



Investigaciones sociotécnicas en el BIG DATA. Un reto para el análisis cualitativo del #8M y #9M en tiempos de pandemia

Abraham Mena Farrera, Angelica Evangelista García

Resumen: Se discute el reto que desde las academias se enfrentan en el análisis cualitativo, el desarrollo y la disposición de información ordenada y producida desde sistemas BIG DATA, que sea útil para el análisis ciudadano de los discursos, mensajes, diálogos y noticias geolocalizadas que en contextos de pandemia sirva para documentar las marchas feministas del 8 de marzo y el paro de mujeres del 9 de marzo de los años 2020, 2021 y 2022, años que marcaron un antes, durante y después del confinamiento por COVID-19. Este artículo muestra de manera preliminar, un novedoso sustento metodológico desde el cual las personas interesadas en abordajes comprensivos desde las ciencias sociales y las humanidades inicien su propio debate. La discusión, estudio y análisis de las interacciones que intercambiamos en la red, es un asunto que se ha pospuesto por más de una década, y se ha promovido que el trabajo de producir y analizar altos volúmenes de datos no le corresponde a los análisis cualitativos, y se sostiene que son técnicas exclusivas de la infotécnica en el cual se representen información textual y tendencias, temas y hashtags. Hoy sabemos que no es así. Que los estudios de comprensión del discurso, el texto y los multimedia también corresponde al campo de los métodos cualitativos para en donde tenemos la tarea de ser protagonistas de temas, iniciativas, reclamos, miedos, denuncias y luchas en torno a los derechos humanos y hacktivismo dentro y fuera de la red. Hacerlo nos lleva a profundizar en los procesos y en los análisis más novedosos surgidos en los últimos años en donde necesitamos dar un salto de escala en donde la gran aceleración del Tecnoceno-Anthropoceno debate en las redes la vigencia de una nueva "gubernamentalidad algorítmica" asociada a la disponibilidad de datos masivos o BIG DATA.

Palabras clave: Pandemia, Cualitativo, México, Big Data, #8M y #9M.

Sociotechnical research in BIG DATA. A challenge for the qualitative analysis of #8M and #9M in times of pandemic

Abstract: The challenge that academies face in qualitative analysis, development and provision of organized information produced from BIG DATA systems, which is useful for citizen analysis of speeches, messages, dialogues and news geolocated in contexts, is discussed. of pandemic serves to document the feminist marches of March 8 and the women's strike of March 9 of the years 2020, 2021 and 2022, years that marked a before, during and after confinement by COVID-19. This paper shows in a preliminary way, a novel methodological support from which people interested in comprehensive approaches from the social sciences and humanities start their own debate. The discussion, study and analysis of the interactions that we exchange on the network is an issue that has been postponed for more than a decade, and it has been promoted that the work of producing and analyzing high volumes of data does not correspond to qualitative analysis and it is argued that they are exclusive techniques of info-technique in which textual information and trends, themes and hashtags are represented. Today we know that is not so. That the studies of discourse, text and multimedia comprehension also correspond to the field of qualitative methods for which we have the task of being protagonists of issues, initiatives, claims, fears, complaints and struggles around human rights and hacktivism inside and outside the network. Doing so leads us to delve into the most innovative processes and analyzes that have emerged in recent years where we need to take a leap in scale where the great acceleration of the Technocene-Anthropocene debates in the networks the validity of a new "algorithmic governance" associated to the availability of massive data or BIG DATA.

Keywords: Pandemic, Qualitative, Mexico, Big Data, #8M and #9M.

1. Introducción

La ponencia tiene como objetivo presentar, discutir y avanzar en la consolidación de metodologías e investigaciones basadas en la producción de narrativas basadas en la tecnología Big Data que hoy pueden consultarse en el sitio <https://kibana.web.ecosur.mx>, en donde, por medio de un portal ciudadano de acceso público y gratuito para la población abierta y comunicad científica y académica, se pueda acceder a distintos análisis mediante búsquedas dirigida al contenido en tiempo real de la producción textual de discursos, mensajes, diálogos y noticias geolocalizados sobre la epidemia COVID-19 en sus fases y la infodemia que se provoca en una red social.

El reto de cosechar de datos producto de las interacciones sociotécnicas por internet, implica un trabajo multidisciplinario que permita almacenar un alto volumen de información en español acerca de las marchas y paro del #8M y #9M, antes, durante y después del paso de la pandemia de COVID-19, que permita a la academia y ciudadanía visualizar, conocer, estudiar y analizar el fenómeno socio digital que los usuarios expresan en la red, a partir de una plataforma de consulta, poderosa en su manejo y búsqueda de información. El análisis de la ciudadanía podrá ser filtrada por el origen de la red redes, fecha, países y estados, temas, hashtag y filtro de Bot.

La producción y análisis de la información cosechada basados en procesos de Big Data tiene el objetivo de ser una herramienta pública para conocer las tendencias, y el contenido de las publicaciones que sobre los fenómenos sociales desde el confinamiento y desconfinamiento producto del aislamiento social en pandemia, para conocer y almacenar una memoria de lo que se discuten en la red y considerarlos insumos confiables de la interacción en internet para ser usada en el periodismo, la academia, la política y la ciudadanía en general.

2. Explicar el Big Data para el uso de análisis de fenómenos sociotécnicos

El COVID-19 representó para la academia en México, en particular para el gobierno federal, la población, el sistema de salud y por supuesto los centros públicos de investigación (entre otros actores e instituciones) uno de los mayores retos de la historia moderna. La atención pronta de una epidemia mortal y desconocida hace que tengamos que proponer esquemas de investigación poco convencionales o a la fecha poco explorados por la ciencia sociales y humanidades.

Las academias se fueron articulando en procesos temporales y permanentes sobre la red de internet. Un ejemplo de ello es, el comercio electrónico, los tramites del gobierno, la información, el debate político, la educación desde el preescolares hasta al posgrado, la telemedicina, y el entretenimiento. En internet se disputan una férrea batalla de los medios electrónicos de comunicación los cuales han explotado una forma inusitada de comunicación, las fake-news, que han logrado desinformar y crear tensiones en la sociedad. Por ultimo los hábitos laborales, médicos, y científicos usan la red para comunicar, informar y transmitir todo tipo de mensajes y contenidos. En resumen, el COVID-19 nos obliga a experimentar una etapa de convivencia digital inédita y aun poco analizadas desde los estudios sociotécnicos en México.

El proyecto se presenta soportado por una experiencias claves. Instalar un laboratorio de producción de datos de Big Data realizados por el Área de Infonomía y El Grupo Académico de Estudios de Género de ECOSUR; realizado el mismo día en que se anunció la fase 1 del COVID-19 en México. La recolección de tweets del día 13 y 14 de marzo de 2020, recolecto 12 millones de tweets de todo el mundo los cuales fueron identificando los países, estados, temas, hashtags, bots e idiomas entre otras etiquetas. El segundo fue el laboratorio permitió capturar los mensajes generados los días 8 y 9 de marzo de 2020, 2021 y 2022, en las marchas feministas como en el día del paro de labores del #8M y #9M, respectivamente, tal y como se ve en la Figura 1.

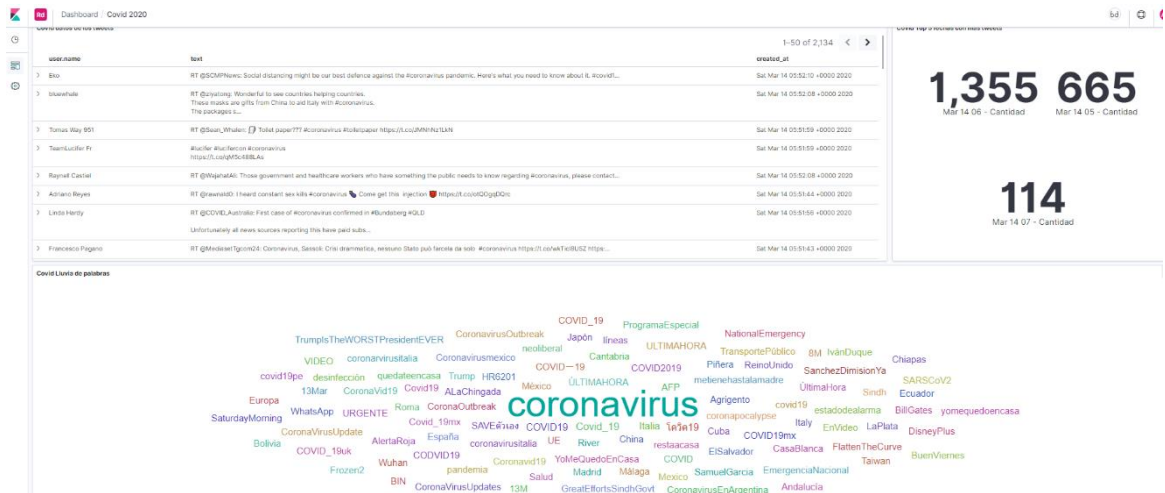


Figura 1 . Laboratorio piloto Big Data en ECOSUR. (Ventana de captura del laboratorio BIG DATA)

2.1 EL impacto del big data en la academia

A partir del surgimiento de la pandemia COVID-19 diversos artículos científicos han expuesto la importante tarea que se debe de realizar para la recolección de datos en el plano de lo clínico y salud social, usando para ello tecnologías Big Data. Los artículos tratan de cómo la información, los casos, los éxitos y fracasos en el tratamiento son compartidos en la red de internet y cosechados en tiempo real. (Martínez, 2020; Wang, C. J., Ng, C. Y., & Brook, R. H. 2020; Grasselli, G., Pesenti, A., & Cecconi, M., 2020; Wu, Z., & McGoogan, J. M. 2020) La presentación de las discusiones, noticias, e interacción de la población en torno a los fenómenos clínicos, socioeconómicos y culturales que la población genera a partir de la pandemia son hoy más que nunca visibles en la red de internet.

Por tanto, las ciencias sociales y las humanidades deberán iniciar su propio debate que por más de una década se ha pospuesto. Deberá poner ya un pie en el futuro por venir, en donde los ciudadanos que más haya de convertirse en datos y tendencias, en temas y hashtags, se configuran como protagonistas de temas, iniciativas, reclamos, miedos, denuncias y luchas en torno a los derechos humanos y hacktivismo dentro y fuera de la red (Gomez L y Mena.F 2019; Sancho, G. R. (2019).

Sin embargo, los y las científicas sociales poco han aprovechado las herramientas Big Data por considerarlas un tema de expertos tecnológicos y no de la interacción sociodigital de millones de personas que habitan el territorio de internet. (Gualda, E., & Díaz, C. R., 2020; Matas Terrón, A., Leiva Olivencia, J. J., Caballero, F., & Daniel, P. 2020; Vitón-Castillo, A. A., & Linares-Cánovas, L. P., 2020;). Los estudios se han confinado a la bibliotecología y al análisis del estudio de las publicaciones. (Terradillos, J. O., i Gras, J. M., & Ureta, A. L., 2020; Cadenas, L. E., 2020; Vessuri, H. (2020). Reyes Espejo, M. I., Cáceres Serrano, P., & Jiménez Vallejos, F., 2020;), y en casos más recientes al estudio de las redes sociales como Twitter, Facebook e Instagram (Sanchez-Duque, J. A., Gaviria-Mendoza, A., Moreno-Gutierrez, P. A., & Machado-Alba, J. E. (2020) Ardini, C., & Mirad, H. N. (2020)) los cuales son utilizados como territorios sociodigitales para conocer las tendencias en los temas que los usuarios usan, el consumo de noticias falsas, y la influencia que tienen los personajes con mayores seguidores.

El impacto se traduce en que la academia ofrezca a la ciudadanía y la comunidad científica y académica un portal web público y gratuito basada en una herramienta informática Big Data en donde se observe en tiempo real las tendencias, temas en debate, cuestionamientos, noticias, estrategias y el comportamiento de bots para la difusión de noticias falsas. Todo lo anterior a partir de las notificaciones e información que genera diariamente los programas de atención en materias de salud, educación, empleo y atención que despliega el gobierno federal.

Con base de la información generada en la plataforma se presenta el trabajo en el que un grupo de especialistas técnicos y académicos: 1) Mejoren su experiencia en el uso de Tecnologías Big Data. 2) Realicen análisis de la información de manera cualitativa. 3) Ofrezcan a la ciudadanía un manual para que por sí mismo tengan el acceso y realicen análisis propios de la información generada en Big Data, además de contar con la síntesis del análisis de 8 expertos en temas relacionados al COVID-19 y su impacto a: 1) Derechos Humanos, 2) Salud, 3) Género, 4) Trabajo, 5) Educación, Tecnología, 6) Violencia en medios digitales, 7) información para el periodismo y 8) Desarrollo de tecnologías Big Data.

El portal público ciudadano de Big Data, con su característica de libre acceso, sería un referente que informe en tiempo real sobre el diálogo que en la red de internet se hace en temas relacionados al COVID-19 y su relación con otros temas, emergentes como lo son la propagación de noticias falsas, las oportunidades de empleo, los tratamientos y rutas de atención, la telemedicina y teleeducación, la ciber vigilancia y movilidad, los mitos y realidades del COVID-19 entre otros temas.

La arquitectura para el proceso de la información es de una construcción discreta, tal y como se observa en la figura 2, está basada en un esquema simple de utilización de recursos en la Nube, para una configuración inicial se plantea la utilización de una máquina virtual con un disco duro el cual tomara el rol de servidor, ambos resguardados en la nube. Dicho servidor tendrá instalado la tecnología de contenedores Docker y dentro de si los contenedores de Elasticsearch, Logstash y Kibana siendo este último nuestra interfaz de presentación para el usuario final a través de internet. Si bien la configuración se visualiza de forma sencilla, dada la capacidad de la nube y de la tecnología de contenedores con simples configuraciones es fácil realizar un escalamiento horizontal o vertical de nuestra arquitectura, según los requerimientos de información y almacenamientos que así sean necesarios para suplir la necesidad del proyecto.

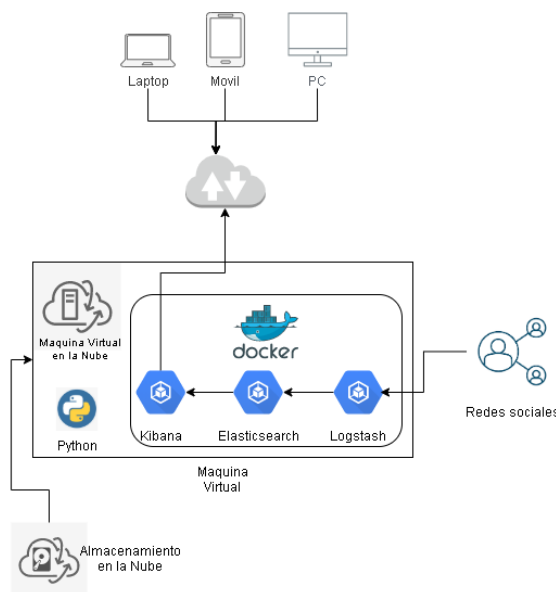


Figura 2. Esquema que ejemplifica la arquitectura de sistema de cosecha de datos. (*Esquema de flujo y presentación de información del BIG DATA*)

3.8M y 9M. De las calles al Big Data

El laboratorio y el análisis del Big Data permite que desde múltiples disciplinas, enfoques y posiciones onto epistémicas, reflexionemos acerca de la experiencia socio digital en la cosecha de datos masivos desde las redes sociales, además de contar con una visualización sencilla y accesible para el público en general para su posterior análisis en el campo de las ciencias sociales. Inédito fue en las redes el debate establecido por los #8M y #9M, e igual de inédito fue para un amplio sector de la academia analizar la información que se produce en las redes sociales en México a partir de tecnologías Big Data.

Los resultados de la investigación soportada por el primer laboratorio de Big Data realizados por el Área de Infonomía y El Grupo Académico de Estudios de Género de ECOSUR que permitió capturar o cosechar los mensajes generados los días 8 y 9 de marzo de 2020, en las marchas feministas como en el día del paro de labores del #8M y #9M, respectivamente en donde se recolectaron más de 2 millones de twitts en español en tres días de cosecha, y el 14 de marzo, día en que entro la fase 1 de la emergencia sanitaria en México por el COVID-19 cosechando más de 12 millones de twitts. Como se muestran en las siguientes imágenes. La cosecha es una magnífica oportunidad consolidar el uso de herramientas basadas en tecnológicas Big Data y visualizados en Elastic Stack, como herramientas para los análisis cualitativos basada en productos y estrategias digitales en donde se promueva el trabajo y el análisis académicos y técnicos que analicen la información y la ofrezcan directamente a la población, tal y como se observa en la Figura 3.

La ponencia resuelve que las tecnologías de información deberán ser comprendidas, analizadas y presentadas para las futuras tomas de decisiones en la transformación de los clásicos conceptos de individualidad, privacidad, comunidad y solidaridad derivados de los reclamos de la sociedad, pero también del uso que la propia sociedad hace desde cualquier espacio de la red, para informarse, acceder a oportunidades y no quedar aislado de la información de nuestro régimen democrático, el cual queda en ocasiones atrapado en la desinformación por medio de noticias falsas (fake-news) en la red y con esto en la sociedad.

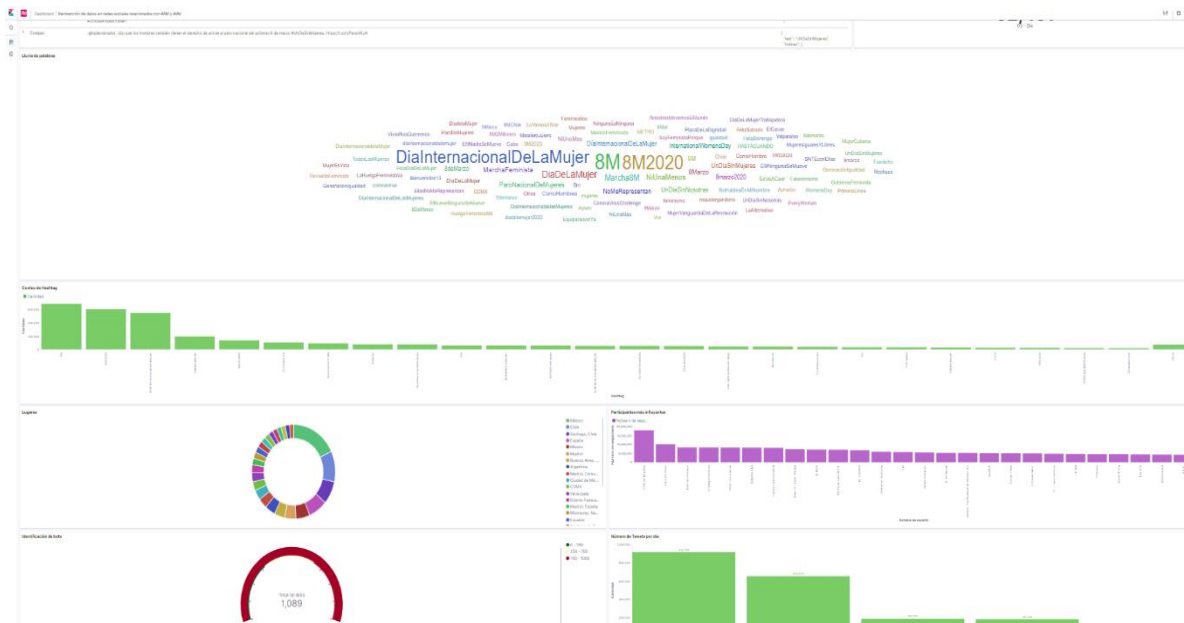


Figura 3. Sitio en Elastic Stack con tecnología Big Data en donde se reporta la discusión en Twitter del #8M día de las mujeres y #9M el paro de mujeres. ECOSUR. *(Ventana de visualización del sitio en donde se realizan las búsquedas de información)*

4. Consideraciones finales

Las ciencias sociales y las humanidades deberán iniciar su propio debate que por más de una década se ha pospuesto. Deberá poner ya un pie en el futuro por venir, en donde los ciudadanos puedan acceder a los sistemas que les permitan visualizar ejercicios de ciencia de datos y con eso puedan analizar, convertir y acceder a datos y tendencias mediante temas y hashtags, se configuran como protagonistas de temas, iniciativas, reclamos, miedos, denuncias y luchas en torno a los derechos humanos y hacktivism dentro y fuera de la red (Gomez L y Mena.F 2019; Sancho, G. R. (2019).

El resultado tangible es un portal público ciudadano que informe en tiempo real sobre el diálogo que en la red de internet a partir de los principales debates en las redes sociales en torno a las dos debates nacionales que se concentran en el #8M y #9M, a partir de análisis de la información cosechada, optimizada y organizada, con la cual el ciudadano podrá hacer búsquedas, filtros, a partir de la geolocalización, hashtags, bots, fechas y demás opciones posibles en el diseño del portal con Elastic Stack.

El usuario final será el ciudadano que contará con una herramienta informática accesible, mediante el acceso a un portal de acceso público realizado con Elastic Stack, en donde pueda observar en tiempo real las tendencias, temas en debate, cuestionamientos, noticias falsas, notificaciones y toda la información que genere el gobierno, las instituciones y el público en general. Además, se desarrollarán habilidades locales y nacionales sobre la tecnología Big Data en los centros públicos y universidades participantes en el proyecto.

5. Agradecimientos:

Este trabajo se hace con la colaboración, trabajo y conocimientos de la Unidad de tecnologías de información en su área de Infonomía. Se agradece la colaboración de los maestros en ciencias Isidra Ocampo Guzman, Eder Valente Toledo Nuñez y Héctor Augusto Santiago Ruiz.


6. Referencias

- Martínez, 2020; Wang, C. J., Ng, C. Y., & Brook, R. H. 2020; Grasselli, G., Pesenti, A., & Cecconi, M., 2020; Wu, Z., & McGoogan, J. M. 2020:
- Martínez, R. M. (2020). Los tratamientos de datos personales en la crisis del covid-19. Un enfoque desde la salud pública. *Diario La Ley*, (9601), 2.
- Wang, C. J., Ng, C. Y., & Brook, R. H. (2020). Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing. *JAMA*.
- Grasselli, G., Pesenti, A., & Cecconi, M. (2020). Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. *JAMA*.
- Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *Jama*.
- Gualda, E., & Díaz, C. R. (2020). Big Data y Twitter para el estudio de procesos migratorios: Métodos, técnicas de investigación y software. *Empiria: Revista de metodología de ciencias sociales*, (46), 147-177.
- Cadenas, L. E. (2020). El rol de las redes nacionales de investigación y educación en las Ciencias Sociales. *Anuario Electrónico de Estudios en Comunicación Social "Disertaciones"*, 13(1).
- Matas Terrón, A., Leiva Olivencia, J. J., Caballero, F., & Daniel, P. (2020). Big Data Irruption in Education. *Pixel-Bit: Revista de Medios y Educación*, 57, 59-90.
- Sanchez-Duque, J. A., Gaviria-Mendoza, A., Moreno-Gutierrez, P. A., & Machado-Alba, J. E. (2020). Big data, farmacoepidemiología y farmacovigilancia. *Revista de la Facultad de Medicina*, 68(1).
- Vitón-Castillo, A. A., & Linares-Cánovas, L. P. (2020). Big data en el contexto de la salud cubana. *Revista Cubana de Salud Pública*, 45, e2012.
- Vessuri, H. (2020). Una historia de la verdad en Occidente. Ciencia, arte, religión y política en la conformación de la cosmología moderna: [A history of truth in the west. Science, art, religion and politics in the construction of modern cosmology], by Mauricio Nieto Olarte, Bogota, Universidad de los Andes Colombia/Facultad de Ciencias Sociales-Fondo de Cultura Económica, 2019, 589 pp., 70.000 colombian pesos (paperback), ISBN 978-958-8249-48-3.

- Reyes Espejo, M. I., Cáceres Serrano, P., & Jiménez Vallejos, F. (2020). Organización social y construcción de subjetividades en los nuevos escenarios sociales en América Latina. *Psicoperspectivas*, 19(1), 1-5.
- Terradillos, J. O., i Gras, J. M., & Ureta, A. L. (2020). Desinformación en redes sociales: ¿compartimentos estancos o espacios dialécticos? El caso Luther King, Quim Torra y El Confidencial. *Revista Mediterránea de Comunicación*.
- Sancho, G. R. (2019). Constelaciones performativas y multitudes urbanas: el activismo en red, la sensibilidad feminista y la contrainsurgencia. *Desacatos: Revista de Ciencias Sociales*, (61), 40-55.
- García, S. A., Carabel, T. C., & Martínez, N. O. (2020). # ElDilemaSalvados, análisis de las reacciones en Twitter al programa de Jordi Évole sobre Cataluña. *Revista Prisma Social*, (28), 1-19.
- Gómez, D. M. L., & Farrera, R. A. M. (2019). El hacktivismo e Internet como territorio en disputa. Una mirada desde los marcos de accion colectiva. *Estudios Políticos*, 9(48), 115-132.
- Ardini, C., & Mirad, H. N. (2020). El uso del big data en política o la política del big data. *Comunicación y Hombre*, 2020(16).
- Mujica, R. Z. Ciencia abierta y datos enlazados.
- Miquel Segarra, S., & Aced Toledano, C. (2020). Big data: la revolución de los datos y su impacto en la comunicación corporativa.

Abraham Mena Farrera

El Colegio de la Frontera Sur, México

 <https://orcid.org/0000-0002-2242-3210>✉ rmena@ecosur.mx**Angelica Evangelista García**

El Colegio de la Frontera Sur, México

 <https://orcid.org/0000-0002-4460-854X>✉ aevangel@ecosur.mx**Fecha de envío:** 02/2022**Fecha de evaluación:** 04/2022**Fecha de publicación:** 07/2022