

Análisis hashtag #Reto2020 en Instagram

1. Descripción del proyecto

El presente trabajo está enfocado en el análisis de las publicaciones que incluyeron el #Reto2020, para identificar palabras claves, temas relacionados y detectar patrones en las publicaciones ya sea en el contenido de la descripción como en las imágenes que se comparten.

2. Fuentes de datos

La fuente primaria de información para esta investigación fueron las publicaciones que usaron #Reto2020 en Instagram, ya que a través de este objeto de estudio se recuperaron los siguientes contenidos:

- Número de publicaciones con #Reto2020.
- Alcances o relaciones con otros hashtag similares.
- Palabras claves y coincidencias.
- Imágenes relacionadas.

3. Tecnologías elegidas y resultados obtenidos

A continuación se agrega el listado de las herramientas con las que se realizó el análisis del #Reto2020

- a) **Google Trends.** El resultado obtenido demuestra que de marzo a julio hubo un pico en la búsqueda del tema Reto2020 en la web en los países de México, Costa Rica, Venezuela y Colombia, coincide con los meses en los que hubo mayor confinamiento.

Enlace de exploración: <https://bit.ly/2ITx3WD>

- b) **Nube de palabras.** Se ingresó la data inicial recopilada con #reto2020 (sin limpieza - 4500 publicaciones hechas durante el mes de noviembre). La exploración reveló las palabras con mayor frecuencia, lo cual ayudó a detectar los tópicos que están

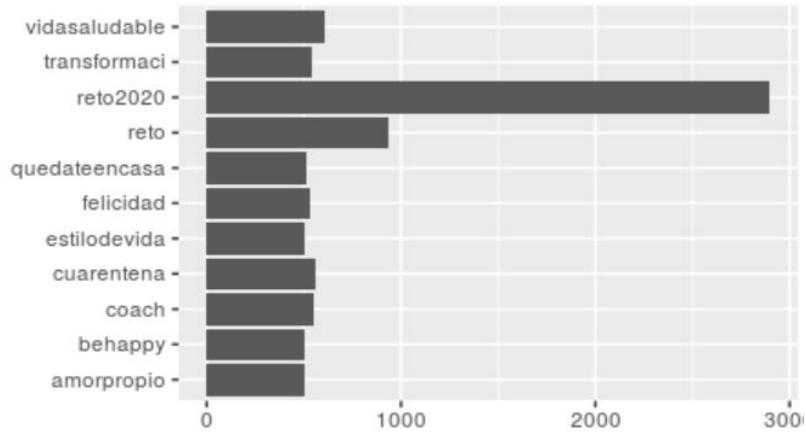
vinculados a este hashtag. Tales como: *quédate en casa, cuarentena, vida saludable, mama fit, coach, amor propio, bienestar y otros.*



- c) **Voyant Tools.** La utilización de esta herramienta se hizo en conjunto con el software R para obtener un análisis más completo de la información. Para ello, en R se separó por ngram y se eliminó stopwords.

El trabajo con R nos ayudó a identificar y confirmar las palabras clave y el uso de otros posibles hashtags ya detectados con la herramienta anterior: *vida saludable, cuarentena, coach, transformación y felicidad*. Y nos dió el detalle de los ngrams.

	palabra	n
1	reto2020	2898
2	reto	937
3	vidasaludable	607
4	cuarentena	559
5	coach	553
6	transformaci	537
7	felicidad	533



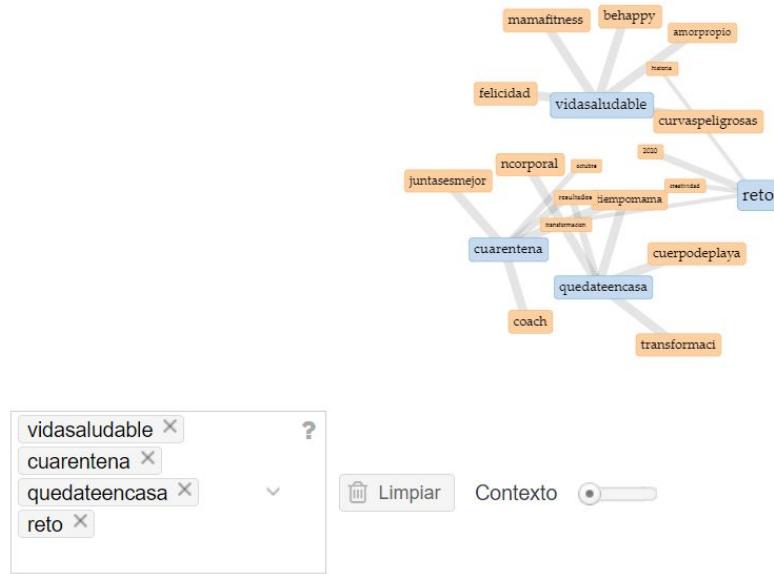
Con la data organizada en R, los resultados obtenidos en Voyant Tools fueron los siguientes:

- Corpus con 112,113 total de palabras y 14,331 palabras únicas.
- Densidad del vocabulario 0.128
- Promedio de palabras por oración: 112113.0
- Palabra más frecuente en el corpus: reto (937); vidasaludable (607); cuarentena (559); coach (553); transformación (537); felicidad (533); quedateencasa (515); amorpropio (506); estilodevida (502); behappy (501). (Estos datos coinciden con los identificados en R)
- Lista de 25 palabras comunes en el corpus:



- En relaciones se eligió los términos con mayor frecuencia (reto2020, cuarentena, quédate en casa y felicidad) para identificar los temas relacionados a estos. Se encontró:
 - Vida saludable está relacionado con felicidad, amor propio, mamafit, behappy.

- Cuarentena relacionado con coach, juntas es mejor, octubre, transformación
- Quédate en casa relacionado con resultados, tiempo de mamá, transformación, cuerpo de playa, corporal.
- Reto relacionado con 2020, transformación, creatividad, curvas peligrosas.

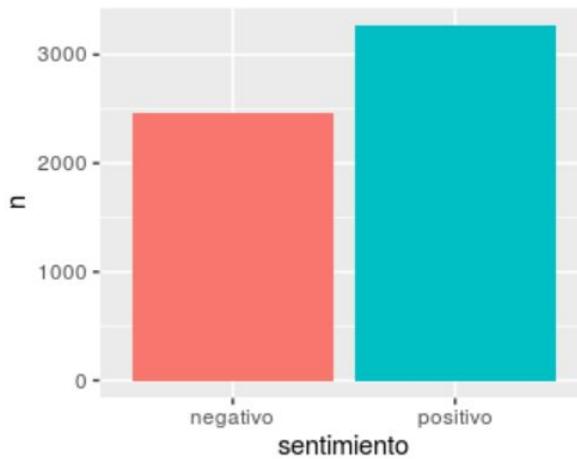


Enlace de exploración: <https://voyant-tools.org/?corpus=8760e06e62e50eb511a28e0a93cc8510>

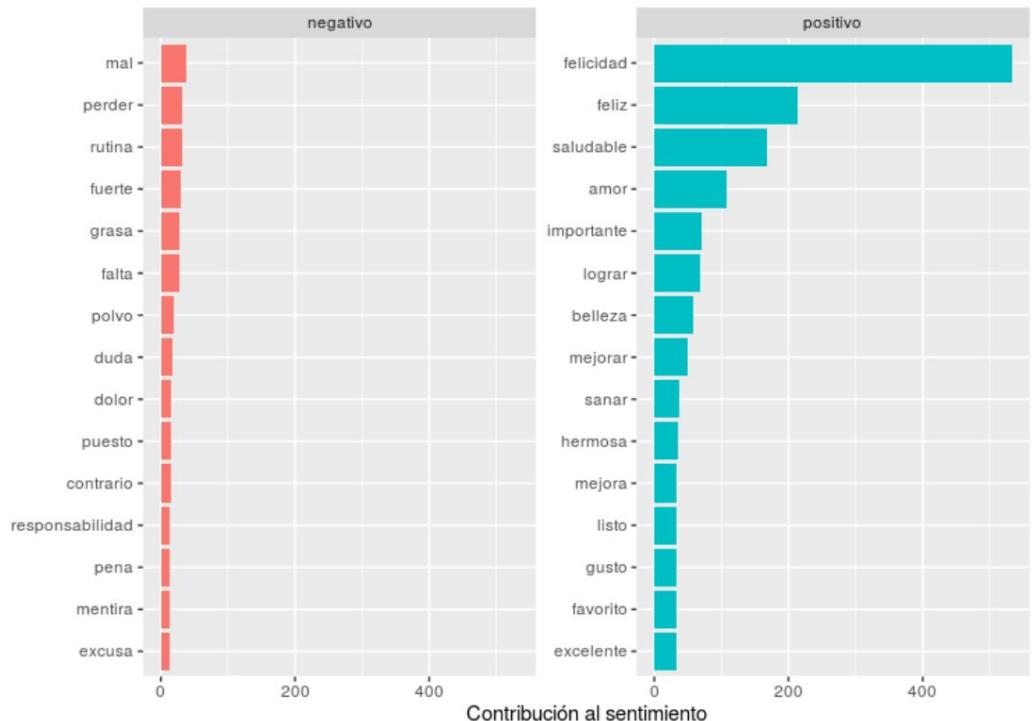
d) R - Análisis sentimental. Se identificaron las palabras del corpus que aportan a sentimientos positivos y negativos, para ello se separó por n-gram y se filtraron los stopwords.

- Primero se hizo un recuento rápido de sentimientos:

	sentimiento	n
1	positivo	3266
2	negativo	2462



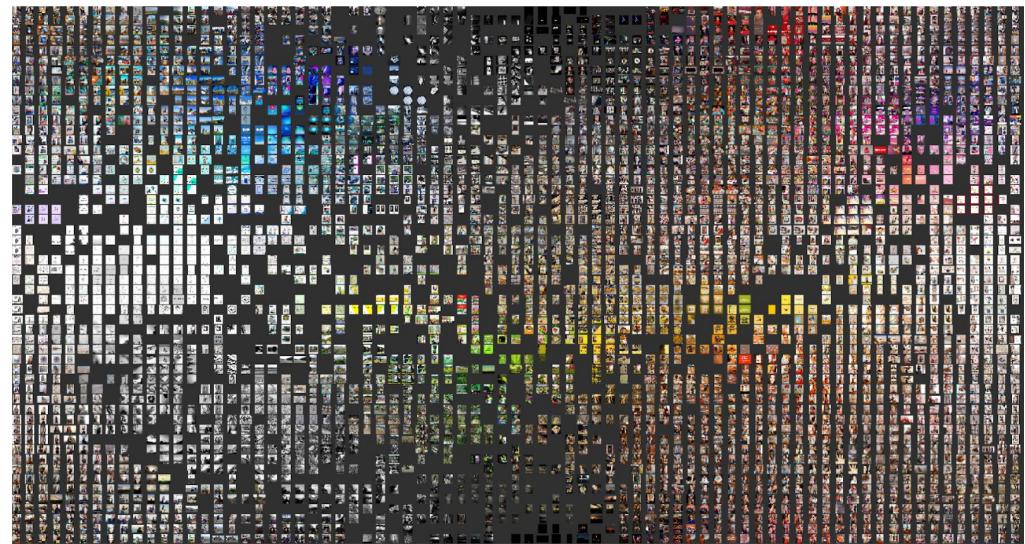
- Se revisó el detalle de las palabras que apoyaban esta valoración, tomando una lista de las palabras positivas y negativas más mencionadas en el corpus.



Nota: aquí se observó que las palabras “cuarentena” y “quédate en casa” destacadas en anteriores búsquedas no aparecen en esta lista de palabras de análisis sentimental, lo cual nos hace concluir que posiblemente aún no se encuentran clasificadas en el archivo de relación de sentimientos, ambas podrían ser consideradas en el grupo de negativos.

e) **Image Sorter.** La descarga de publicaciones de Instagram desde R integraba archivos de imágenes, esto se aprovechó para hacer una lectura distante de las imágenes que integraban las publicaciones, los resultados fueron estos:

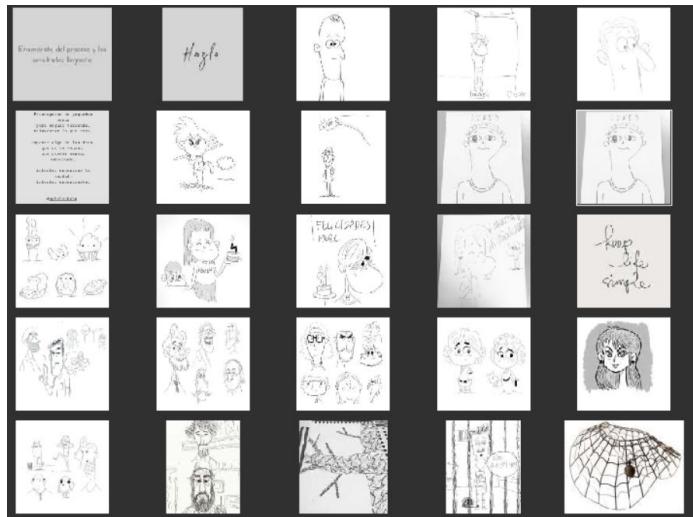
- Se organizaron por escala de color, lo cual agrupó las imágenes que compartían colores similares, se pudo destacar la predominancia de los azules, blancos, rojos y amarillos.



- Revisión de imágenes cercana relacionada al color amarillo, lo cual demuestra que la publicación estuvo relacionada con el # y el tema de belleza, ejercicios



- Imágenes con blancos tienden a ser dibujos mano alzada podrían estar relacionadas con el tema de creatividad.



- Escala de grises relacionado con fotos en blanco y negro, lo cual tendría relación con el #black&white.



4. Conclusiones

En la exploración del hashtag reto2020 durante el mes de noviembre, se aprecia que está ligado a publicaciones con temas de vida saludable, ejercicios, amor propio, belleza y creatividad, temas de connotación positiva.

Aunque la hipótesis inicial del proyecto era encontrar patrones en el texto descriptivo, la exploración en las herramientas dieron resultado relacionados a otros hashtag que suelen acompañar a reto2020, lo cual también nos ayudó a tener una idea mejor del tipo de publicaciones.

A pesar que la exploración de las publicaciones se hizo en noviembre, mes en el que el confinamiento había bajado, hubo un gran número de publicaciones que aún se enmarcaron en tópicos relacionados al covid-19, como “quédate en casa” y “cuarentena”.

Si bien el hashtag reto2020 connota un sentimiento de dificultad, es posible decir que las publicaciones en su mayoría abordaban un sentimiento de positivismo ya que hubo un porcentaje mayor de palabras que contribuyeron al análisis sentimental. Sin embargo, queda pendiente identificar si las palabras “cuarentena” y “quédate en casa” (las cuales tuvieron una frecuencia alta de uso) se enmarcan en sentimientos negativos.

Referencias

Burdick, A. (Ed.). (2012). *Digital humanities*. MIT Press.
Fradejas, J. (2020) *Cuentapalabras. Estilometría y análisis de texto con R para filólogos*. Universidad de Valladolid. Recuperado de:
<http://www.aic.uva.es/cuentapalabras/>

Google Trends (14 de noviembre de 2020). *Ret02020*. Recuperado de
<https://bit.ly/2ITx3WD>

Kemp, S. (21 de Julio de 2020). *Digital 2020 July Global Stashot*. www. datareportal.com. Recuperado de
<https://datareportal.com/reports/digital-2020-july-global-statshot>

Sued, G. E. (2020). Repertorio de técnicas digitales para la investigación con contenidos generados en redes sociodigitales. *Paakat: Revista de Tecnología y Sociedad*. 10(19). <http://dx.doi.org/10.32870/Pk.a10n19.498>

Silge, J & Robinson, D. (2020). *Text Mining with R: A Tidy Approach*. O'Reilly. Recuperado de <https://www.tidytextmining.com/index.html>

Voyant Tools (24 de noviembre de 2020). *Corpus Ret02020*. Recuperado de
<https://voyant-tools.org/?corpus=8760e06e62e50eb511a28e0a93cc8510>