

# MODELAMIENTO DE TÓPICOS UTILIZANDO MENSAJES DE TWITTER RELACIONADOS CON EL CÁNCER CERVICAL

RUTH MARÍA REÁTEGUI ROJAS

[rmreategui@utpl.edu.ec](mailto:rmreategui@utpl.edu.ec)

<https://orcid.org/0000-0002-7304-4413>

Universidad Técnica Particular de Loja, Ecuador

**RESUMEN.** El cáncer cervical es un problema de salud mundial que genera una gran cantidad de información que circula por las redes sociales. El modelado de tópicos permite automáticamente identificar aquellos que tratan sobre un tema en un conjunto de documentos. En el presente trabajo se ha aplicado el modelamiento de tópicos para identificar los tópicos de un conjunto de tuits que tratan sobre al cáncer cervical. Se aplicó el algoritmo LDA y la métrica de la coherencia para la evaluación. Se identificaron siete tópicos relacionados con el efecto de las vacunas contra el VPH, la relación que existe entre el VPH y otras enfermedades, las formas de prevención como vacunas y test de Papanicolaou, programas que prestan servicios médicos para la prevención y eliminación de esta enfermedad, historias de mujeres que han padecido de cáncer cervical y estudios dirigidos a mujeres latinas.

**PALABRAS CLAVE:** text mining, Twitter, cáncer cervicouterino, modelado de tópicos

## TOPIC MODELING USING TWITTER MESSAGES RELATED TO CERVICAL CANCER

**ABSTRACT.** As a global health problem, cervical cancer generates much information that circulates through social networks. Modeling allows us to automatically identify the topics that deal with a specific subject matter in a set of documents. This research used the LDA algorithm and the coherence metric for topic modeling and identified seven topics in a set of tweets on cervical cancer. The topics were related to the effect of HPV vaccines, the relationship between HPV and other diseases, forms of prevention such as vaccines and Papanicolaou tests, programs that provide medical services for the prevention and elimination of this disease, stories of women who have had cervical cancer and studies aimed at Latina women.

**KEYWORDS:** text mining, Twitter, cervical cancer, topic modeling

## 1. INTRODUCCIÓN

El cáncer cervical es el cuarto tipo de cáncer más frecuente en las mujeres a nivel mundial, pero genera un mayor impacto en países de bajos recursos económicos. Este tipo de cáncer afecta al cuello del útero y el 99 % de las veces se debe al virus del papiloma humano (VPH) (Organización Mundial de la Salud, 2022). Entre los tipos de VPH considerados de alto riesgo están el 16, 18, 31, 33, 45, 52 y 58, mientras que los tipos considerados de bajo riesgo son el 6 y el 11.

El cáncer cervical es prevenible y curable si es diagnosticado a tiempo. Como método de prevención primaria, existe la vacuna contra el VPH que se recomienda aplicarla antes del inicio de la vida sexual. Además, como método secundario, se tiene el tamizaje y tratamiento de lesiones precancerosas. A pesar de la existencia de estos métodos, la desigualdad económica, social, cultural y racial incide en el acceso a servicios médicos, incluso en los países desarrollados (Hernández et al., 2021).

Puesto que el cáncer cervical es un problema de salud mundial, son muchos los programas, estudios y eventos que se realizan en torno a esta enfermedad. A través de las redes sociales (por ejemplo, Twitter), se puede observar cómo el personal del área de la salud, los investigadores e incluso los pacientes comparten temas de interés. Las redes sociales se han convertido en una fuente popular para la difusión y debate de información sobre salud (Lyu et al., 2021), pues contienen opiniones que permiten comprender la percepción popular sobre temas como diagnóstico médico, medicamentos, instalaciones y reclamos (Asghari et al., 2019); además, actúan como un sensor del comportamiento de una persona (Hinduja et al., 2022) o incluso para establecer comunidades de soporte para ciertas enfermedades (Jansli, 2022).

Debido a la gran cantidad de información que puede circular por las redes sociales, la aplicación de la inteligencia artificial y sus diferentes ramas, como la minería de texto y procesamiento de lenguaje natural, hace posible extraer información para inferir conocimiento relevante en torno a un determinado tema. El modelado de tópicos (*topic modeling*, en inglés) permite automáticamente identificar temas o tópicos de un conjunto de documentos con formato no estructurado. La asignación latente de Dirichlet (LDA, por sus siglas en inglés) es un modelo probabilístico generativo de tópicos, en el cual los documentos se representan como una combinación aleatoria sobre tópicos latentes, y cada tópico es representado por una distribución de probabilidades de palabras (Blei et al., 2003).

El modelado de tópicos ha sido utilizado ampliamente para obtener información de las redes sociales en temas relacionados con la salud. Musaeu et al. (2019) examinan las conversaciones de Twitter sobre la salud cardiovascular para

evaluar el enfoque temático de estas conversaciones, así como el papel de varios departamentos de salud estatales. Alahmari et al. (2022) aplican modelamiento de tópicos para analizar la red social Twitter en temas relacionados con el cáncer. Además, Joshi et al. (2022) proponen un modelo basado en LDA para extraer e identificar las reacciones adversas de medicamentos basado en *posts* provenientes de las redes sociales. El trabajo presentado por Sidana et al. (2018) se enfoca en el monitoreo de la salud de las personas a lo largo del tiempo en la red Twitter y utiliza algunos modelos de tópicos.

En los últimos años, también se han desarrollado varios trabajos que utilizan el modelado de tópicos para el análisis de los tuits relacionados con la pandemia del COVID-19. El estudio de Ma et al. (2021) aplicó dos métodos de modelado de tópicos: LDA y Top2Vec, para analizar comentarios de usuarios con respecto a las dudas sobre la vacunación. Kwok et al. (2021) y Bogdanowicz y Guan (2022) utilizaron el modelo de tópicos para identificar temas que se discuten respecto al COVID-19 tanto en Australia como en Estados Unidos, respectivamente. Lyu et al. (2021) emplearon LDA para identificar percepciones, actitudes y preocupaciones relacionadas con el vapeo, y para discernir posibles desinformaciones y conceptos erróneos sobre esta acción durante la pandemia. Otro reciente estudio publicado por Oliveira et al. (2022) usó el mismo algoritmo LDA para analizar nuevos temas en tuits y luego proceder a identificar emociones en respuestas a tuits de noticias.

En vista de la utilidad del modelado de tópicos en el área de la salud, en el presente trabajo se propone la aplicación del modelo LDA para identificar temas relacionados con el cáncer cervical en los mensajes de la red social Twitter.

## 2. METODOLOGÍA

### 2.1 Obtención de datos

Los datos fueron recolectados de la red social Twitter en los meses de noviembre y diciembre del 2021. Se buscó tuits con las palabras *cervical cancer* en el idioma inglés, puesto que se constató que hay una mayor interacción en esta lengua. Se obtuvo un total de 630 registros únicos, y para el análisis se utilizó el contenido textual de los tuits.

### 2.2 Preprocesamiento

En este trabajo se utilizó el lenguaje Python, que proporciona varias librerías relacionadas con el procesamiento de lenguaje natural y la aplicación del algoritmo LDA. Durante el preprocesamiento, se usaron las librerías NLTK y Spacy. Algunos pasos de esta fase se detallan en la Tabla 1.



que cada una aparece en el documento y su frecuencia en el conjunto de documentos. De esta forma, no solo las palabras más comunes serán las más relevantes.

## 2.4 Identificación de tópicos

Una vez obtenidos los valores TF-IDF, se procedió a aplicar el modelo LDA definido en la librería Gensim. LDA maneja los parámetros alfa y beta; para el presente trabajo, se consideraron los valores que vienen establecidos por defecto. Alfa se refiere a la distribución de documentos por tema; y beta, a la distribución de palabras por tema (Blei et al., 2003). Otro de los parámetros es el valor  $k$ , que hace referencia al número de tópicos que se desea obtener. Para conocer cuál es el mejor valor de  $k$ , se utilizó la métrica de coherencia.

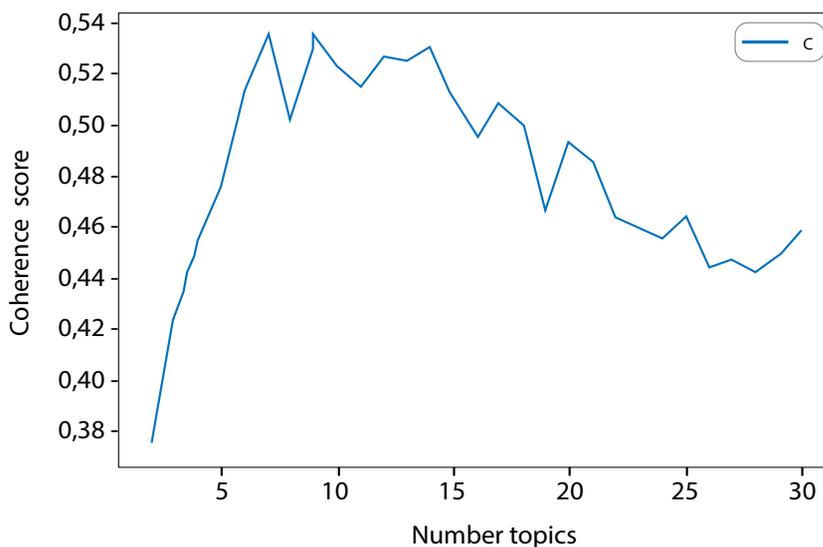
## 2.5 Evaluación

La métrica de coherencia permite calificar un tema o tópico midiendo el grado de similitud semántica entre las palabras de alto puntaje en el tema. Se utilizó, igualmente, la librería Gensim que se configuró para que se use la métrica 'c\_v'. Esta métrica trabaja con una ventana deslizante y combina la medida del coseno de similitud con la información mutua normalizada por puntos (Röder et al., 2015; Syed & Spruit, 2017).

La Figura 2 muestra los valores de la coherencia para valores de  $k$  que van de 2 a 30. Se puede observar que el valor más alto de la coherencia (0,54) se obtiene cuando  $k = 7$ . Por tanto, se procedió a identificar siete tópicos.

Figura 2

Métrica de coherencia



### 3. RESULTADOS

Una vez identificado el mejor número de tópicos  $k = 7$ , las palabras que corresponden a cada tópico se encuentran en la Tabla 2. Los números que preceden a cada término son la probabilidad de que cada término corresponda a dicho tópico.

La Tabla 3 muestra la cantidad de tuits por cada tópico; se ha tomado en cuenta el tópico dominante en cada tuit. Se puede observar que la mayoría de los tuits corresponden a los tópicos 2 y 3.

**Tabla 2**

*Tópicos identificados y palabras por tópico*

Número de tópico	Palabras por cada tópico
1	0,014**"vaccination" + 0,013**"may" + 0,012**"study" + 0,012**"prevention" + 0,012**"smoke" + 0,011**"reduce" + 0,010**"need" + 0,010**"support" + 0,010**"risk" + 0,010**"late"
2	0,024**"vaccine" + 0,019**"hpv" + 0,017**"woman" + 0,013**"get" + 0,011**"pap" + 0,011**"tell" + 0,011**"live" + 0,011**"year" + 0,010**"prevent" + 0,010**"help"
3	0,024**"screening" + 0,018**"perspective" + 0,017**"health" + 0,011**"campaign" + 0,010**"survivor" + 0,009**"available" + 0,008**"join" + 0,008**"still" + 0,007**"develope" + 0,007**"session"
4	0,020**"life" + 0,016**"elimination" + 0,013**"test" + 0,010**"breast" + 0,009**"save" + 0,009**"clinical" + 0,009**"important" + 0,009**"come" + 0,009**"radiomic" + 0,008**"screen"
5	0,018**"hispanic" + 0,017**"go" + 0,017**"cervicalcancer" + 0,014**"give" + 0,014**"treatment" + 0,012**"pregnant" + 0,010**"week" + 0,009**"know" + 0,008**"oncology" + 0,008**"cause"
6	0,018**"new" + 0,018**"photographer" + 0,016**"be" + 0,014**"stage" + 0,011**"work" + 0,009**"story" + 0,009**"journey" + 0,008**"release" + 0,007**"australian" + 0,007**"challenge"
7	0,023**"global" + 0,016**"doctor" + 0,014**"die" + 0,014**"day" + 0,012**"reason" + 0,010**"could" + 0,010**"early" + 0,009**"action" + 0,009**"leader" + 0,008**"earlystage"

**Tabla 3**

*Documentos por tópico*

Número de tópico	Número de tuits
1	78
2	161
3	105
4	74
5	89

(continúa)

(continuación)

Número de tópico	Número de tuits
6	73
7	50
Total	630

#### 4. DISCUSIÓN DE LOS RESULTADOS

Una de las dificultades de este tipo de modelos es la interpretación humana que se puede dar a los diferentes tópicos encontrados. Con base en los términos que forman los tópicos y una revisión manual de algunos tuits que tienen como dominante el tópico que se va a analizar, se presenta a continuación la interpretación de cada uno de los tópicos. La Tabla 4 ofrece algunos ejemplos de tuits por tópico.

- *Tópico 1.* Entre las palabras que conforman este tópico están *vaccination, may, study, prevention, smoke, reduce, need, support, risk, late*. Este tópico principalmente abarca tuits que tratan sobre los efectos de la vacunación contra el cáncer cervical. También varios textos hacen referencia a estudios relacionados con el cáncer cervical.
- *Tópico 2.* Algunas de las palabras que se encuentran en este tópico son *vaccine, hpv, woman, get, pap, tell, live, year, prevent, help*. Dentro de este tópico se encuentran tuits que mencionan algún tipo de relación entre los virus de VPH y el virus de la inmunodeficiencia humana (VIH). Además, se comenta sobre algunas formas de prevención como la vacuna y los test de Papanicolaou.
- *Tópico 3.* Algunas de las palabras que forman parte de este tópico son *screening, perspective, health, campaign, survivor, available, join, still, develop, session*. Dentro de este tópico se encuentran tuits que comentan sobre programas o instituciones que prestan servicios para la prevención del cáncer cervical; entre los servicios que se mencionan está el tamizaje (*screening*, en inglés).
- *Tópico 4.* Entre las palabras que conforman este tópico están *life, elimination, test, breast, save, clinical, important, come, radiomic, screen*. Este tópico hace referencia a los exámenes de prevención para el cáncer cervical y de seno. Menciona algunos estudios o programas en países de bajos recursos para eliminar el cáncer cervical, así como estudios científicos basados en radiómica.
- *Tópico 5.* En este tópico se encuentran palabras como *hispanic, go, cervical-cancer, give, treatment, pregnant, week, know, oncology, cause*. Este tópico

abarca tuits que hablan sobre algunos estudios o programas dirigidos a mujeres latinas o hispanas.

- *Tópico 6.* Entre las palabras que conforman este tópico están *new, photographer, be, stage, work, story, journey, release, Australian, challenge*. Este tópico abarca tuits que tratan de historias de mujeres que han padecido de cáncer cervical. Además, algunos textos comentan sobre nuevos estudios, programas o historias relacionados con el cáncer cervical.
- *Tópico 7.* Algunas de las palabras que conforman este tópico son *global, doctor, die, day, reason, could, early, action, leader, earlystage*. Este tópico considera tuits que tratan sobre programas o acciones globales para eliminar el cáncer cervical. Además, nombran algunos casos de personas que han fallecido por este tipo de cáncer.

**Tabla 4**

*Tuits por tópicos dominantes*

Tópico	Tuits
1	<p>The effects of the national HPV vaccination programme in England, UK, on cervical cancer and grade 3 cervical intraepithelial neoplasia incidence: a register-based observational study <a href="https://t.co/HrRnTAL9wz">https://t.co/HrRnTAL9wz</a></p> <p>A new study reports a 38 % drop in cervical cancer incidence and a 43 % decline in mortality among young women and girls after HPV vaccination was introduced in the U.S. <a href="https://t.co/1wt929xQRO">@Medscape @slusom</a></p>
2	<p>Women living with HIV/AIDS have up to 5 times higher risk of acquiring cervical cancer compared to those without. This World AIDs day, let us regularize cervical cancer screening. We also call upon Government and partners to integrate HIV/AIDs and Cancer services. <a href="https://t.co/XvDvRDZNbh">https://t.co/XvDvRDZNbh</a></p> <p>Women with HIV should be screened regularly for HPV. Women have shown a higher preference to urine for HPV testing for cervical cancer. Read more in our blog - <a href="https://t.co/ouURCvhKKu">#WorldAIDSDay</a></p>
3	<p>Cancer is preventable, treatable, &amp; beatable. Erie County Cancer Services provides free breast, cervical, &amp; colon cancer screenings &amp; education to eligible un- &amp; under-insured men &amp; women. More info at: <a href="https://t.co/2MfXrISsGI">https://t.co/2MfXrISsGI</a></p> <p>"Title X is the only federal grant program that funds family planning and counseling programs to help patients access contraception, as well as breast and cervical cancer screenings, screenings and treatments for sexually transmitted ... <a href="https://t.co/QhaEI73tMr">https://t.co/QhaEI73tMr</a> <a href="https://t.co/Q83tQT9r5I">https://t.co/Q83tQT9r5I</a>"</p>
4	<p>"Day by day women are getting aware about Breast and Cervical Cancer early detection and prevention. Canapp is Playing an important role in educating women about Breast Self Exam Thanks to @madhuchauhanbjp for the initiative and helping us and encouraging women to come for checkup <a href="https://t.co/icPrTmlrFW">https://t.co/icPrTmlrFW</a>"</p> <p>"Machine learning based evaluation of clinical and pretreatment (18)F-FDG-PET/CT radiomic features to predict prognosis of cervical cancer patients <a href="https://t.co/QqIFXO2hVH">https://t.co/QqIFXO2hVH</a>"</p>

(continúa)

(continuación)

Tópico	Tuits
5	<p>Hispanic women born outside the U. S. have less frequent cervical cancer screenings than Hispanic women born in the U. S.</p> <p>Hispanic women who live in Tarrant County are more likely to die from cervical cancer than Hispanic women who live in Dallas County. for @FortWorthReport  <a href="https://t.co/cZAMzSw5bW">https://t.co/cZAMzSw5bW</a></p>
6	<p>An inspiring story from HK-based Australian photographer Julia Broad, who has just gone through a tough cancer recovery journey <a href="https://t.co/DlpuHdCCm9">https://t.co/DlpuHdCCm9</a></p> <p>A GoFundMe has been started for an #EnfieldNS single mom of four, Donna Boutilier, who has had her cervical cancer return. Here's how you can help her out: #EastHants #Halifax #NS <a href="https://t.co/lGj5Qxbxvo">https://t.co/lGj5Qxbxvo</a></p>
7	<p>Global leaders call for cervical cancer elimination on Day of Action <a href="https://t.co/FK8f0UzShu">https://t.co/FK8f0UzShu</a>  <a href="https://t.co/hkczFjorjc">https://t.co/hkczFjorjc</a></p> <p>CA587:A621horeography written by: William Masters As soon as Albert Andrew Fischer's wife died of cervical cancer, after a final, twenty-two-day hospitalization, he felt relief. Such relief derived not from the end of his <a href="https://t.co/AjBteLwFne">https://t.co/AjBteLwFne</a></p>

## 5. CONCLUSIONES

Actualmente, en las redes sociales como Twitter, profesionales de la salud, pacientes e investigadores comparten información de su preferencia. Esta puede ser analizada automáticamente con técnicas de inteligencia artificial y generar conocimiento de interés para el área de la salud. Aplicando el algoritmo LDA, fue posible identificar cuáles son los temas principales o tópicos que se comenta en Twitter. Entre los temas de interés están el efecto de las vacunas contra el VPH (tópico 1), la relación que existe entre el VPH y otras enfermedades como el sida (tópico 2) y el cáncer de seno (tópico 4), formas de prevención como vacunas y test de Papanicolaou (tópico 2), programas o instituciones que prestan servicios médicos para la prevención del cáncer cervical (tópico 3), programas o estudios para eliminar esta enfermedad (tópico 4, tópico 7), historias de mujeres que han padecido de cáncer cervical (tópico 6) y estudios dirigidos a mujeres latinas o hispanas (tópico 5).

En este estudio, a pesar de que se consideraron tuits en inglés, los tópicos identificados pueden servir de ejemplo para que instituciones de salud y gobiernos puedan difundir información que sea de mayor interés para sus pacientes y para definir estrategias de prevención de este tipo de enfermedades. Además, la Organización Panamericana de la Salud recalca que la mortalidad por cáncer cervical es tres veces más alta en América Latina y el Caribe en comparación con Norteamérica. Este tipo de desigualdad en la salud se mencionó en uno de los tópicos identificados en este trabajo.

## 6. TRABAJO FUTURO

Tomando en cuenta el alto nivel de muertes entre las mujeres hispanas por este tipo de cáncer, en un inicio se consideró analizar cuáles son los temas que en Twitter se comentan sobre esta enfermedad en Ecuador o Latinoamérica, pero al no haber una cantidad importante de tuits se realizó el análisis de tuits en inglés, según lo indicado en la sección 2. Por tanto, para un trabajo futuro, se puede intentar analizar otras redes sociales donde haya mayor participación, o también la información sobre otro tipo de enfermedades, de acuerdo con las necesidades locales o nacionales. Esto puede dar respuesta a algunas preguntas de investigación como las planteadas a continuación:

- ¿Cuáles son los tópicos de interés para los pacientes que se discuten en una red social determinada?
- ¿Cuáles son los tópicos de interés para los médicos que se discuten en una red social determinada?
- ¿Qué tópicos discutidos en una red social pueden orientar a la creación de políticas públicas?

## REFERENCIAS

- Alahmari, N., Alswedani, S., Alzahrani, A., Katib, I., Albeshri, A., & Mehmood, R. (2022). Musawah: A data-driven AI approach and tool to co-create healthcare services with a case study on cancer disease in Saudi Arabia. *Sustainability*, 14(6). DOI: 10.3390/su14063313
- Asghari, M., Sierra-Sosa, D., & Elmaghraby, A. (2019). Trends on health in social media: Analysis using Twitter topic modeling. En *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)* (pp. 558-563). <https://doi.org/10.1109/ISSPIT.2018.8642679>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Bogdanowicz, A., & Guan, C. (2022). Dynamic topic modeling of Twitter data during the COVID-19 pandemic. *PLoS ONE*, 17(5). DOI: 10.1371/journal.pone.0268669
- Hinduja, S., Afrin, M., Mistry, S., & Krishna, A. (2022). Machine learning-based proactive social-sensor service for mental health monitoring using twitter data. *International Journal of Information Management Data Insights*, 2(2). DOI: 10.1016/j.jjime.2022.100113

- Hernández, J. A., Ramírez, P. X., Valbuena-García, A. M., Acuña, L., & González-Díaz, J. A. (2021). Factors associated with delays in time to treatment initiation in Colombian women with cervical cancer: A cross-sectional analysis. *Gynecologic Oncology Reports*, 35. <https://doi.org/10.1016/j.gore.2021.100697>
- Jansli, S. M., Hudson, G., Negbenose, E., Erturk, S., Wykes, T., & Jilka, S. (2022). Investigating mental health service user views of stigma on twitter during COVID-19: A mixed-methods study. *Journal of Mental Health*, 31(4), 576-584. doi: 10.1080/09638237.2022.2091763
- Joshi, C., Attar, V. Z., & Kalamkar, S. P. (2022). An unsupervised topic modeling approach for adverse drug reaction extraction and identification from natural language text. En S. Tiwari, M. C. Trivedi, M. L. Kolhe, K. Mishra & B. K. Singh (Eds.), *Advances in data and information sciences. Lecture notes in networks and systems* (vol. 318, pp. 505-514). Springer. [https://doi.org/10.1007/978-981-16-5689-7\\_44](https://doi.org/10.1007/978-981-16-5689-7_44)
- Kwok, S. W. H., Vadde, S. K., & Wang, G. (2021). Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: Machine learning analysis. *Journal of Medical Internet Research*, 23(5). <https://doi.org/10.2196/26953>
- Lyu, J. C., Luli, G. K., & Ling, P. M. (2021). Vaping discussion in the COVID-19 pandemic: An observational study using twitter data. *PLoS ONE*, 16(12). <https://doi.org/10.1371/journal.pone.0260290>
- Ma, P., Zeng-Treitler, Q., & Nelson, S. J. (2021). Use of two topic modeling methods to investigate covid vaccine hesitancy. En *International Conferences ICT, Society, and Human Beings 2021; Web Based Communities and Social Media 2021; and e-Health 2021* (pp. 221-226). [https://www.ict-conf.org/wp-content/uploads/2021/07/04\\_202106C030\\_Ma.pdf](https://www.ict-conf.org/wp-content/uploads/2021/07/04_202106C030_Ma.pdf)
- Musaeu, A., Britt, R. K., Hayes, J., Britt, B. C., Maddox, J., & Sheinidashtegol, P. (2019). Study of twitter communications on cardiovascular disease by state health departments. En J. Miller, E. Stroulia, K. Lee & L. J. Zhang (Eds.), *Web Services - ICWS 2019. Lecture notes in computer science* (vol. 11512). [https://doi.org/10.1007/978-3-030-23499-7\\_12](https://doi.org/10.1007/978-3-030-23499-7_12)
- Oliveira, F. B., Haque, A., Mougouei, D., Evans, S., Sichman, J. S., & Singh, M. P. (2022). Investigating the emotional response to COVID-19 news on twitter: A topic modelling and emotion classification approach. En *IEEE Access* (vol. 10, pp. 16883-16897). <https://doi.org/10.1109/ACCESS.2022.3150329>

- Organización Mundial de la Salud. (2022). *Cáncer cervicouterino*. [https://www.who.int/es/health-topics/cervical-cancer#tab=tab\\_1](https://www.who.int/es/health-topics/cervical-cancer#tab=tab_1)
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. En *Proceedings of the 8th ACM International Conference on Web Search and Data Mining (WSDM '15)* (pp. 399-408). <https://doi.org/10.1145/2684822.2685324>
- Sidana, S., Amer-Yahia, S., Clausel, M., Rebai, M., Mai, S. T., & Amini, M. (2018). Health monitoring on social media over time. *IEEE Transactions on Knowledge and Data Engineering*, 30(8), 1467-1480. <https://doi.org/10.1109/TKDE.2018.2795606>
- Syed, S., & Spruit, M. (2017). Full-text or abstract? Examining topic coherence scores using Latent Dirichlet Allocation. En *IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 165-174). <https://doi.org/10.1109/DSAA.2017.61>