

# Google Académico en la visibilidad de revistas científicas

Adriana Peña Pérez Negrón<sup>1</sup> y Graciela Lara López<sup>2</sup>

## Resumen

La cantidad cada vez más grande de información disponible, ha hecho que encontrar literatura académica de manera organizada se convierta en un requerimiento creciente. En este contexto, los motores de búsqueda escolares en Internet se han convertido en una poderosa herramienta en el ámbito académico, en dónde uno de los más utilizados es Google Académico. Este motor de búsqueda especializado en literatura académica, presenta numerosas y documentadas deficiencias; sin embargo, sus poderosos algoritmos de clasificación lo han convertido en un importante medio de visibilidad, que las revistas no debemos pasar por alto.

## Motores de búsqueda en Internet

Internet, la red electrónica de comunicación que conecta redes y organiza computadoras alrededor del mundo, se ha convertido en uno de los principales medios de información. Debido a la gran cantidad que por este medio está disponible hoy en día, los usuarios dependen de los motores de búsqueda que, además de clasificarla, la presentan de manera ordenada. Los motores operan mediante algoritmos que se van modificando conforme el uso que le damos los humanos.

Los primeros motores de búsqueda en Internet datan de finales de los ochenta. En 1991 aparece una primera herramienta para buscar contenido en Internet, llamada Archie (por *archive*, archivo en inglés, sin la *v*). Archie fue creada por Alan Emtage, Bill Heelan y Peter Deutsch, estudiantes de computación de la Universidad McGill en Montreal, Canadá. Esta

---

<sup>1</sup> Adriana Peña Pérez Negrón es profesora investigadora del Departamento de Ciencias Computacionales de la División de Electrónica y Computación del CUCEI de la Universidad de Guadalajara, y miembro del Sistema Nacional de Investigadores Nivel I. Es Editora General de la revista ReCIBE, publicada por la División de Electrónica y Computación de la Universidad de Guadalajara. Sus principales líneas de investigación son la realidad virtual multiusuario y la interacción mediada por computadora. CE: [adriana.ppnegron@academicos.udg.mx](mailto:adriana.ppnegron@academicos.udg.mx).

<sup>2</sup> Graciela Lara López es profesora investigadora del Departamento de Ciencias Computacionales de la División de Electrónica y Computación del CUCEI de la Universidad de Guadalajara, y miembro del Sistema Nacional de Investigadores Nivel I. Es Editora de la Sección de Informática y Computación de la revista ReCIBE, publicada por la División de Electrónica y Computación de la Universidad de Guadalajara. Su principal línea de investigación es la realidad virtual. CE: [graciela.lara@academicos.udg.mx](mailto:graciela.lara@academicos.udg.mx).

herramienta, si bien generaba una lista de archivos FTP (archivos con protocolo cliente-servidor), no los indexaba. Este último proceso no era necesario, debido a la pequeña cantidad de archivos de que se trataba (Kuyoro *et al.* 2012, 942). Archie se encuentra todavía en funcionamiento y su última modificación fue hecha en 2017.<sup>3</sup>

Dos años después, el 2 de septiembre de 1993, Oscar Nierstrasz de la Universidad de Génova, lanzó el que es reconocido como el primer motor de búsqueda de Internet. Y ese mismo año, Matthew Gray de la MIT produjo el primer robot web utilizado para generar un índice, cuyo propósito era medir el tamaño de la WWW (World Wide Web). Sin embargo, esto no pudo concretarse sino hasta 1995 (Kuyoro *et al.* 2012, 943).

En enero de 1996 Google inició como un proyecto de investigación de Larry Page y Sergey Brin, estudiantes de la Universidad Stanford de California, en Estados Unidos. Y en 1998 Google, basada en una pequeña compañía de motor de búsqueda llamada goto.com, consideró la idea de vender la búsqueda de términos. Este hecho impactó en un negocio que se convirtió en uno de los más lucrativos de Internet.

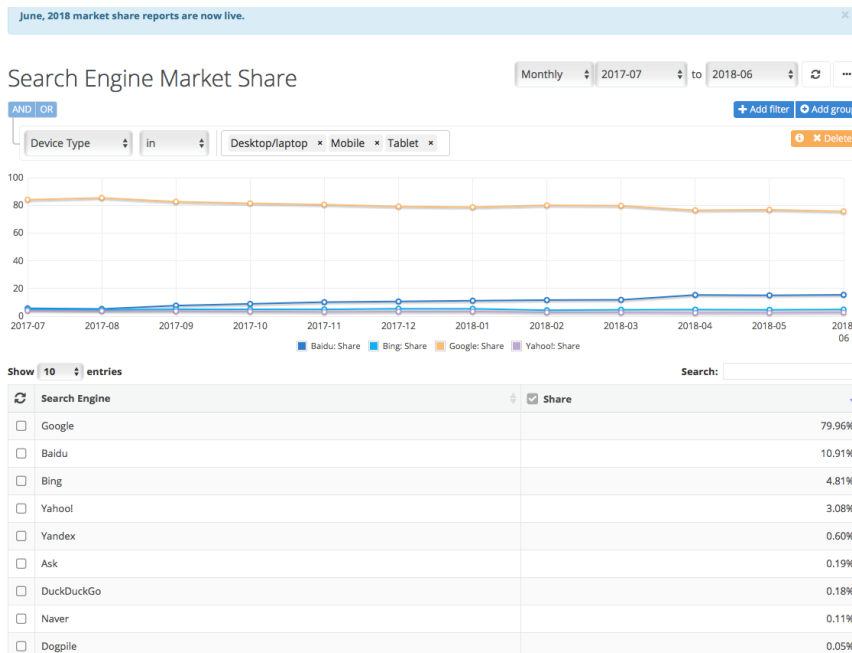
Alrededor del año 2000, el motor de búsqueda de Google se convirtió uno de los más prominentes, y consiguió muy buenos resultados mediante su algoritmo PageRank. Este algoritmo crea un *ranking* (o clasificación de páginas) basado en el número de *links* o ligas que redirigen al usuario a esas páginas (Kuyoro *et al.* 2012, 943). Al día de hoy, Google es el motor de búsqueda más popular, con un 79.96 por ciento del mercado total, que incluye computadoras, tabletas y móviles —dato tomado de NetMarketshare, véase Figura 1 (NetMarketshare, 2018). Una línea de tiempo completa, con la historia de los motores de búsqueda, puede consultarse en la página de la empresa WordStream.<sup>4</sup> Entre las aplicaciones con las que cuenta Google, está la especializada en literatura académica, Google Scholar, que en español se traduce como Google Académico.

**Figura 1. Consulta en Netmarketshare del motor más utilizado en el mercado**

---

<sup>3</sup> [http://archie.icm.edu.pl/archie\\_eng.html](http://archie.icm.edu.pl/archie_eng.html)

<sup>4</sup> <https://www.wordstream.com/articles/internet-search-engines-history>



## Antecedentes de Google Académico

En 2004, Google lanzó en formato beta su motor de búsqueda web, Google Académico. De acuerdo con información de su página web,<sup>5</sup> este motor indexa texto completo y metadatos de literatura escolar ,entre los que se incluyen revistas científicas, libros, resúmenes y dictámenes de editores académicos, sociedades profesionales, repositorios en línea, universidades, entre otros sitios web. Su lanzamiento tuvo el objetivo encontrar trabajos relevantes en relación a la academia escolarizada. Sus principales funciones son:

- Buscar literatura escolarizada en un solo lugar
- Explorar trabajos relacionados, citas, autores y publicaciones
- Localizar de documentos completos a través de bibliotecas o en la web
- Mantenerse al día con los más recientes desarrollos en cualquier área de investigación
- Verificar quién cita tus publicaciones, crear un perfil público de autor

De acuerdo con Google Académico, el *ranking* de sus documentos se lleva a cabo de la misma forma que lo hacen los científicos, considerando el documento completo, dónde fue

<sup>5</sup> <https://scholar.google.com.mx>

publicado, quién lo escribió, así como qué tan frecuentemente y qué tan recientemente ha sido citado en algún otro documento similar.

Es importante hacer notar que aunque las personas en general no vayan directamente a Google Académico, éste puede dirigirlos. En cambio, en el ámbito de la academia sí existe una fuerte tendencia hacia su uso.

Desde su lanzamiento y hasta la fecha, Google Académico ha sido reconocido y utilizado en la comunidad escolar, particularmente en licenciatura (Fagan 2017, 33). Entre sus herramientas están los *ranking* o clasificadores, que son de mucha utilidad cuando se está aprendiendo en un nuevo campo, ya que proporciona la lista de textos y las páginas más importantes en el área, lo que permite echar un vistazo rápido a lo más relevante. Además, Google Académico incorpora investigación histórica desde inicios del siglo XX (y más vieja) aplicando la tecnología OCR, que permite buscar en textos aun si éstos no están indexados en bases de datos tradicionales (Warner 2012).

Google Académico también ofrece la posibilidad obtener resultados basados en libros digitalizados e indexados de Google Books (Moskovkin 2013). Sin embargo, no todo son ventajas.

### **Ventajas y desventajas**

Aun cuando Google Académico tiene el respaldo de Google, contiene deficiencias que han impedido que se convierta en un sustituto de otras bases de datos. Esto es, su recopilación de datos automatizada es menos confiable que la de otras de publicación académica, como Scopus y Web of Science (WoS), que son de paga.

En un estudio realizado en 2013 (Boeker, Vach yMotschall 2013), para realizar búsqueda sistematizadas de literatura particularmente, se encontró que Google Académico no estaba listo para ser utilizado como herramienta profesional en una búsqueda metodológica de este tipo. Esto se debe principalmente a algunas cuestiones técnicas en las búsquedas; entre ellas, el número de caracteres permitido de la cadena de búsqueda, las falta de operación con palabras trucas, y la falta de opciones para construir conjunciones y disyunciones.

Algunas limitaciones técnicas de este motor de búsqueda han ido siendo corregidas a través de los años. Sin embargo, existen barreras logísticas documentadas entre las que se incluyen las siguientes (Fagan 2017, 8):

- Los filtros de búsqueda están limitados a doscientos cincuenta y seis caracteres
- No se pueden obtener más de mil resultados
- No se pueden desplegar más de veinte resultados por página
- No se pueden bajar lotes de resultados (e.g. llamar en un manejador de citas)
- No se indica el formato que tienen los resultados (e.g. si son artículos de conferencia)
- Sólo tiene operadores básicos booleanos
- Se desconoce el algoritmo mediante el que se clasifican los resultados

Una de las críticas más fuertes hacia Google Académico es que genera duplicidad de citas, cuya causa tiene origen en las múltiples versiones que contiene de un mismo artículo (Fagan 2017,3). Sin embargo, a este respecto, en uno de los primeros estudios de este tipo realizado por Huang y Yuan (2012), se encontró que las citas de Google Académico proveen un registro bien organizado en términos de disciplina y autoría; es decir, es una herramienta útil como fuente para el análisis de citas. Los resultados encontraron una fuerte correlación entre el *h*-índice, el *i10*-índice y el número total de citas. Otro estudio más reciente confirma estos resultados (Thelwall y Kousha 2017): Google Académico encuentra más citas que otras bases de datos como ResearchGate, WoS y Scopus.

Aun cuando Google Académico asegura que “clasifica sus resultados de la forma en que lo hacen los investigadores, considerando el texto completo, dónde fue publicado, quién lo escribió, así como que tan frecuente y recientemente ha sido citado en la literatura escolar”<sup>6</sup>, se ha encontrado una fuerte correlación entre la lista de clasificación o *ranking*, y las veces que un artículo científico ha sido citado (Martín-Martín *et al.* 2017, 159), lo que

---

<sup>6</sup> <https://www.google.com/intl/en/scholar/about.html>.

ocasiona que un texto influyente se haga más influyente, ello causa que textos meritorios queden prácticamente enterrados en estas listas.

Otra fuerte crítica que ha recibido Google Académico se refiere a la inclusión de revistas predatorias; es decir, las revistas que cobran al autor por el acceso abierto sin tener una rigurosa revisión por pares. Al parecer, no existe un estricto control de calidad de las fuentes de origen. Además, no hay información sobre la frecuencia de sus actualizaciones. Por otro lado, que Google Académico haga la cobertura de un editor, no implica que todos los artículos o revistas de dicho editor sean tratados (Fagan 2017, 23-24).

Uno de los principales problemas de Google Académico, que nos atañe a las revistas latinoamericanas, es que se ha demostrado su tendencia a dar preferencia a textos en inglés, aun cuando las publicaciones hayan sido muy citadas en otros idiomas. Al parecer, Google Académico utiliza el lenguaje del *host* en el que se aloja la página Internet, lo que significa que, si el documento está alojado en un sitio de otro idioma que no sea el inglés, incluso cuando si se encuentra escrito en inglés, aparecerá en posiciones menores (Martín-Martín *et al.* 2017, 161).

A esto se le agrega que los repositorios de Latinoamérica no son muy visibles en Google o Google Académico debido tanto a limitaciones en la descripción de esquemas que las revistas latinoamericanas seleccionan, como a la confiabilidad del motor de búsqueda (Orduña-Malea y Delgado López-Cózar 2015). La indexación de repositorios de acceso abierto en Google Académico varía conforme a la plataforma del repositorio institucional que se use, así como al hecho de que éste haya sido estructurado conforme a los lineamientos de Google. En algunos casos, aun cuando el texto ha sido indexado, no se encuentra completo. Ello se debe más comúnmente a que se guardan los escritos en un solo documento PDF (en lugar de crear archivos HTML con las recomendaciones de Google), y a que no se someten directamente a las páginas a Google Académico. Los lineamientos para la inclusión de textos académicos pueden encontrarse directamente en Google Académico.<sup>7</sup>

---

<sup>7</sup> <https://scholar.google.com/intl/en/scholar/inclusion.html>.

## **Visibilidad de revistas latinoamericanas en Google Académico**

Como ya fue señalado por varios autores (Farga 2017), Google Académico tiene importantes debilidades. Sin embargo, hasta el día de hoy no hay una herramienta que, sin costo, lo haga mejor. Aun cuando Google Académico presenta un sesgo hacia las publicaciones en inglés, debemos considerar que la mayoría de las bases bibliográficas también lo tienen. Google Académico, incluye mayor diversidad en tipos de documentos relevantes, comparado con servicios de indexación y resúmenes. A pesar de que para las revisiones sistemáticas se recomienda usar otras fuentes o no usarlo como única fuente, es un hecho que Google Académico sí agrega valor a otras bases de datos. En general, muchos investigadores consideran tolerable el ruido en Google Académico, dados su beneficios (Ştirbu *et al.* 2015).

## **Conclusión**

A pesar de sus limitantes, Google Académico ha incrementado la visibilidad de la investigación en la comunidad internacional, y su uso va en crecimiento. Consideramos que las revistas, especialmente las que no están en bases de datos de pago, deben hacer uso de esta ventaja, pero deben modificar sus lineamientos de publicación o bien someter su información directamente a las páginas de Google Académico.

## **Referencias**

- Boeker, Martin, Werner Vach y Edith Motschall. 2013. "Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough." *BMC medical research methodology* 13.1: 131.
- Fagan, J. 2017. "An evidence-based review of academic web search engines, 2014–2016: Implications for librarians' practice and research agenda". *Information Technology and Libraries*. 36: 7–47.
- Huang, Z. y B. Yuan. 2012. "Mining Google Scholar Citations: An Exploratory Study". En *International Conference on Intelligent Computing*. Springer, Berlin, Heidelberg: 182-189.

Kuyoro Shade, O., O. Okolie Samuel, U. Kanu Richmond y Awodele Oludele. 2012. "Trends in web-based search engine". *Journal of Emerging Trends in Computing and Information Sciences*, 3:6.

Martín-Martín, Alberto, Enrique Orduña-Malea, Anne-Wil Harzing y Emilio Delgado López- Cózar. 2017. "Can we use Google Scholar to Identify Highly-Cited Documents?" *Journal of Informetrics* 11.1: 152-163. doi:10.1016/j.joi.2016.11.008.

Moskovkin, V. M. 2013. "The construction of academic publishing and terminological structures using the Google Scholar search engine: An example of environmental terms in publications at the classical universities of Kharkiv and Skopje". *Scientific and Technical Information Processing* 40.1: 11-16.

NetMarketShare. 2018. "Market Saher Statistics for Internet Technologies". NetMarketShare. Acceso el 18 de enero del 2018. <https://www.netmarketshare.com/desktop-search-engine-market-share>.

Orduña-Malea, Enrique y Emilio Delgado López-Cózar. 2015. "The Dark Side of Open Access in Google and Google Scholar: The Case of Latin-American Repositories." *Scientometrics* 102.1: 829-846. doi:10.1007/s11192-014-1369-5.

Știrbu, Simona, Paul Thirion, Serge Schmitz, Gentiane Haesbroeck, y Ninfa Greco. 2015. "The Utility of Google Scholar when Searching Geographical Literature: Comparison with Three Commercial Bibliographic Databases." *The Journal of Academic Librarianship* 41.3: 322-329. doi:10.1016/j.acalib.2015.02.013.

Thelwall, M., y K. Kousha. 2017. "ResearchGate versus Google Scholar: which finds more early citations?" *Scientometrics* 112.2: 1125-1131.

Warner, B. "Is Google Scholar a reliable tool?" 2012. Acceso el 2 de febrero del 2018. <https://submissions.scholasticahq.com/conversation/questions/is-google-scholar-a-reliable-tool>.