1962) y *“Understanding Media”* (1964); y d) Como afirma Scolari (2010), el establecimiento de una mirada ambiental de los medios de comunicación estuvo fuertemente anclada en las ideas ecologistas que surgen en los años 60’.

Incorporar a la comunicación máquinas capaces de manipular y procesar símbolos abre la puerta a considerar los medios de comunicación como parte de la naturaleza humana, sin ellos los humanos no pueden relacionarse con el mundo (Trujillo y Llamas, 2017). El entramado entre esfera e información puede ser pensado como un intento de unificación, una búsqueda de producir totalidad; idea castigada por el horror post 2º Guerra Mundial, por la fractura en la política internacional expresada en la Guerra Fría y sostenida por el discurso tecnotriunfalista (Winner, 2016) que alentaba “la creencia de que el bienestar humano depende en última instancia de la continuación del avance tecnológico” (p. 130). En este contexto la infoesfera hace su presentación e inicia su devenir

Gráfico N°1: Usos del término Infoesfera (Infosphere y infosphere en inglés) según período de tiempo.



Fuente: Google N-Gram.

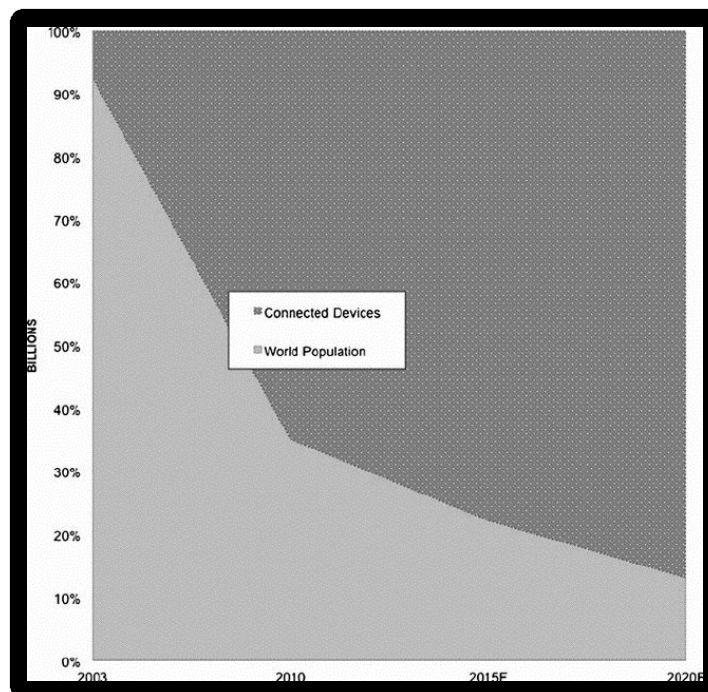
En el Gráfico N°1 puede observarse que los usos del término (en inglés) comienzan en la década del sesenta. Un conocido ejemplo de su uso data del año 1979, cuando Alvin Toffler (1980) utiliza el significante de este modo: “we are altering our info-sphere fundamentally. We are not merely de-massifying the Second Wave media, we are adding whole new strata of communication to the social system” (p. 172). Es notoria la veloz evolución a partir de los años 90s. Esta aceleración tiene por causa principal la vertiginosa expansión de Internet, que revitalizó las preocupaciones y abrió nuevos interrogantes acerca del hombre y su entorno técnico.

Con la finalidad de identificar las principales características del concepto de Infoesfera. Se recupera aquí la recomendación de Deleuze y Guattari (1993) quienes definen a la filosofía como el acto-arte de crear conceptos. Además, se elaborará una definición de Infoesfera siguiendo la premisa del

filósofo italiano Franco Berardi, preocupado por los fenómenos informacionales y las transformaciones que se experimentan en la cultura contemporánea.

Es necesario destacar que el concepto de infoesfera es considerado también desde la filosofía de la información. El trabajo de Floridi (2002), uno de sus principales exponentes, presenta a la Infoesfera como: “the whole system of services and documents, encoded in any semiotic and physical media, whose contents include any sort of data, information and knowledge” (Floridi, 2002, p. 8). El autor afirma que se produce un giro informacional que tiene por efecto el fin de la distinción entre lo real y lo virtual. Lo humano se desdibuja y lo artificial pasa a tener un rol de agencia en los procesos sociales: “All entities are informational entities, some informational entities are agents, some agents are artificial, some artificial agents are moral, and moral artificial agents are accountable but not necessarily responsible” (Floridi, 2013, p. 110).

Gráfico N°2: cantidad de dispositivos conectados a Internet según la población mundial.



Fuente: Floridi (2014) en base a Evans (2011).

Sin embargo, en un ejercicio de lectura del trabajo de Floridi Almendros y Echeverría (2019) sostienen que: “no presta atención a las dimensiones sociales, políticas ni axiológicas de la infoesfera” (p. 152). Siendo compartida esta interpretación y buscando hacer foco en dimensiones de lo político, se optó por trabajar la noción de infósfera desde el análisis de la sociedad post-alfabética de Berardi (2007). Son varios los puntos en común entre Floridi y Berardi, uno de los más destacados es el considerar la difusión exponencial de las tecnologías digitales como una de las causas del proceso de ampliación de la infoesfera. Un ejemplo de este proceso puede verse en el

Gráfico N°2, en el cual se ilustra cómo en 15 años se invirtió la relación entre la población mundial y la cantidad de dispositivos conectados a Internet.

La relación información-esfera en la propuesta teórica de Berardi se denomina: Infósfera. La diferencia en la sintaxis con respecto a Floridi, puede ser un resultado de la traducción o la intención creadora del autor. Pero sin lugar a dudas, la solución entre los dos significantes responde a la idea de entorno, información y transformación radical.

En Berardi (2007) sostiene que la Infósfera es un conjunto de sistemas en red que penetran en la mente colectiva. Es un espacio topológico creado a partir de la circulación de señales portadoras de intención cultural emitidas por un creciente número de emisores (p. 167). Un flujo de estímulos nerviosos (palabras, imágenes, audios y videos) que envuelven a los sujetos y que transforman el modo en que se producen las subjetividades; siguiendo a Deleuze y Guattari (1985), el inconsciente no es un teatro (versión psicoanalítica), es más bien, una fábrica (Berardi, 2007, p. 164).

Preocupado por los cambios que la información pueda producir en las máquinas que operan en los talleres de las mentes contemporáneas. La definición de Infósfera, en esta aproximación, desplaza el foco de la indiferenciación entre humanos y entorno informativo, y lo concentra en la relación de ajuste entre la infósfera y los sujetos. Preguntándose por la “transmisión de la herencia cultural y política del siglo XX a las generaciones que se forman en un ambiente tecnológico y comunicativo completamente transformado” (Berardi, 2007, p. 24). También se destacan los cuestionamientos al rol de los grandes medios de comunicación y de los movimientos de resistencia creativa y de información independiente.

La infoesfera ha crecido en densidad y variedad. Sin embargo, Berardi hace foco en la velocidad como la dimensión más problemática del fenómeno: “La catástrofe temporal se produce sobre todo en el plano cognitivo. Es consecuencia de un colapso en la relación entre la velocidad de la infósfera y los tiempos de elaboración racional y emotiva” (p. 185). Los sujetos estamos expuestos a un cibertiempos y a una dinámica multitasking. El hombre entre máquinas que describió Simondon (2007) se encuentra con un problema de acoplamiento. La velocidad y densidad de la Infósfera supera con creces la capacidad de los receptores humanos y esa saturación tiene fuertes consecuencias sociales, ya que: “Las tecnologías de la mente no son propiedad común de todos los seres humanos, sino propiedad privada de unos pocos grupos económicos mundiales extremadamente poderosos. Estos grupos se han vuelto capaces de canalizar la atención, el comportamiento, las expectativas, las elecciones de consumo y las elecciones políticas” (Berardi, 2007, p. 187).

Para Berardi el centro del problema está situado en la categoría de sujeto y la preocupación por distinguir el formato de las mentes post-alfabéticas, propias de las generaciones que han aprendido más palabras de una máquina que de su madre (Berardi, 2007, p. 25). Es por ello que la infósfera tiene límites definidos: “La infósfera es la interfaz entre el sistema de los medios y la mente que recibe sus señales” (p. 175). Esta puede entonces ser definida como el universo de enunciados que han circulado y dejado huella mnémicas; ya sea en estructuras psíquicas, en libros, en grafitis, en bases de datos o en redes neuronales artificiales. En otras palabras, se puede definir como el entorno compuesto por la totalidad de entidades informacionales, agencias con características propias, procesos y relaciones mutuas.

Los grandes medios de comunicación tradicionales, los nuevos medios como las redes sociales y los movimientos de resistencia de información independiente adquieren un lugar privilegiado. Sobre todo para estudiar el fenómeno de la constitución de la subjetividad, su vínculo con lo político y la tensión con la experiencia de una Infósfera cada vez más variada, densa y, sobre todo, veloz.

Culturomics y Big Data.

Definida conceptualmente la infoesfera, ¿Cuál es la mejor estrategia para convertirla en un problema de estudio abordable empíricamente desde las ciencias sociales? Para Berardi, y también para Floridi, la infoesfera creció en volumen y, además, en densidad. Es decir, un mayor número de elementos informativos por unidad espacial. La cantidad de emisores participantes y la velocidad en que se producen y circulan las “señales portadoras de intención cultural” también se han incrementado.

Un fenómeno reciente asociado estructuralmente con la infoesfera es el Big Data. Este término es usualmente relacionado con un aluvión de datos (Bell et. al., 2009) que, asociado con algoritmos de machine learning, han transformado principalmente algunos ámbitos de la industria (Schwab, 2016; Srnicek, 2018) y las ciencias (Manovich, 2011; Anderson, 2008). El Big data “surge como una nueva fase del paradigma intensivo en información y comunicación que abarca no sólo su dimensión tecnológica, sino también una dimensión social, económica, política y cultural” (Malvicino y Yoguel, 2015).

Al igual que la infoesfera, las bases de datos catalogadas como Big Data se caracterizan por: un gran volumen y densidad; amplia variedad de formatos; y la necesidad de operar a gran velocidad para obtener respuestas en un tiempo cercano a la experiencia real (Bell et. al., 2009; Kitchin, 2013; Laney, 2001; Gartner, 2011). Administrar y manipular esta clase de datasets requiere un abordaje

multidisciplinario, en el que no puede escasear la capacidad de almacenamiento, el poder de procesamiento y técnicas de cómputo avanzadas como el machine learning.

En el marco de la investigación en ciencias sociales, el Big Data ha desdibujado algunos límites disciplinares, permitiendo la emergencia de propuestas como: la Sociología Digital (Lupton, 2014; Lupton, 2016; Orton-Johnson y Prior, 2013) y las Humanidades Digitales (Schreibman, 2004). En los últimos años este campo ha crecido y se ha institucionalizado, entre los casos más destacados se cuenta el Alliance of Digital Humanities Organizations³ y el centerNet⁴, este último es una red internacional de más de 100 centros universitarios de estudios en humanidades digitales.

Lo novedoso de estos estudios y lo variado de los formatos de sus fuentes ha conducido a que se establezcan diversos términos para encuadrar sus producciones. Ejemplo de ello son los *Cultural Analytics*⁵ (Hochman y Manovich, 2013; Manovich, 2011) principalmente ocupados del análisis automático de grandes bancos de imágenes; o los estudios de *Culturomics* (Michel Et. al., 2011; Moretti, 2003) que suelen considerar como fuente textos digitales o digitalizados, buscando identificar patrones culturales desplegados en el tiempo. En este trabajo se profundizará en esta última clase de estudios, en la búsqueda de identificar algunas estrategias apropiadas para operacionalizar la infoesfera.

La infoesfera, como ya se dijo, incluye materiales semióticos analógicos y digitales. Las imágenes y los textos son uno de sus principales elementos. Sin embargo, los registros textuales predominan por sobre las imágenes y los registros sonoros a lo largo de la historia. Mientras la imprenta multiplicaba industrialmente los textos escritos las imágenes todavía eran portadoras de un aura (Benjamin, 2019). Aunque desde el siglo XIX, las tecnologías de la reproductibilidad técnica, como la fotografía y el cinematógrafo, pusieron en jaque esa característica esencial.

El objetivo general de los estudios culturomics es identificar patrones culturales de forma longitudinal en los textos escritos y digitalizados. La propuesta es análoga a la Escuela de los Annales en Historia, como sostiene Moretti (2003): “qué ocurriría si también nosotros cambiáramos nuestro centro de atención de los textos excepcionales a «la gran masa de los hechos [literarios]»” (p. 60). Este autor propone el concepto de lectura distante en contraste con la tradicional lectura cercana propia de los estudios clásicos de análisis de contenido (Melo Flórez, 2017; Moretti, 2003).

En la búsqueda de tendencias literarias que expresen ciclos culturales y cambios en la sintaxis y la gramática, Michel, Aiden et al. (2011) escriben uno de los trabajos más conocidos del culturomics:

³ Sitio web institucional: <http://adho.org>

⁴ Sitio web institucional: <http://dhcenternet.org/>

⁵ Un conjunto de trabajos en esta área pueden verse en: <http://lab.culturalanalytics.info/>

“Quantitative Analysis of Culture Using Millions of Digitized Books”. En esa investigación crean y analizan, apoyados en los recursos de Google Labs, un corpus de 5.195.769 libros, según ellos, el 4% de todos los libros alguna vez publicados. Uno de los principales logros es la herramienta Google Books Ngram Viewer⁶, abierta al público y utilizada en este trabajo para ilustrar el recorrido en la literatura en inglés del término “infosphere”. Los usos de esta herramienta se han multiplicado en los ámbitos académicos, sobre todo en temas como los cambios idiomáticos (Ophir, 2016; Juola, 2013; De Luca, 2016) o los cambios en paradigmas científicos (Pettit, 2016; Montagne y Morgan, 2013; Kumar y Sahu, 2011).

Cabe destacar que se han realizado algunos cuestionamientos a la herramienta presentada en el párrafo anterior por no presentar los metadatos del corpus (Koplenig, 2017); y por algunos sesgos en la selección de libros de las últimas décadas del siglo XX (Pechenick et. al., 2015; Zhang, 2015). Esta última crítica se debe a que los libros fueron provistos en su mayoría por bibliotecas universitarias, por lo tanto presentan mayor contenido científico. Aunque esto último buscó ser solucionado agregando la posibilidad de seleccionar el corpus “English Fiction”.

Sin embargo, muchas investigaciones de culturomics utilizan otros corpus. Los trabajos sobre medios de comunicación (Phani, et. al., 2012; Suchanek y Preda, 2014) que consideran la gran cantidad de noticias publicadas por los medios de comunicación tradicionales. Además, están los estudios de Stanford Literary Lab⁷, en el que se investigan por ejemplo, las emociones volcadas en la literatura británica sobre los barrios londinenses y como estas se modifican al transformarse la ciudad (Heuser, Moretti y Steiner, 2016; Moretti, 1997; Moretti, 2004a; Moretti, 2004b; Richterich, 2011).

Hasta el aquí se ha presentado una de las formas de cristalizar el etéreo concepto de infoesfera. Los estudios de culturomics, apoyados en los grandes corpus textuales, han habilitado un modo de indagar en algunos cambios históricos en los usos del lenguaje escrito. Sin embargo la mayoría de las investigaciones citadas contemplan materiales digitalizados, es decir, los nacidos analógicos y posteriormente modulados. En el siguiente apartado se hará foco en un proyecto que se inscribe en los culturomics pero que se sustenta en grandes datos nacidos digitales como las noticias de periódicos web.

El Culturomics 2.0: un paso más hacia la Infoesfera.

⁶ Ver: <https://books.google.com/ngrams>

⁷ Ver: <https://litlab.stanford.edu/>

En el apartado anterior se relevó la bibliografía disponible de culturomics a fin de identificar modos de construir materialmente a la infoesfera. Estos estudios presentan en general recortes temporales, temáticos y suelen enfocarse en alguno de los idiomas disponibles en los grandes corpus. La elaboración de estos últimos depende de la disponibilidad de los materiales textuales (libros, revistas, artículos) y de tecnologías apropiadas para digitalizar los textos.

La propuesta de la Web 2.0⁸ resultó en la expansión de los productores de contenidos. Se popularizan los comentarios en la web, los blogs y los sitios que permiten la colaboración entre usuarios. Jugando con esta figura, Leetaru (2011) propone los estudios de culturomics 2.0. El autor elabora el proyecto al que denomina “Global Database of Events, Language, and Tone” (GDEL). Este es presentado como una plataforma abierta que comparte la base de datos más grande sobre la sociedad humana.

El proyecto monitorea los medios de comunicación impresos, los *broadcast* y los formatos webs con el fin de construir una base de datos que permita consultar los sucesos de todos los países del mundo, en más de 100 idiomas y con una frecuencia de actualización de 15 minutos. Además cubre un periodo de 30 años ya que realiza una reconstrucción de las noticias desde el año 1979 considerando los bancos de datos de los principales periódicos internacionales. Sin embargo el autor afirma que en los últimos 15 años la proporción de noticias web ha superado notoriamente a la de noticias impresas (Leetaru, 2011, p. 2).

Teniendo como referencia proyectos gubernamentales como el Foreign Broadcast Information Service⁹ y Summary of World Broadcasts, GDEL se propuso utilizar técnicas de análisis computacional de contenido como análisis de sentimientos, geolocalización, análisis automático de tópicos y análisis de redes para procesar una amplia variedad de fuentes: miles de medios de comunicación digitales, las closed caption de más de 100 estaciones de televisión, colecciones de noticias de grandes medios, libros digitalizados y 70 años de literatura académica.

Del proceso que denominan “Computing on The Entire World: Events & Networks” se desprenden los dos resultados más importantes de GDEL, a saber: i) una base de datos de eventos globales geo-referenciados, catalogadas con más de trecientas categorías de tópicos; y ii) además se presenta una base de datos denominada “Global Knowledge Graph” (GKG) que expone un grafo de personas, organizaciones, empresas y lugares relacionadas con miles de temas y emociones presentes en las noticias analizadas. Según el proyecto: “The resulting network diagram constructs a

⁸ Ver más detalles en:

<https://web.archive.org/web/20071025235914/http://sociedaddeinformacion.telefonica.es/jsp/articulos/detalle.jsp?elem=2146>

⁹ Ver más detalles en: https://en.wikipedia.org/wiki/Foreign_Broadcast_Information_Service

graph over the entire world, encoding not only what's happening, but what its context is, who's involved, and how the world is feeling about it, updated every single day.” (Proyecto GDELT, 2019).

El tamaño de un año de resultados en el GKG se estima en 2.5 Terabytes. Este gran volumen impone ciertas dificultades para trabajar con los datos. Por lo tanto el proyecto habilita tres modalidades de acceso: a) un servicio de informes cuyos parámetros se cargan en un formulario web y se reciben vía email los resultados, cabe mencionar que esta modalidad está restringida a la versión 1.0 del proyecto; b) usando el servicio BigQuery de la plataforma Google Cloud, que permite la ejecución de consultas en lenguaje SQL (lenguaje de consulta estructurado); y c) la descarga directa de los archivos para ser procesados de modo local.

El carácter global del proyecto, el gran volumen de datos procesados y de resultados expuestos, la variedad de fuentes y de idiomas, sumados a la velocidad de actualización permite afirmar que este proyecto no solamente encuadra como Big Data sino que representa uno de los más ambiciosos intentos de cartografiar la infoesfera.

Son muchas las posibilidades que abre esta clase de proyectos, sobre todo para la comunidad académica interesada en los posibles efectos de los medios de comunicación. Sin embargo, es necesario mencionar algunos ítems acerca de su uso y posibilidades. En primera instancia, se destaca que las fuentes son traducidas automáticamente al inglés previamente a ser procesadas. Esto puede ser problemático al momento de interpretar resultados, por ejemplo el análisis de sentimientos. En segundo lugar, y relacionado con el primer ítem, no se presentan las fuentes originales ni el texto completo de la noticia. Esto dificulta la comparación con otros resultados o la puesta en práctica de otras técnicas de procesamiento de lenguaje natural apropiadas para el idioma y el tipo de fuente primaria.

La dimensión del problema que se presenta al querer materializar la infoesfera hace visible la necesidad de apoyo de grandes centros de procesamiento y almacenamiento de datos. Los dos proyectos más grandes de los estudios de culturomics requirieron el soporte de Google para presentar una interfaz lo suficientemente amigable para estudiar sus resultados. Un problema que se desprende de ello es que la herramienta BigQuery de Google Cloud no es gratuita y requiere conocimientos de lenguajes de consulta específicos. Otro aspecto que se desprende de la exploración es el apoyo institucional de la National Science Foundation. Esta agencia federal norteamericana brindó soporte económico a dos proyectos¹⁰¹¹ que colaboraron con el desarrollo de

¹⁰ Ficha disponible en: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1539302

GDELT. El monto del apoyo suma aproximadamente 480.000 dólares por año entre ambos proyectos.

En síntesis, el proyecto GDELT es un claro avance en los estudios del culturomics. Sus datos permiten modelar cambios en los modos de cobertura informativa sobre eventos, personas e instituciones con detalle y siguiendo la evolución a lo largo del tiempo. Esto último es un requisito para permitir la lectura distante y la identificación de ciclos en la oceánica infoesfera. Aunque las posibilidades que habilita son tan variadas como sus fuentes, su desarrollo e implementación expone la necesidad de un fuerte soporte institucional, tecnológico y económico.

Palabras finales.

El acelerado despliegue de la infoesfera en la últimas décadas ha impactado en la vida contemporánea. Como sostuvo Berardi, la velocidad y la densidad de sus flujos obturan la capacidad de los sujetos para precisar los fenómenos que los rodean, cada vez más interconectados y mundializados. Esto último se debe a las propiedades intrínsecas de la información, que según Floridi, tiene la capacidad de permear toda la existencia, constituyendo un nuevo fondo ontológico del que emergen un sin número de entidades informacionales.

El estudio de este fenómeno requiere el abordaje desde una perspectiva multidisciplinaria. En este trabajo, debido a la materialidad textual de sus fuentes se seleccionó el conjunto de investigaciones enmarcadas como culturomics. Destacándose el proyecto GDELT, que busca reconstruir en breves intervalos de tiempo un mapeo de la enunciación de los medios de comunicación a nivel internacional. Con el objetivo de hallar patrones espacio temporales de riesgos políticos, situaciones de catástrofes y eventos varios con influencia regional (Leetaru, 2011).

Berardi menciona dos aspectos relevantes sobre la infoesfera. El primero refiere a la propiedad privada de los medios de comunicación y su capacidad para darle forma a la infoesfera condicionando las elecciones de los agentes. Para problematizar dicha dimensión el culturomics y su versión 2.0 pueden resultar de suma utilidad, permitiendo distinguir el ritmo de los flujos informativos según las entidades referenciadas, el tono de los términos usados y su geolocalización. En segundo lugar, Berardi considera que la densidad y velocidad de circulación de señales con intención cultural produce un colapso cognitivo en las mentes de la sociedad contemporánea descrita como post-alfabética. Esta dimensión es difícil de abordar desde el culturomics debido a la necesidad de identificar el proceso de recepción de los mensajes. Sin embargo, a futuro, se

¹¹ Ficha disponible en: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1004414

pueden diseñar investigaciones que superpongan recursos como GDELT con insumos de grandes datos provenientes de redes sociales.

La complejidad para utilizar recursos como GDELT pone de manifiesto la dificultad que afrontan las ciencias sociales, sobre todo en latitudes lejanas a los centros hegemónicos de poder, para elaborar reflexiones empíricas sobre un proceso tan amplio y variado. Más aún si se consideran los posibles sesgos idiomáticos y las limitaciones tecnológicas. Sin embargo, puede resultar de interés el uso de estos datasets para realizar corroboraciones de recursos locales sobre medios de comunicación, imposición de agendas mediáticas y análisis de encuadre.

Bibliografía:

- Almendros, L. y Echeverría, J. (2019). Ingenierías, sociedades digitales e infoesfera. Una interpretación de la filosofía y la ética de la información de Luciano Floridi. *Revista Iberoamericana de Ciencia, Tecnología y Sociedad-CTS*, 14(41).
- Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16(7), 16-07.
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132.
- Bell, G., Hey, T., & Szalay, A. (2009). Beyond the data deluge. *Science*, 323(5919), 1297-1298.
- Benjamin, W. (2019). La obra de arte en la época de su reproductibilidad técnica. Ediciones Godot.
- Berardi, F. (2007). Generación post-alfa: patologías e imaginarios en el semiocapitalismo. Buenos Aires: Tinta Limón.
- Castells, M. (2004). La era de la información: economía, sociedad y cultura (Vol. 3). siglo XXI.
- Deleuze, G., Guattari, F. (1993). ¿Qué es la filosofía?. Barcelona: Anagrama.
- Deleuze, G., Guattari, F. (1985). El antiedipo: capitalismo y esquizofrenia. Barcelona: Paidós.
- De Luca, J. (2016). Incidencia de idiomas populares en la lengua española con Big Data: análisis masivo de datos mediante Amazon Elastic MapReduce y Google N-grams (Doctoral dissertation, Facultad de Informática, Universidad Nacional de La Plata).

- Derrida, J. (1989). Firma, acontecimiento, contexto. En *Márgenes de la filosofía*. Madrid: Cátedra.
- Diuk, C (2014). The Formation of love. Recuperado de: <https://www.facebook.com/notes/facebook-data-science/the-formation-of-love/10152064609253859/>
- Floridi, L. (2002). What is the Philosophy of Information?. *Metaphilosophy*, 33(1-2), 123-145.
- Floridi, L. (2002). *Philosophy and computing: An introduction*. Routledge.
- Floridi, L. (2011b): *The Philosophy of Information*, Nueva York, Oxford University Press.
- Floridi, L. (2013): *The Ethics of Information*, Oxford University Press.
- Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.
- Gartner (2011). “Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data”. Junio 2011
- Juola, P. (2013). Using the Google N-Gram corpus to measure cultural complexity. *Literary and linguistic computing*, 28(4), 668-675.
- Haugeland, J. (1988). *La inteligencia artificial*. México: Siglo XXI editores.
- Hochman, N., & Manovich, L. (2013). Zooming into an Instagram City: Reading the local through social media. *First Monday*, 18(7).
- Kitchin, R. (2013) Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3), 262-267.
- Koplenig, A. (2017). The impact of lacking metadata for the measurement of cultural and linguistic change using the Google Ngram data sets—Reconstructing the composition of the German corpus in times of WWII. *Digital Scholarship in the Humanities*, 32(1), 169-188.
- Kumar, N., & Sahu, M. (2011). The evolution of marketing history: a peek through google ngram viewer.
- Laclau, E. (1996). Sujeto de la política, política del sujeto. En *Emancipación y diferencia*. Buenos aires: Ariel.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META group research note, 6(70), 1.
- Latour, B, (2007). Beware, your imagination leaves digital traces. *Times Higher Literary Supplement*. Sciences-Po, Paris.
- Latour, B (2008). *Reensamblar lo social: una introducción a la teoría del actor-red*. Buenos Aires: Manantial.

- Leetaru, K. (2011). Culturomics 2.0: Forecasting large-scale human behavior using global news media tone in time and space. *First Monday*, 16(9).
- Lupton, D. (2014). *Digital sociology*. Routledge.
- Lupton, D. (2016). *The quantified self*. John Wiley & Sons.
- McLuhan, M. (1962). *The Gutenberg galaxy: The making of typographic man*. University of Toronto Press.
- McLuhan, M. (1964). *Understanding media: The extensions of man*. MIT press.
- Manovich, L. (2011). Trending: The promises and the challenges of big social data. *Debates in the digital humanities*, 2, 460-475.
- Manovich, L. (2011). What is visualisation?. *Visual Studies*, 26(1), 36-49.
- Malvicino, F., & Yoguel, G. (2015). Descubriendo big data en argentina. In *Simposio Argentino de GRANdes DATos (AGRANDA 2015)-JAIIO 44* (Rosario, 2015).
- Montagne, M., & Morgan, M. (2013). Drugs on the Internet, part IV: Google's Ngram viewer analytic tool applied to drug literature. *Substance use & misuse*, 48(5), 415-419.
- Moretti, F. (1997). *Atlas of the European Novel, 1800–1900*, Verso, New York.
- Moretti, F. (2003) Gráficos, mapas, árboles – 1, *NLR* 24.
- Moretti, F. (2004a) Gráficos, mapas, árboles – 2, *NLR* 26.
- Moretti, F. (2004b) Gráficos, mapas, árboles – 3, *NLR* 27.
- Melo Flórez, J. A. (2017). Lectura distante, fragmentada y colaborativa en el archivo infinito. *Relaciones. Estudios de historia y sociedad*, 38(149), 169-189.
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva P. Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, et al. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014): 176-182.
- Morán Reyes, A. A. (2013). La ética de la información y la infoesfera. *Escritos*, 21(46), 21-37.
- Ophir, S. (2016). Big data for the humanities using Google Ngrams: Discovering hidden patterns of conceptual trends. *First Monday*, 21(7).
- Orton-Johnson, K. y Prior, N. (Eds.). (2013). *Digital sociology: Critical perspectives*. Springer.
- Phani, S., Lahiri, S., & Biswas, A. (2012). Culturomics on a bengali newspaper corpus. In *2012 International Conference on Asian Language Processing* (pp. 237-240). IEEE.
- Pettit, M. (2016). Historical time in the age of big data: Cultural psychology, historical change, and the Google Books Ngram Viewer. *History of Psychology*, 19(2), 141.

- Proyecto GDELT (2019) The GDELT Project. Recuperado de: <https://www.gdelproject.org/#watching>
- Toffler, A. (1980). *The third wave* (Vol. 484). New York: Bantam books.
- Trujillo y Llamas (2017). La distopía de la tecnología, *Opción*, vol. 33, núm. 83.
- Richterich, A. (2011). Cartographies of digital fiction: Amateurs mapping a new literary realism. *The Cartographic Journal*, 48(4), 237-249.
- Schwab, K. (2016). *La cuarta revolución industrial*. Debate.
- Scolari, C. 2010. “Ecología de los Medios. Mapa de un nicho teórico”. *Quaderns del CAC*. Disponible en: http://www.cac.cat/pfw_files/cma/recerca/quaderns_cac/Q34_Scolari_ES.pdf
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3), 379-423.
- Simondon, G. (2007). *El modo de existencia de los objetos técnicos*. Prometeo Libros Editorial.
- Srnicek, N. (2018). *Capitalismo de plataformas*. Buenos Aires: Caja Negra.
- Schreibman S., Siemens R., Unsworth J. (2004). *A Companion to Digital Humanities*. Oxford: Blackwell.
- Suchanek, F. y Preda, N. (2014). Semantic culturomics. *Proceedings of the VLDB Endowment*, 7(12), 1215-1218.
- Viani Sallaberry, J. (1995). Sinópsis histórica de la Inteligencia Artificial. *Revista Actuarios*, 12, 66-74. Recuperado de <http://www.actuarios.org/espa/revista12/17-art08.pdf>
- Wiener, N. (1958). *Cibernética y sociedad*, Sudamericana. Buenos Aires.
- Winner, L. (2016). Decadencia y caída del tecnotriunfalismo. *Revista Redes*, vol. 22, nº 43, pp. 127-142.
- Zhang, S. (2015). The pitfalls of using Google ngram to study language. *Wired*. Disponible en: <https://www.wired.com/2015/10/pitfalls-of-studying-language-with-google-ngram/>
-