

Optimización de la Revisión de literatura académica mediante Textmining en R: Un enfoque automatizado

Henry Osorto

Facultad de Postgrado, Universidad Tecnológica Centroamericana, UNITEC, Tegucigalpa, Honduras

INTRODUCCIÓN

La revisión de literatura es muy importante en el desarrollo de investigaciones, que debido a la enorme rapidez con la que está creciendo la publicación digital de información académica, se ha vuelto complicada de realizar sin evitar caer en sesgos. Al llevar a cabo este proceso es común encontrarse con una abrumadora cantidad de artículos relacionados al tema que

se está investigando. A estos sistemas que contienen información de millones de autores, artículos, citas, figuras, tablas, así como redes académicas y bibliotecas digitales se les denomina Big Scholarly Data. El objetivo de este estudio fue desarrollar y aplicar un enfoque automatizado de revisión de literatura académica utilizando técnicas de Textmining en el lenguaje de programación R.

METODOLOGÍA/RESULTADOS

Se recuperó una base de datos de 86,820 artículos publicados en revistas científicas que contenían los términos de búsqueda (“modelo” AND “crecimiento” AND “económico”) alojados en la base de datos de acceso abierto Redalyc. Para acceder a la base de datos se recurrió a las herramientas de desarrollo del navegador web y consultar la API.

En el lenguaje de programación R se desarrollaron los Scripts que permitieron la optimización de la recuperación, procesamiento/limpieza y análisis del texto (palabras clave, resumen) mediante Textmining. Con ello se diseñaron análisis de frecuencias de palabras, relación de palabras en el tiempo y origen de la producción, y ngramas.

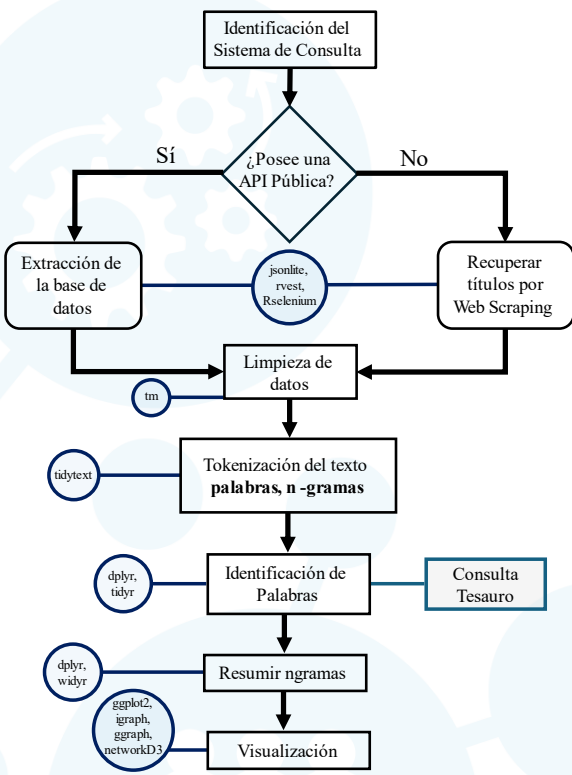


Figura 1. Modelo de optimización de RL en R

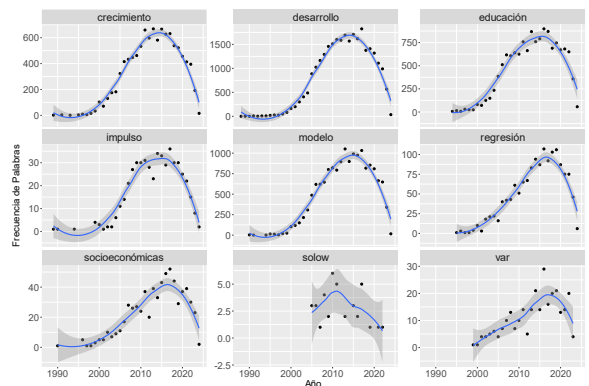


Figura 2. Análisis del texto por año de publicación

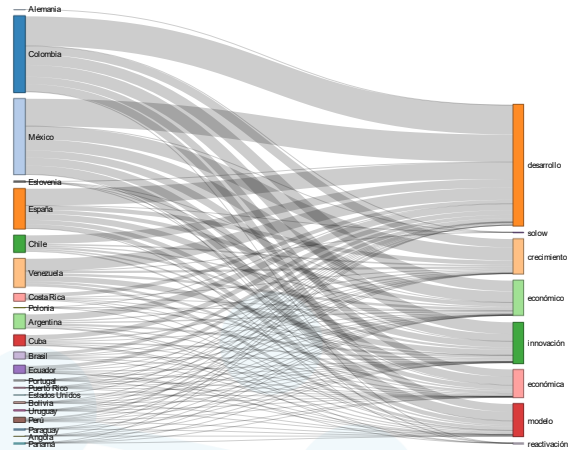


Figura 3. Análisis del texto por origen de publicación

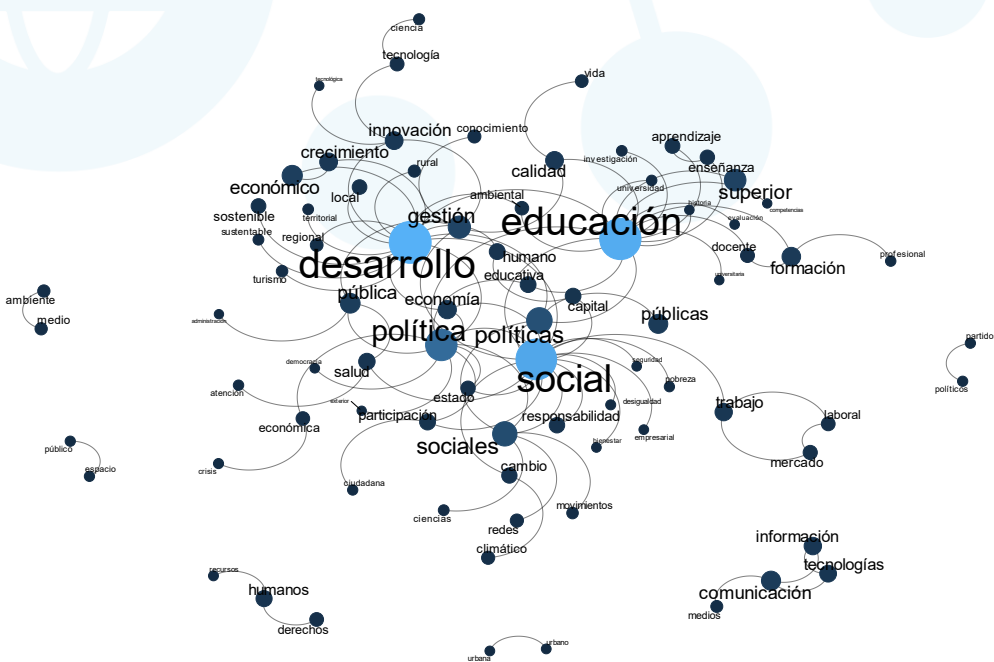


Figura 4. Análisis de texto mediante diagrama de red de palabras clave

CONCLUSIONES/RECOMENDACIONES

Este estudio demuestra el potencial del uso de herramientas de textmining y análisis bibliométrico automatizado, utilizando el lenguaje de programación R, para optimizar el proceso de revisión de literatura en el ámbito científico. Parte del aporte metodológico que

brinda esta investigación, consistió en la elaboración de Scripts de R que pueden ser consultados en el estudio completo para replicar o potenciar los análisis en otras revisiones de literatura. Se insta a la comunidad científica y académica a poder hacer uso de este enfoque de mediante el modelo propuesto.

Contacto: henry.osorto@unitec.edu.hn

Conflicto de interés: ninguno