

NEWS TIME SERIES

Palabras clave: Noticias, API, Serie de Tiempo

Esta charla relámpago presenta un nuevo paquete de R en funcionamiento, desarrollado para analizar la cobertura mediática de noticias desde 1851 hasta la fecha actual. Específicamente, este paquete se comunica con la API del NYT para optimizar la recolección para largos períodos de tiempo y es uno de los primeros desarrollos en diseñar un software optimizado para recolectar artículos de noticias y sus imágenes durante períodos de tiempo de más de 100 años.

El paquete de R interactúa con la API del NYT para recopilar una serie temporal de artículos, ofreciendo una forma de reunir y preprocesar grandes volúmenes de datos textuales. El paquete ya está escrito en GitHub (no está público aún, podría mostrarlo sin problemas sabiendo que será público antes de la presentación) y en funcionamiento.

El resultado de la función principal es un tibble que, además del texto del artículo dividido por secciones, también incluye sus respectivos metadatos y, lo más importante, las imágenes incluidas en el artículo. Aunque existe un paquete similar (Kearney, 2018), este ha estado sin actualizaciones durante mucho tiempo, la API del NYT ha cambiado, y, especialmente, no está optimizado para recolectar una serie temporal ni imágenes. News Time Series es un trabajo que ofrece datos de noticias que incluyen tanto el texto del artículo en lenguaje natural como sus imágenes, aportando nuevas oportunidades para analizar dos vías de comunicación al mismo tiempo con datos reales.

Esta charla presentará dos elementos principales: el funcionamiento del nuevo paquete y los resultados empíricos de una muestra de los datos que se pueden obtener con un enfoque especial en el estudio del sesgo mediático a largo plazo en América Latina.

La primera parte de la charla mostrará las funciones del paquete y los diferentes argumentos que se pueden utilizar. Trabajar con la API implica una tasa limitada de artículos por día y un número limitado de páginas de búsqueda; el paquete tiene en cuenta estas limitaciones para optimizar la recolección y, en unos pocos días, obtener todos los artículos sobre un tema desde 1851 hasta la fecha, que podrían ser más de 100.000 en muchos casos.

La segunda parte muestra resultados exploratorios utilizando los datos obtenidos. El análisis se realizó sobre más de 350,000 artículos sobre la cobertura del NYT de importantes eventos políticos en América Latina. Específicamente, se investiga si hubo algún sesgo en la cobertura de la inestabilidad política en Brasil, Argentina, Chile y Colombia desde 1851 hasta 2024.

NEWS TIME SERIES

Los resultados exploratorios indican una tendencia del NYT a dar prioridad particular a la cobertura de la inestabilidad de los regímenes políticos, a menudo a expensas de otros temas significativos. Este sesgo hacia noticias relacionadas con la inestabilidad resalta un posible sesgo en la cobertura de los medios internacionales, donde el conflicto y la crisis se enfatizan.

La charla ilustrará la usabilidad y el alcance del paquete de R a través de visualizaciones y resultados clave del análisis de más de 350,000 artículos. Al proporcionar tanto un enfoque metodológico como hallazgos empíricos, esta presentación tiene como objetivo demostrar el potencial del paquete como una herramienta poderosa para el análisis de series de tiempo de noticias en la investigación científica en ciencias sociales.