Cómo aplicar la cienciometría a la investigación ecológica

Y. Añino1, J. Monge2, D. Murillo3, L. Michán-Aguirre4

1. Museo de Invertebrados G. B. Fairchild, Universidad de Panamá, Panamá.
2. Laboratorio Ecología Urbana, UNED, 2050 San José, Costa Rica.
3. Universidad Tecnológica de Panamá, Panamá.
4. Laboratorio de Bioinformación, Universidad Autónoma de México, México.

Autor para correspondencia: J. Monge [[julianmonge@gmail.com](mailto:julianmonge@gmail.com)]

Con más de 300 000 revistas científicas publicando millones de artículos cada año (Ulrich 2021), es imposible medir manualmente la producción científica. Así, han surgido la webmetría, que mide las vistas en internet, y la alt-metría, que mide menciones en redes sociales, *Wikipedia*, blogs y otros. Consideramos que estas técnicas, junto a la bibliometría, son parte de lo que hoy conocemos como cienciometría o guardan una estrecha relación con ella. La bibliometría es una rama multidisciplinaria que analiza la producción científica por sus publicaciones, en función de su estructura, desarrollo, dinámica, tendencias y relaciones (Michán y Muñoz-Velasco 2013).

En esta nota eco-informática explicamos cómo medir la producción ecológica usando estas técnicas, a la vez que advertimos sobre los errores comunes que hemos notado en la literatura reciente. Presentamos, al final, un ejemplo concreto sobre ecología de organismos panameños.

# Qué podemos aprender de la ecología con estas técnicas

Con técnicas cienciométricas podemos estudiar proyectos y productos —como patentes — para conocer la interacción entre científicos e instituciones y disciplinas, así como las líneas de investigación con mayor repercusión en un momento dado. Con las técnicas bibliométricas podemos identificar y medir las características de los artículos y revistas con indicadores cuantitativos de la producción especifica de una ciencia como las ciencias ecológicas (Michán 2011).

# Etapa 1. Cómo elegir el tema

Las opciones más comunes en este tipo de análisis son ecosistemas (ej. producción científica sobre ecología de arrecifes coralinos), organismos (ej. qué se publica sobre el lince ibérico) y países (ej. la producción en ecología de Brasil). Pero también puede hacerse cienciometría sobre otros temas recuperables en las bases de datos, como la producción de una persona particular (ej. la obra de Ramón Margalef) o instituto en particular (ej. producción científica del Instituto de Ecología de la UNAM).

## Etapa 2. Cómo recopilar los datos

## Es muy numerosa la literatura ecológica: por ejemplo, buscar ecología y *ecology* en *Google Scholar* muestra más de 5 000 000 de resultados. Los datos en masa tienen estructuras y formatos de diferentes fuentes y se llaman en inglés *Big Data*, es decir datos que, por su gran volumen, complejidad y velocidad de crecimiento, dificultan recopilarlos, administrarlos y analizarlos. Esta información puede ser filtrada a través de una técnica llamada *minería de datos*, la cual busca identificar de manera automática patrones o correlación de un conjunto de datos. Si la fuente de datos consiste, por ejemplo, en páginas web, redes sociales, artículos de revista y correos electrónicos, se habla de *minería de texto* (Jiawei et al. 2012), específicamente utilizando la técnica llamada en inglés *Web Scraping,* la cual tiene dos pasos, en primer lugar, “raspar” el código HTML, o sea, usar un algoritmo para extraer los datos de interés y, en segundo lugar, limpiarlos, procesarlos y analizarlos (Ferrara et al. 2014).

## Existen varias formas de *Web Scraping*, Murillo et al. (2018) comparan algunas formas para la extracción de perfiles de investigadores en *Google Scholar*, incluyendo algoritmos en *R* y *Rstudio* con paquetes como *Scholar (*Keirstead 2015). De la misma manera, podemos utilizar programas como *Publish or Perish*, el cual nos permite realizar búsquedas en bases de datos como *Crossref*, *Google Scholar*, *PubMed*, *Microsoft Academic*, *Scopus* y *Web of Science*. En estas y otras bases de datos se puede descargar la información directamente.

## Etapa 3. Como “limpiar” los datos

Este tipo de estudios (cienciometría) suele sufrir de datos “sucios”, que se deben a resultados espurios, por ejemplo, una búsqueda sobre comportamiento territorial en mariposas, basado en *butterfly+territory*, no solo dará las respuestas que buscamos, sino otras como *The butterflies of the Territory of Boznia.* Además de eliminar esos casos espurios, hay que unificar los registros de personas e instituciones que aparecen con diferentes nombres, por ejemplo, una misma institución puede aparecer en los artículos como INIECO, Inst. Inv. Ecól, o Instituto de Investigaciones Ecológicas de Málaga; y para la computadora, cada uno es un instituto diferente, lo que falsea las estadísticas. Por el contrario, hay que poder distinguir entre todas las personas que comparten un nombre, por ejemplo, es muy probable que estén activos varios ecólogos llamados Juan Pérez y antes de existir el número de identidad ORCID era necesario distinguirlos manualmente.

## Etapa 4. Las dos maneras de analizar datos

Los datos recopilados por las diversas bases de datos pueden ser desde decenas hasta miles de artículos, y pueden analizarse de dos formas. La primera, es puramente descriptiva y la segunda, llamada inferencial, por prueba de hipótesis. La descriptiva, que por el momento es la gran mayoría de lo que se publica, es la más débil, y se limita a indicar frecuencias y porcentajes, indicando, por ejemplo, que la Universidad Central publica el 36% de los artículos ecológicos de Venezuela. En la manera inferencial, se proponen hipótesis y se aplican pruebas estadísticas para evaluarlas, por ejemplo, usando una U de Mann-Whitney para evaluar la hipótesis de que ciertos mamíferos europeos son más estudiados que sus parientes africanos.

### Etapa 5. Cuidados a la hora de redactar

Es fácil que, cuando se llega a esta etapa, se comience a hablar de factor de impacto, temas más estudiados, investigadores más prolíficos y revistas más citadas. Pero nada de esto puede extraerse de estos estudios, porque las bases disponibles actualmente, sean la extremadamente incompleta *Web of Science*, la casi igualmente insuficiente *Scopus*, o incluso *Google Scholar*, dejan por fuera la gran mayoría de las publicaciones científicas. Por ejemplo, *Web of Science* y *Scopus*, únicamente cubren poco más de 10 000 revistas de las 300 000 que se publican, o sea, dejan por fuera el conteo del 97 % de las publicaciones *y de las citas* al medir el supuesto “factor de impacto” (Monge-Nájera 2014; Millán et al. 2017; Goñi 2021). Por lo tanto, al redactar, debemos aclarar en todo momento que lo que encontramos solo es válido para el 3 % de las publicaciones.

### Etapa 6. Un consejo final

Cuando nuestro análisis sea sobre temas en que no tenemos experiencia directa, conviene agregar al equipo de investigación una persona con amplia experiencia en el tema ecológico estudiado por nuestros análisis bibliométricos y cienciométricos. Alguien que conozca a las personas e instituciones que estudian el campo, para que pueda interpretar mejor los patrones que encontremos. Por ejemplo, es fácil hallar un aumento reciente en el número de publicaciones y pensar que se debe a alguna política institucional o estatal de apoyo a la investigación ecológica, cuando en realidad se trata de un patrón mundial que se ha documentado en todos los países, independientemente de cuánto apoyen sus gobiernos la investigación ecológica.

### Ejemplo concreto de aplicación

Hemos incluido la figura 1 para mostrar el ejemplo de un análisis bibliométrico de estudios sobre abejas de las orquídeas en Panamá, basado en *PubMed*, que cubre una fracción de la literatura del *Web of Science*, la cual a su vez deja fuera el 97 % de las publicaciones científicas (Fig. 1a). Como es de esperar, predominan los artículos enfocados a ciencias médicas (Fig. 1b), sería mejor incluir también repositorios institucionales, portales de revistas locales, motores de búsqueda (*Google Scholar, Dimensions, Microsoft Academic*, etc.) y técnicas como *web scraping, big data* y minería de datos. Los metadatos permiten identificar palabras frecuentes (Fig. 1b), y temas asociados (Fig. 1c). También se pueden identificar temas más estudiados y patrones en el tiempo (Fig. 1d).

# Referencias

Goñi, F. 2021. La cienciometría, una calamidad histórica para la ciencia. Sociedad Española de Bioquímica y Biología Molecular. (Consultado: 15-2-2021): <https://www.sebbm.es/revista/articulo.php?id=419&url=la-cienciometria-una-calamidad-historica-para-la-ciencia>

Ferrara, E., De Meo, P., Fiumara, G., Baumgartner, R. 2014. Web data extraction, applications and techniques: A survey. *Knowledge-Based Systems*, 70, 301–323. <https://doi.org/10.1016/j.knosys.2014.07.007>

Jiawei, H., Kamber, M., Han, J., Kamber, M., Pei, J. 2012. Data Mining: Concepts and Techniques. In San Francisco, CA: Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>

Keirstead, J. 2015. Package Scholar. Disponible: <https://cran.rproject.org/web/packages/scholar/index.html>

Michán, L. 2011. Cienciometría, información e informática en ciencias biológicas: enfoque interdisciplinario para estudiar interdisciplinas. *Ludus Vitalis: Revista de Filosofía de Las Ciencias de La Vida,* 19(35), 239–243.

Michán, L., Muñoz-Velasco, I. 2013. Cienciometría para ciencias médicas: definiciones, aplicaciones y perspectivas. *Investigación en Educación Médica*, 2(6), 100–106. <https://doi.org/10.1016/s2007-5057(13)72694-2>

Millán, J.D., Polanco, F., Ossa, J.C., Béria, J.S., Cudina, J.N. 2017. La cienciometría, su método y su filosofía: Reflexiones epistémicas de sus alcances en el siglo XXI. *Revista Guillermo De Ockham*, 15(2), 17–27. <https://doi.org/10.21500/22563202.3492>

Monge-Nájera, J. 2014. La invalidez del Factor de Impacto como indicador del impacto de las revistas científicas latinoamericanas. *Revista de Biología Tropical*, 62(1), 407-412.

Murillo, D., Saavedra, D., Quintero, E. 2018. Extracción de datos de perfiles en Google Scholar utilizando un algoritmo en el lenguaje R para hacer minería de datos. I+D Tecnológico, 14(1), 94-104. <https://doi.org/10.33412/idt.v14.1.1807>

Ulrich. 2021. (<https://www.ulrichsweb.com/ulrichsweb/faqs.asp> )

###### 

**Figura 1.** Ejemplo de la aplicación de la bibliometría a un tema ecológico. (a) extracción de información en una base de datos con artículos en *Web of Science*; (b) representación de la conexión de palabras clave de artículos panameños en distintas diciplinas; (c) gráfico de red de las palabras claves de estudios con abejas de las orquídeas en Panamá, mediante el uso de VOSviewer; (d) análisis bibliométricos exploratorios derivados de la compilación de información.