

# Cómo aplicar la ciencia métrica a la investigación ecológica

Yostin J. Añino<sup>1,\*</sup> , Julián Monge-Nájera<sup>2</sup> , Danny Murillo-González<sup>3</sup> , Layla Michán-Aguirre<sup>4</sup>

(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 de correspondencia: Julián Monge [[julianmonge@gmail.com](mailto:julianmonge@gmail.com)]

> Recibido el 25 de mayo de 2021 - Aceptado el 21 de junio de 2021

Cómo citar: Añino, Y. J., Monge-Nájera, J., Murillo, D., Michán, L. Cómo aplicar la ciencia métrica a la investigación ecológica. *Ecosistemas* 30(2): 2256. <https://doi.org/10.7818/ECOS.2256>

¿Cuántos artículos científicos se publican cada año? ¿Cuántas revistas científicas hay en el mundo? Sorprendentemente, nadie sabe la respuesta a estas preguntas tan básicas; la respuesta, por otra parte, cambia constantemente con la aparición de nuevos artículos y con el frecuente cierre y apertura de revistas (Laaks et al. 2021). Estas preguntas, aplicables también a ramas concretas como la ciencia ecológica, son propias de la *ciencia métrica* y la *bibliometría*. En breve, la ciencia métrica es el estudio cuantitativo de toda la producción científica, y una de sus ramas, la *bibliometría*, mide específicamente las publicaciones. Con decenas de miles de revistas científicas publicando cada una entre decenas y centenares de artículos cada año, es claro que la producción mundial es inmensa y que es imposible medirla manualmente. Así, han surgido en años recientes métodos automatizados, como la *webmetría*, que mide las descargas de artículos en internet, y la *alt-metría*, que mide menciones de los artículos científicos en redes sociales, *Wikipedia*, blogs y otros.

El tema que consideraremos en esta nota eco-informática, la bibliometría, se define como aquella rama multidisciplinaria de la ciencia métrica que analiza la producción científica por sus publicaciones, en función de su estructura, desarrollo, dinámica, tendencias y relaciones (Michán y Muñoz-Velasco 2013).

Acá explicamos cómo medir la producción de literatura científica en ecología, usando técnicas bibliométricas, a la vez que advertimos sobre los errores que hemos notado en la literatura reciente. Presentamos, al final, un ejemplo concreto usando la ecología de organismos panameños, aunque la parte central de esta nota usa fundamentalmente ejemplos españoles.

## Qué podemos aprender sobre la ecología con estas técnicas

Con técnicas científicas 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 específica de una ciencia como las ciencias ecológicas (Michán 2011).

## Etapas 1. Elección de temas ecológicos de interés bibliométrico

Las opciones más comunes en este tipo de análisis son ecosistemas (p.ej. producción científica sobre ecología de arrecifes coralinos), organismos (p.ej. qué se publica sobre el lince ibérico) y países (p.ej. la producción en ecología de Brasil). Pero también puede hacerse ciencia métrica sobre otros temas recuperables en las bases de datos, como la producción de una persona (p.ej. la obra de Ramón Margalef) o instituto en particular (p.ej. producción científica del Instituto de Ecología de la UNAM). Nuestra recomendación es elegir un tema conocido, porque así se podrá hacer una interpretación mejor fundamentada de los datos.

## Etapas 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 (“Big Data”) tienen estructuras y formatos diferentes según sus fuentes; por su gran volumen, complejidad y velocidad de crecimiento, resulta difícil 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*; por ejemplo, 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. Es importante men-

cionar que la implementación de técnicas de búsqueda y cosecha de información también pueden ser complementadas por las búsquedas tradicionales en internet (búsqueda generalizada) y búsquedas sistematizadas como es el caso del método PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) que recopila información de manera estandarizada (Moher et al. 2009). De hecho, la aplicación de todas las técnicas anteriormente mencionadas contribuye favorablemente al estudio bibliométrico (Fig. 1).

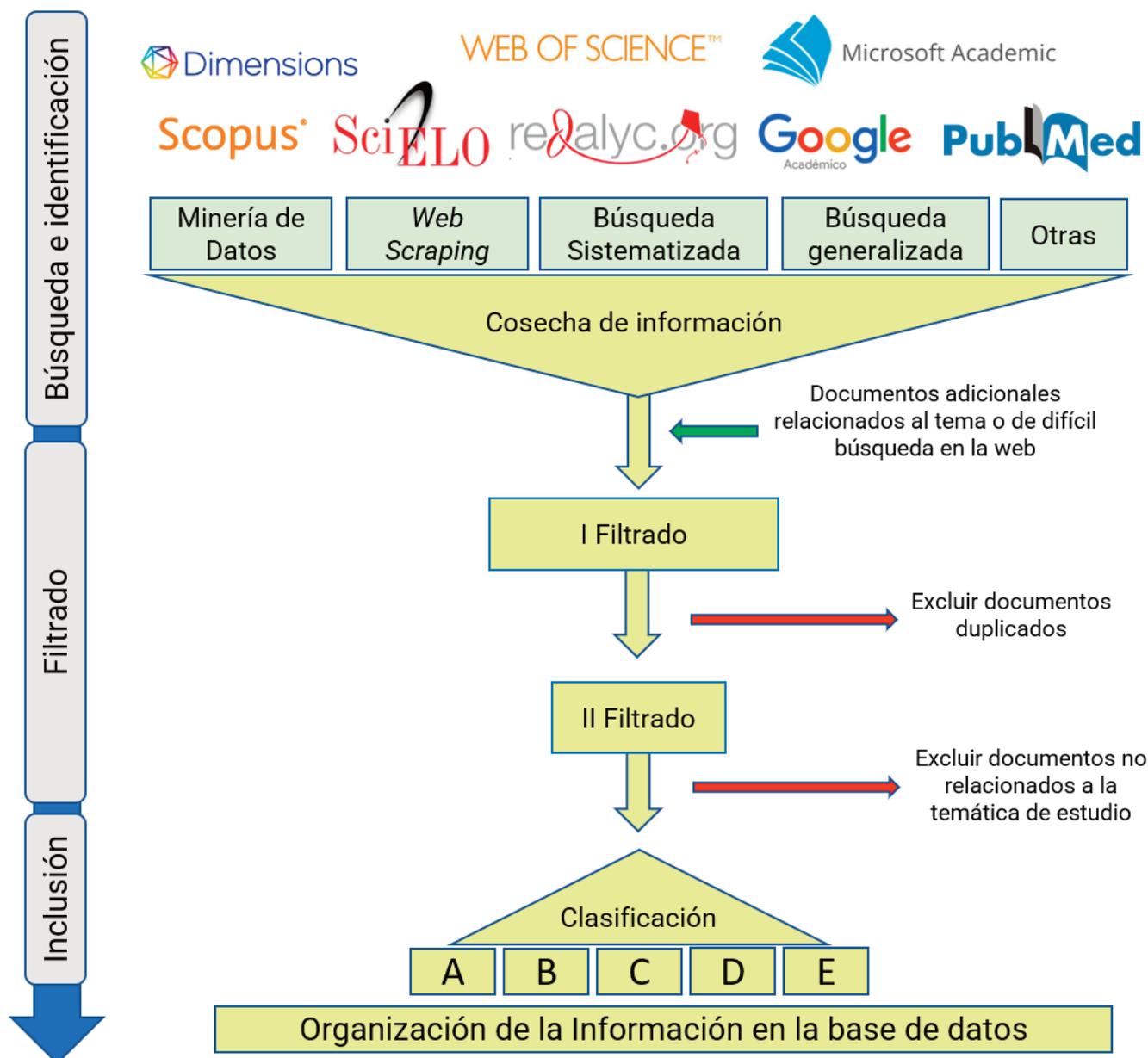
### Etapa 3. Cómo “limpiar” los datos

Los estudios de cienciometría suelen 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 sobre el comportamiento de estos insectos, sino otras que no son sobre el tema pero que incluyen las mismas palabras, como *The butterflies of the Territory of Bosnia*. 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”, pero también

como “Inst. Inv. Ecol.”, o incluso como “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 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, se centra en frecuencias y porcentajes, indicando, por ejemplo, que la Universidad Central publica el 36% de los artículos ecológicos de Venezuela (Rojas-Sola y Jordá-Albiñana 2010). En la manera inferencial, se proponen hipótesis y se aplican pruebas estadísticas para evaluarlas, por ejemplo, en un ejemplo imaginario, usar una U de Mann-Whitney para evaluar la hipótesis de que ciertos mamíferos europeos son más estudiados que sus parientes africanos.



**Figura 1.** Diagrama de flujo para la obtención de datos en un análisis bibliométrico en tres etapas. Tras revisar las fuentes principales conviene consultar fuentes secundarias, para luego descartar material inadecuado o repetido, y proceder al análisis de los datos.

Existen varios programas para análisis bibliométricos, por ejemplo: R (paquetes: *Bibliometrix* e *igraph*), Vosviewer (Van Eck y Waltman 2010), Sci2 (Sci2 Team 2009) y Gephi (Bastian et al. 2009).

## Etapa 5. Cuidados al interpretar los resultados

Es fácil que cuando se llega a esta etapa, se comience a hablar de temas frecuentes en este tipo de estudio: factor de impacto, temas más estudiados, investigadores más prolíficos y revistas más citadas. Pero debe tenerse un cuidado, porque las fuentes de uso común, *Web of Science*, *Scopus* y *Google Scholar*, están lejos de incorporar todas las revistas y todas las citas, particularmente por su poca cobertura de revistas de países pequeños y de artículos que no estén en inglés (Monge-Nájera 2014; Millán et al. 2017; Goñi 2021). Un ejemplo del efecto del idioma corresponde al vietnamita: según Hien (2010), el *Science Citation Index Expanded* deja por fuera el 95 % de los artículos científicos de Vietnam, por hallarse en vietnamita. Un ejemplo del problema a nivel de país es Cuba, donde solo un 22% de la producción científica aparece en revistas indexadas internacionalmente (Galbán-Rodríguez et al. 2019).

## Etapa 6. Un consejo final

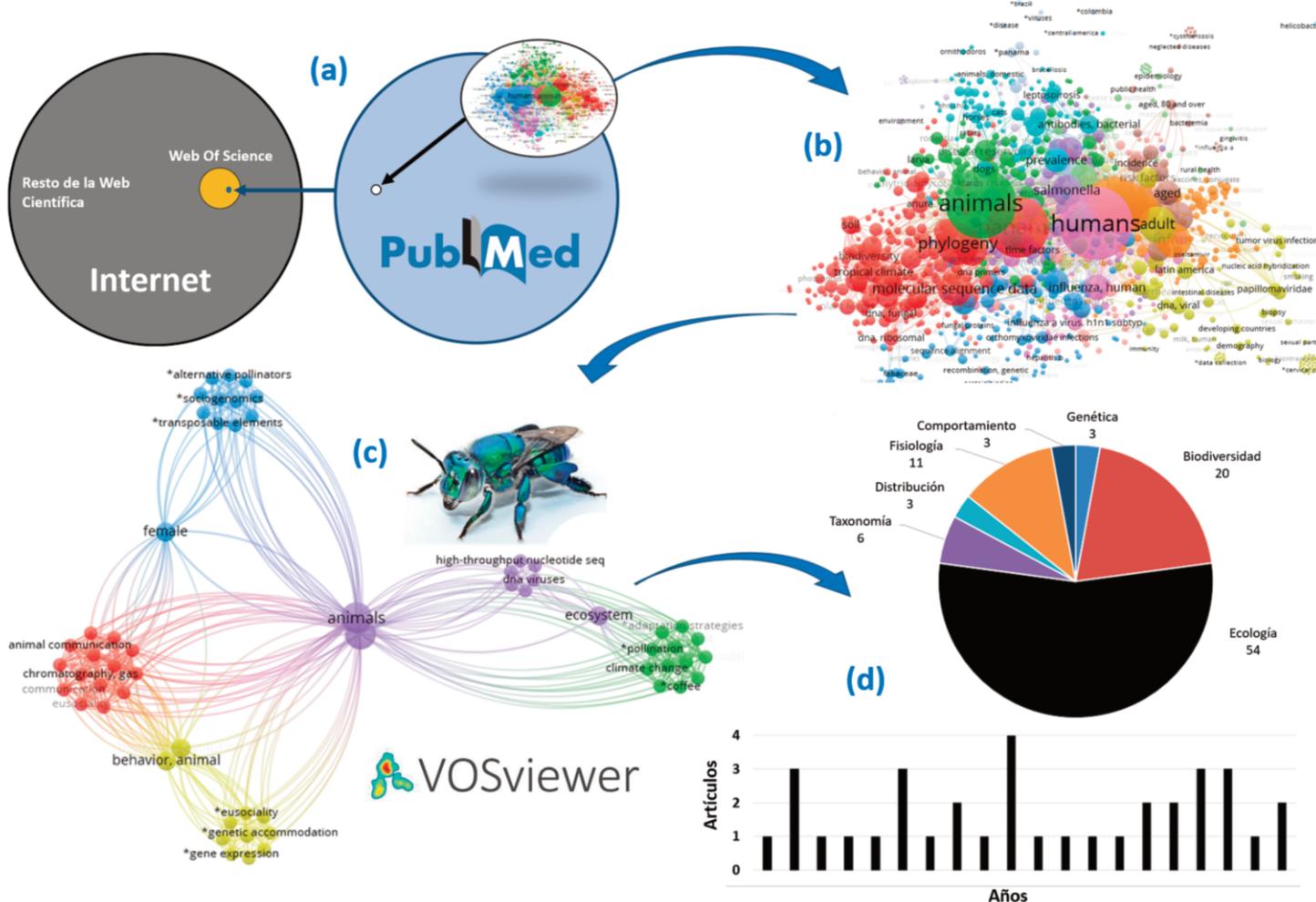
Cuando nuestro análisis sea sobre temas en que no tenemos experiencia directa, conviene agregar al equipo de investigación una persona con amplia experiencia en el tema ecológico estudiado por nuestros análisis bibliométricos y cienciométricos: alguien que conozca a las personas e instituciones que estudian 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 (Monge-Nájera et al. 2021).

## Ejemplo concreto de aplicación

Hemos incluido la **Figura 2** para mostrar el ejemplo de un análisis bibliométrico de estudios sobre abejas de las orquídeas en Panamá, basado en *PubMed*, índice que se concentra en los temas de salud, pero no se limita a ellos y cubre mejor que otros la región centroamericana (**Fig. 2a**). Los metadatos permiten identificar palabras frecuentes (**Fig. 2b**); temas asociados (**Fig. 2c**); temas más estudiados y patrones temporales (**Fig. 2d**). Como es de esperar, los resultados para abejas de las orquídeas no son muy amplios por ser una base de énfasis en salud, y habría que interpretar los resultados con esa limitación. Sería aún mejor incluir también bases más generales, como *Scopus*; repositorios institucionales; portales de revistas locales; motores de búsqueda (*Google Scholar*, *Dimensions*, *Microsoft Academic*, etc.) y aprovechar técnicas nuevas como *web scraping*, *big data* y minería de datos, las cuales mencionamos al inicio de esta nota.

En conclusión, con los cuidados del caso, podemos hacer una gran cantidad de estudios bibliométricos sobre investigación científica en el área de la ecología, tema oportuno como nunca en esta época de grandes cambios ambientales.



**Figura 2.** 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 disciplinas; (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. Nota: el uso de *PubMed* en este ejemplo se debió a la fácil colecta de datos y a que es una base de acceso gratuito.

## Agradecimientos

Agradecemos a Francisco Rodríguez, Hugo Saiz y David García por sus valiosas recomendaciones para mejorar esta nota.

## Referencias

- Bastian, M., Heymann, S., Jacomy, M. 2009. Gephi: an open-source software for exploring and manipulating networks. EN: International AAAI Conference on Weblogs and social media, San José, EE.UU.
- 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>
- Galbán-Rodríguez, E., Torres-Ponjuán, D., Martí-Lahera, Y., Arencibia-Jorge, R. 2019. Measuring the Cuban scientific output in scholarly journals through a comprehensive coverage approach. *Scientometrics* 121(2): 1019-1043.
- Goñi, F. 2021. La cienciometría, una calamidad histórica para la ciencia. Sociedad Española de Bioquímica y Biología Molecular. [usado el 15 de febrero de 2021]. Disponible en: <https://www.sebbm.es/revista/articulo.php?id=419&url=la-cienciometria-una-calamidad-historica-para-la-ciencia>
- Hien, P.D. 2010. A comparative study of research capabilities of East Asian countries and implications for Vietnam. *Higher Education* 60(6): 615-625.
- Jiawei, H., Kamber, M., Han, J., Kamber, M., Pei, J. 2012. Data Mining: Concepts and Techniques. Morgan Kaufmann, Wlatham, EE.UU.
- Keirstead, J. 2015. Package Scholar. Disponible en: <https://cran.rproject.org/web/packages/scholar/index.html>
- Laaks, M., Matthias, L., Najko, J. 2021. Open is not forever: A study of vanished open access journals. *Journal of the Association for Information Science and Technology* 2021:1-14.
- 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](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>
- Moher, D., Liberati, A., J., Tetzlaff, Altman, D., PRISMA Group. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine* 151(4): 264. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>
- 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.
- Monge-Nájera, J., Yi, S., Ho, Y.S. 2021. Bibliometrics of Cambodian Science: subjects, researchers and impact in the Science Citation Index Expanded. *Revista de Biología Tropical* 69(2): 678-687.
- 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>
- Rojas-Sola, J. I., Jordá-Albiñana, B. 2010. Análisis bibliométrico de las publicaciones científicas venezolanas en la categoría de ecología en la base de datos del Web of Science (1997-2008). *Interciencia* 35(8): 619-623.
- Sci2 Team 2009. Science of Science (Sci2) Tool. Indiana University and SciTech Strategies, Bloomington, EE.UU. Disponible en: <https://sci2.cns.iu.edu>
- Van Eck, N., Waltman, L. 2010. Software Survey: Vosviewer, A Computer Program for Bibliometric Mapping. *Scientometrics* 84(2): 523-538.