

Seguimiento diario de la producción científica sobre COVID-19

Juan Pablo Sokil

OCTS-OEI; Universidad de Buenos Aires

Palabras Claves: Bibliometría, Shiny, Visualización de Datos, COVID-19

Introducción

En diciembre de 2019 se registró en Wuhan, China, el primer paciente afectado por una neumonía atípica, el 7 de enero de 2020 las autoridades chinas confirmaron que correspondía a un nuevo virus, el 2019-ncov. Tan solo diez días después, apareció la primera publicación científica sobre el tema en la base de datos bibliométrica PubMed¹.

Si bien al principio, las publicaciones se concentraron en unos pocos países, al hacerse más claro el peligro de un brote global, la comunidad científica mundial comenzó a trabajar en el tema. El 30 de enero se definió al 2019-ncov como un problema de salud pública, en ese momento ya existían 29 publicaciones firmadas por 20 países. El 11 de marzo se lo clasificó como pandemia, para ese entonces ya había 777 publicaciones sobre el tema, al 30 de agosto existen más de 45 mil publicaciones y 188 países trabajando sobre el tema.

Objetivos

La velocidad con que la temática se expande dentro del campo de la ciencia requiere de un análisis dinámico, por eso, se diseñará un tablero que permita realizar un seguimiento diario de la producción científica sobre covid-19: Conocer su magnitud, evolución, países involucrados, colaboración conjunta y las líneas de investigación.

Materiales y Métodos

Se utilizaron las publicaciones científicas sobre Covid-19 extraídas de la base de datos PubMed². La descarga y procesamiento se realizó a través de la librería RisMed de R. Se desarrolló un código que recopila los datos de forma diaria y los incluye en una base de datos acumulada, de forma automática.

El campo que presentó mayor dificultad para procesar fue el país: este dato se extrae de la afiliación institucional de cada autor; usualmente el texto después de la última coma de cada firma es el país, para los casos donde esto no funcionó se identificaron una serie de expresiones regulares para extraerlo correctamente. Semanalmente se revisa el proceso y se incluyen nuevas expresiones regulares, de ser necesarias.

¹ <https://pubmed.ncbi.nlm.nih.gov/>

² Estrategia de Búsqueda: "COVID-19"[All Fields] OR "severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "2019-nCoV"[All Fields] OR "SARS-CoV-2"[All Fields] OR "2019nCoV"[All Fields] OR ("Wuhan"[All Fields] AND ("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields])) AND 2019/12[PDAT] : 2030[PDAT])

Resultados

El tablero diseñado con Shiny consta de 5 solapas, cada una, además de las visualizaciones correspondientes, brinda la posibilidad de descargar los datos para realizar análisis propios. La primera solapa muestra totales diarios y acumulados, la segunda solapa permite comparar en forma dinámica la evolución de los distintos países (Figura 1)

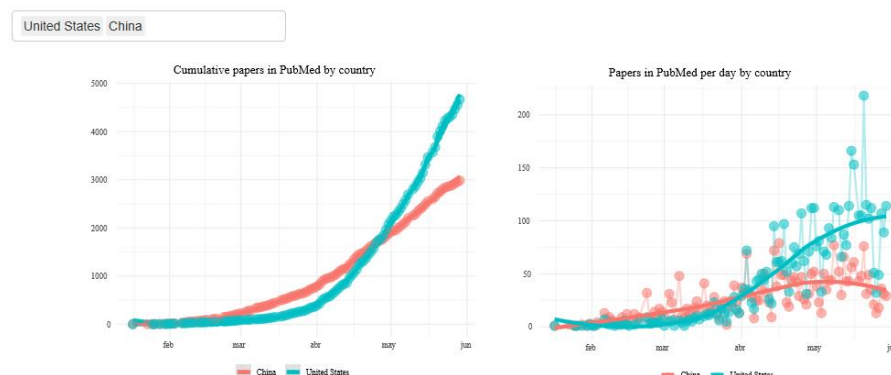


Figura 1: Producción científica sobre COVID-19 por país

La tercera solapa presenta un mapa que permite ver las redes de colaboración³ entre los países (Figura 2), la cuarta solapa presenta un mapa conceptual sobre el contenido de las publicaciones realizado a través de los conceptos más relevantes del resumen de las publicaciones, la quinta solapa es un navegador que permite filtrar las publicaciones por país, acceder a cada una o descargarlas.



Figura 2: Redes de Colaboración Internacional sobre COVID-19

El tablero se encuentra disponible en la siguiente dirección: <https://observatorio-cts.shinyapps.io/Covid-19/>. Las bases y el código utilizado se encuentran disponibles en el siguiente repositorio: <https://github.com/juansokil/Covid-19>

³ Se entiende por colaboración al trabajo conjunto de dos (o más autores) de distintos países (definido a partir de su afiliación institucional)