



Visibilizando la ciencia: indexación en bases de datos

Aliena Núñez González, Residente de Segundo Año de Bioquímica Clínica, Universidad de Ciencias Médicas de Granma, Departamento de Ciencias Fisiológicas, Manzanillo, Cuba.
Correo electrónico: angjmyoon9357@gmail.com.

Resumen

En la actualidad la información ya no se considera como algo inerte o estático, sino, como un ente dinámico en constante expansión. La ciencia no escapa de este marco de globalización, y con la aparición de los sistemas de información en los años setenta se comenzó a expandir el abanico de oportunidades que hoy representa la web para la difusión de los productos científicos a gran escala, siendo una necesidad otorgarle visibilidad al nuevo conocimiento producido. De esta manera aparecen herramientas de uso prioritario como parte de los sistemas informacionales, a fin de facilitar el proceso editorial y conectar a los autores de todo el planeta. Con el propósito de evidenciar la estrecha relación entre el proceso de indexación en bases de datos y la visibilidad de las revistas científicas en la arena internacional se realizó esta investigación de tipo revisión bibliográfica. Desde la sistematización del estudio en el análisis bibliográfico efectuado se obtuvo la conclusión de que es prioritario familiarizarse con el proceso de indexación en bases de datos editoriales, puesto que no es solo tarea del cuerpo editorial de las revistas conocer los criterios que garantizan una mejor socialización y crecimiento del conocimiento científico, sino una máxima para los autores, que en la mayoría de las ocasiones desconocen las características del proceso que opera tras bambalinas desde la concepción de un artículo hasta su publicación.

Palabras clave: indexación, base de datos, ciencia abierta, visibilidad, información.

Introducción

En la actualidad la información ya no se considera como algo inerte o estático, sino, como un ente dinámico en constante expansión. En el mundo de hoy este hecho representa el quehacer de multitud de personas y organizaciones. Puesto que todos estamos expuestos a gran



volumen de información, surge la necesidad de clasificar, ordenar, almacenar y validar esta, afín de facilitar el acceso al conocimiento y con él, el desarrollo de nuestra especie.

Estados Unidos fue uno de los pioneros en la aparición de los sistemas de información, siendo Medline el mejor ejemplo en el área de las Ciencias Médicas. Posteriormente, se comenzaron a desarrollar otros sistemas en varias partes del mundo, como el EUDISED en Europa y en América Latina el REDUC, el cual constituyó un esfuerzo por integrar el conocimiento de toda la región¹. De esta manera, se fue construyendo un verdadero paradigma integrado con el objetivo común de difundir el conocimiento contenido en el trabajo realizado por la comunidad científica internacional.¹

Hoy en día, Internet es la puerta para el desarrollo de la mayoría de las publicaciones académicas. Por ende, fue necesario perfeccionar la forma de clasificar, almacenar y validar todo ese nuevo flujo de información generado y dirigido por los usuarios de la web. Asimismo, surge la necesidad de ampliar el espectro de visibilidad de las publicaciones científicas con el propósito de lograr una mayor cobertura y socialización de los resultados obtenidos. Como parte del movimiento del Open Acces la información dejó de ser parte del núcleo selectivo que constituían las primeras revistas científicas convirtiéndose en un multiverso de posibilidades, capaz de sumar mayor número de colaboradores y hacer evolucionar el proceso de gestión editorial a un ritmo colosal.^{1, 2}

De esta forma, cabe preguntarse a qué criterios habría que atenerse para valorar la calidad de una publicación científica y en mayor medida a la procedencia de esta, es decir, la revista o sitio web que avale la verosimilitud de la comunicación científica publicada. Así aparece el término “indización” que es utilizado en el campo de la bibliotecología, tanto para el proceso de análisis y extracción de términos o frases representativas de un contenido informativo como para la construcción del índice que contendrá dichas expresiones y se utilizará como herramienta de acceso a esa información.²

En este sentido, cuando se dice que una revista está “indizada” en una base de datos se refiere al hecho de que su contenido (los artículos y/o su registro madre) fue debidamente registrado (descrito y analizado) y que sus elementos más representativos (palabras clave, títulos, autores, etc.) forman parte de su índice, por el cual podrán eventualmente, ser accedidos. Este



proceso se observa comúnmente en Internet, a través de la indexación web que lleva adelante el buscador Google Académico de los documentos alojados en servidores del rubro en toda la red.³

El primer paso para proyectar la visibilidad de una revista es su incorporación a una fuente documental secundaria, encargada de referenciar a publicaciones de características similares o del mismo campo temático con el objeto de aumentar su acceso y difusión. Técnicamente, este proceso se denomina “indizar una revista” en una base de datos, índice, repertorio u otra fuente secundaria. Existe una gran variedad de sitios en donde incluir una publicación, tanto desde el punto de vista del alcance geográfico y cobertura temática como de los aspectos técnicos y normativos requeridos.⁴

Objetivo: Evidenciar la estrecha relación entre el proceso de indexación en bases de datos y la visibilidad de las revistas científicas en la arena internacional.

Desarrollo

El hecho que una publicación o revista científica se encuentre indexada significa que la misma denota calidad y ha sido encasillada en alguna base de datos de consulta con un alcance determinado. La indexación de las revistas o publicaciones científicas en índices, repositorios o bases de datos de alto impacto es vital para la reputación y alcances de los documentos publicados en estas. Sin embargo, para la inclusión de las revistas en algunos de estos sistemas es necesario que se cumplan con ciertos criterios de calidad, los cuales estarán sujetos al tipo de sitio que se pretenda indexar.⁴

Dentro de los criterios generales que evalúan los sistemas de información se encuentran:

- La calidad informativa: se examina que las revistas cuenten con un cuerpo editorial el cual debe ser identificable en su plataforma, así como, establecer de forma clara las orientaciones para los autores, el proceso de evaluación, traducción de los resúmenes en otro idioma distinto al de origen y el uso de palabras claves.
- Las características formales: es necesario que las revistas cumplan con la periodicidad que se han impuesto, también, que realizan revisión por pares, contar con un comité científico externo y utilizar los canales adecuados para dar a conocer las decisiones editoriales.⁵



- El uso que se le da a la información y calidad científica de los contenidos: se evalúa que los autores que realizan publicaciones sean externos a la entidad editora, así como, un alto contenido científico.

Una vez revisados estos aspectos, los editores están en condiciones de evaluar si se cumplen con los requisitos generales para indexar sus publicaciones en la base de datos que mejor se adapte a la temática escogida. De esta manera es muy útil conocer que existe una clasificación de estos sistemas de información, que van de lo básico como los catálogos de bibliotecas hasta lo complejo como las bases de datos de alto impacto Scopus, Scimago Journal y Web of Science.⁴

La indización busca la identificación y asignación de rubros, descriptores o encabezamientos de materia representativos. En esencia, pretende facilitar el acceso y la recuperación de información sobre un tema determinado. Además, es capaz de incrementar la relevancia de los resultados de las búsquedas de información en base de datos a partir de la representación precisa del contenido del documento y las solicitudes de información de los usuarios.⁵

Este proceso implica dos fases. Primero la identificación de los contenidos tratados en el documento objeto de interés y segundo, la conversión de las palabras y frases utilizadas en el lenguaje natural para representar una realidad en términos de un lenguaje controlado que pretende elevar la calidad de la recuperación de la información. Entre los factores determinantes del éxito en la recuperación de información se encuentra el dominio del usuario del vocabulario empleado para representar los documentos de interés y las solicitudes de búsqueda.^{5, 6}

La indización es una de las etapas del procesamiento analítico sintético de la información, que se ubica por su complejidad intelectual entre la clasificación y la confección de resúmenes. El producto final de este proceso es generalmente un índice bibliográfico, una base de datos automatizada o simplemente un catálogo alfabético de materias, indispensables para asegurar el acceso y consulta de la información a los usuarios. Esta aseveración significa que se toman en cuenta los mismos criterios para la ejecución. Ellos son: exhaustividad, multiplicidad, especificidad, coherencia, imparcialidad, fidelidad y buen juicio.⁶



En la indización, *la especificidad* y *la exhaustividad* constituyen los dos atributos cardinales. Estos son esenciales para determinar la clase de recuperación que se realizará.

- La exhaustividad: se define por el grado de cobertura que presentan los tópicos o asuntos tratados en los documentos o solicitudes en los resultados de la indización, es decir, en el producto. Cuanto mayor sea la representación de los tópicos expuestos en un documento o solicitud en su indización, mayor será la exhaustividad del proceso. Se expresa cuantitativamente por el número de términos utilizados para realizar la representación. Comprende tópicos centrales y colaterales.
- La especificidad: es el nivel de detalle y exactitud con que se representa cada tópico tratado en un documento o solicitud en la indización. La recuperación de información de manera eficiente requiere de la determinación de los términos más específicos que representan los conceptos presentes en un documento o solicitud.^{7, 8}

El tesauro es un tipo de vocabulario controlado, compuesto esencialmente por una lista de términos autorizados, que se utilizan para indizar (representar semánticamente) los registros que ingresan a una base de datos o sistema de información, así como las solicitudes de búsqueda que formulan sus usuarios. Es el medio más eficaz para realizar una búsqueda por materia. Sin embargo, los tesauros, son incapaces de reunir el total de los conceptos tratados en la literatura especializada. Por esta razón, los índices y los estudios de la frecuencia de empleo de las palabras y frases relacionadas con un tema de interés son fundamentales.⁷

En el área de la salud, uno de los tesauros más empleados es el MeSH (Medical Subject Headings), un vocabulario controlado y jerárquico de encabezamientos de materias elaborado por la Biblioteca Nacional de Medicina de los Estados Unidos. Este, a su vez, presenta una versión en español: el DeCS (Descriptores en Ciencias de la Salud) muy útil para aquellos usuarios que no dominan la terminología médica en inglés.⁸

El MeSH y el DeCS sugieren relaciones entre los términos a partir de una estructura jerárquica de términos genéricos y específicos, que enlaza términos no autorizados y autorizados, así como los relacionados entre ellos. Los tesauros, al igual que el índice, ayudan a la localización de los términos más específicos. Sus referencias cruzadas remiten de los vocablos no autorizados muchas veces sinónimos o subordinados a los términos autorizados. Los tesauros



referidos permiten recorrer una jerarquía de términos en dos direcciones, con vista a delimitar o ampliar los resultados de la búsqueda.⁸

SciELO

Una de las iniciativas más innovadoras en Latinoamérica respecto al proceso de digitalización de las revistas y el acceso abierto al conocimiento científico producido en la región fue el desarrollo del proyecto SciELO (acrónimo de Scientific Electronic Library Online). Se trata de una biblioteca virtual de revistas científicas surgida inicialmente en Brasil en formato electrónico, accesible en texto completo, que además produce un módulo de evaluación basado en indicadores de uso e impacto.⁹⁻¹¹

SciELO entró en funcionamiento en 1998 como resultado de una alianza entre la Fundação de Amparo à Pesquisa do Estado de São Paulo, el Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (BIREME) y editores de revistas científicas, denominándose originalmente Biblioteca Científica Eletrônica Online.¹⁰ Resulta interesante observar la convergencia de los distintos intereses y objetivos que llevaron a crear lo que en la actualidad se denomina el modelo SciELO de publicación electrónica, ampliamente extendido en muchos países de Latinoamérica. Uno de los componentes de mayor peso en el impulso de este proyecto fue la necesidad de aumentar la visibilidad de la producción científica nacional y la de crear mecanismos de evaluación complementarios a los ya existentes como forma de equiparar los estándares utilizados internacionalmente.¹⁰

Por otro lado, la oportunidad de desarrollar una metodología de publicación electrónica de acceso abierto que operara como infraestructura de apoyo a la investigación científica y que produjera un modelo para ser replicado en los países de la región constituyó otro móvil consistente. SciELO ha procurado además que los registros almacenados en sus bases de datos obtengan la mayor visibilidad internacional posible por medio de la extracción y exportación de los metadatos de los textos completos, a bases de datos y repositorios externos.¹¹

La función de publicación va acompañada de la interoperabilidad de los contenidos en la web a través de una distribución masiva de metadatos a los principales índices bibliográficos y los servicios que proporcionan acceso universal eficiente a los textos completos. Para ello



implementa técnicas de indexación por medio de protocolos de interoperabilidad con fuentes externas como Google y Google Académico, CrossRef, PubMed, DOAJ, entre otras.¹²

Este aspecto reviste mucha importancia porque transforma a la plataforma en un sistema que amplía considerablemente los circuitos de circulación de las producciones científicas nacionales al tiempo que provee información estadística de uso e impacto de las colecciones. A partir del 2007, más del 70% de los accesos realizados en el sitio SciELO Brasil (primera base de datos indexada por terceros) se obtienen de los resultados de búsqueda de Google. Esto significa un aumento exponencial en la cantidad de consultas realizadas pasando de 103 millones de accesos y descargas anuales en 2007 a 252 millones en 2012.^{12, 13}

Redalyc (Red de Revistas Científicas de América Latina y el Caribe, España y Portugal)

Redalyc es una de las bases de datos que más se destaca por su crecimiento y cobertura. Fue puesta en marcha en el 2003 por la Universidad Autónoma del Estado de México cuyo objetivo original fue la creación de una hemeroteca de acceso gratuito en línea de revistas regionales en acceso abierto. Aunque inicialmente incluía solo publicaciones de disciplinas de Ciencias Sociales y Humanas, por considerarlas las más marginadas en los circuitos de circulación de información científica, a partir de 2006 el proyecto se abrió a todas las áreas de conocimiento.¹³

En la actualidad Redalyc es una plataforma que indiza 1138 36 títulos de veintidós países iberoamericanos y se estructura sobre estándares internacionales de acceso abierto. De esta manera, garantiza la implementación de mecanismos de interoperabilidad que promueven una mayor difusión y visibilidad de las revistas y contenidos asociándose a actores importantes en la comunicación científica, como el Directory of Open Access Journals (DOAJ), Journal Tocs, ScientificCommons.org y Google Scholar, entre otros.¹³

Redalyc es en la actualidad un portal autodenominado “*Sistema de información científica*”, que provee no solo la búsqueda y acceso a miles de artículos científicos de fuentes iberoamericanas, sino a una serie de productos relacionados con el análisis estadístico y bibliométrico de la información almacenada accesible en línea y difundida por medio de informes de producción. El proceso de evaluación de las revistas para ser indizadas en esta base de datos es exhaustivo y consta de sesenta criterios.^{13, 14}



Dialnet

Es un servicio de indización de revistas nacido para dar cobertura y visibilidad a publicaciones en Iberoamérica. El proyecto se inicia en el 2000 en la Universidad de La Rioja, España, con el objetivo de ofrecer inicialmente la disponibilidad en línea de las fuentes y registros contenidos en bibliotecas universitarias, públicas y especializadas. Con el tiempo, Dialnet, se convirtió en una plataforma de recursos y servicios documentales con la ambición de transformarse en la mayor hemeroteca de artículos científicos hispanos en Internet.¹⁴

A partir de 2009, este objetivo cobra verdadero impulso con la creación de la Fundación Dialnet y la puesta en marcha de un proyecto más ambicioso orientado no solo a ensanchar la base de indización de revistas sino a agregar valor a los servicios ofrecidos. Así nace Dialnet Plus, que ofrece una gama más amplia de opciones de acceso a su base de datos y presenta diferentes alternativas de personalización en la gestión de datos por parte del usuario. También ofrece la posibilidad de obtener estadísticas e informes relacionados con algunos indicadores bibliométricos y de uso de las publicaciones.¹⁵

En la actualidad Dialnet indiza aproximadamente 10000 títulos de revistas a texto completo en su gran mayoría de acceso abierto. También contiene publicaciones por suscripción tanto en formato electrónico como en papel, aunque este último formato ya no es aceptado. El proceso de inclusión puede realizarse por medio de la institución editora de la revista, que deberá registrarse como entidad colaboradora o individualmente.¹⁵

Google Académico

De las opciones existentes en la aplicación de herramientas que potencien las estrategias de difusión y visibilidad de una revista científica, Google Académico (GA) se ubica entre las mejores al ser simple, de cobertura universal, multilingüe, rápido, y sobre todo, gratuito. Las principales fortalezas que pueden esgrimirse se relacionan con los siguientes ítems:

- La alta recuperación de literatura científica escrita en otros idiomas por fuera del inglés, especialmente el español, la cual suele ser dispersa y pobremente estructurada.



- El acceso a producción científica de las Ciencias Sociales y Humanas por fuera de la corriente principal y con formas de publicación que todavía no utilizan al artículo como vía preferente de comunicación.
- La posibilidad de medir el impacto mediante el recuento de citas recogidas en un universo mucho más amplio de fuentes.
- El hecho de ser una herramienta en constante evolución y crecimiento.¹⁶

En esencia, GA es un buscador especializado en información científica y académica nacido en 2004 como aplicación beta derivada de su hermano mayor Google que proporciona, además de la búsqueda y acceso a documentos, una gama de posibilidades relacionadas con el seguimiento del impacto y rendimiento editorial de la producción científica. Opera como un rastreador en el contexto de la denominada *web académica* recopilando e indexando referencias bibliográficas, texto completo y citas recibidas por los documentos.^{15, 16}

Su cobertura se centra especialmente en artículos de revistas y libros, comunicaciones y ponencias de congresos, informes, tesis y otros tipos de materiales depositados en repositorios. El filtrado lo realiza desde los espacios web de información alojados en sitios académicos. Para ello toma en consideración perfiles y webs académicas, repositorios temáticos e institucionales, editoriales, bibliotecas y catálogos, bases de datos bibliográficas, servicios de distribución o almacenaje, entre otros.¹⁶

Google Académico es una de las herramientas más utilizadas en el mundo académico para la búsqueda y acceso a la información científica. Muy por encima de otros servicios de información consagrados como WoS y SCOPUS. Su cobertura aproximada supera en millones de documentos, en más de tres veces a la almacenada en WoS y SCOPUS así como en más de cuatro al buscador Microsoft Academic Search, su principal competidor en este segmento.¹⁶

Conclusiones

Es prioritario familiarizarse con el proceso de indexación en bases de datos editoriales, puesto que no es solo tarea del cuerpo editorial de las revistas conocer los criterios que garantizan una mejor socialización y crecimiento del conocimiento científico, sino una máxima para los autores, que en la mayoría de las ocasiones desconocen las características del proceso que opera tras bambalinas desde la concepción de un artículo hasta su publicación.



Referencias bibliográficas

1. Aguillo, I., Open Science: Retos y Oportunidades, VI Congreso Internacional de Tecnologías Emergentes y Sociedad, Logroño, España (2018).
2. Valdés Abreu MC. Necesidad de las notas de alcance de los calificadores en la "Introducción" del tesoro Descriptores en Ciencias de la Salud (DeCS) Acimed [Internet]. 2006 enero-abril [Citado 29 enero 2008]; 4(1): [aprox. 15 p.]. Disponible en: http://www.bvs.sld.cu/revistas/aci/vol4_1_06/aci04196.htm
3. Rodríguez Camiño R. MeSH o DeCS: algunas consideraciones sobre la indización biomédica. Acimed [Internet]. 1998 [Citado 28 ene 2008]; 6(3): [aprox. 19 p.]. Disponible en: http://www.bvs.sld.cu/revistas/aci/vol6_3_98/aci04398.htm
4. Rodríguez Suárez A, Bermello Navarrete R, Pinillo León AL. Indización en línea: ¿capricho o necesidad? Acimed [Internet]. 2007 [Citado 28 enero 2008]; 15(1): [aprox. 13 p.]. Disponible en: http://www.bvs.sld.cu/revistas/aci/vol15_1_07/aci17107.htm
5. Ainsworth, S. y Russell, J., Has hosting on science direct improved the visibility of Latin American scholarly journals? A preliminary analysis of data quality. <https://doi.org/10.1007/s11192-018-2725-7>, Scientometrics, 115, 1463-1484 (2018)
6. Björk, B.-C., Journal Portals-An Important Infrastructure for Non-commercial Scholarly Open Access Publishing. <https://doi.org/10.1108/OIR-03-2016-0088>, Online Information Review, 41(5), 643-654 (2017)
7. Castro-Romero, A., González-Sanabria, J. S. y Ballesteros-Ricaurte, J. A., Technologies for Metadata Management in Scientific Articles. <https://doi.org/10.25100/iy. v17i2.2195>, Ingeniería y Competitividad, 17 (2), 123-134 (2015)
8. Collazo-Reyes, F., Luna-Morales, M. y Luna-Morales, E., Change in the publishing regime in Latin America: from a local to universal journal Archivos de investigación Médica/Archives of Medical Research (1970-2014). <https://doi.org/10.1007/s11192-016-2207-8>, Scientometrics, 110, 695-709 (2017)
9. González-Sanabria, J. S., Díaz-Peñuela, J. S., y Castro-Romero, A, Análisis de los Indicadores de Citación de las Revistas Científicas Colombianas en el Área de Ingeniería. <https://doi.org/10.4067/S0718-07642019000200293>, Información Tecnológica, 30 (2), 293-302 (2019)



10. Graf, C., Wager, E. y otros cuatro autores, Best Practice Guidelines on Publication Ethics: a Publisher's Perspective. <https://doi.org/10.1111/j.1742-1241.2006.01230.x>, International Journal of Clinical Practice, 61 (s152), 1-26 (2007)
11. ISO 14721:2012: International Organization for Standardization (Suiza), Space Data and Information Transfer Systems- Open Archival Information System (OAIS); 1-25, Ginebra-Suiza (2012)
12. Keller, T., Review of Digital Preservation Essentials, Journal of Western Archives, ISSN: 2154-7149, 8 (1), 2 (2017)
13. Larivière, V., Haustein, S. y Mongeon, P., The Oligopoly of Academic Publishers in the Digital Era. <https://doi.org/10.1371/journal.pone.0127502>, PLoS ONE, 10 (6), e0127502 (2015)
14. Ortega, J. L. y Aguillo, I. F., Microsoft Academic Search and Google Scholar Citations: Comparative Analysis of Author Profiles. <https://doi.org/10.1002/asi.23036>, J Assn Inf Sci Tec, 65: 1149-1156 (2014)
15. Pavas, A., Report About the Preliminary Results of the National Bibliographic Index – Publindex, Ingeniería e Investigación, ISSN: 0120-5609, 37 (1), 1-6 (2017).
16. BIREME. Manual de Indización. Capítulo de calificadores. [Internet]. [Citado 2 julio 2010]. Disponible en: <http://www.cpicmha.sld.cu/metodologias/manual.%20indizacalificadores.pdf>