

Similitudes y diferencias entre *PubMed*, *Embase* y *Scopus*

Similarities and differences between PubMed, Embase and Scopus

Lic. Rubén Cañedo Andalia,^I Lic. Mario Nodarse Rodríguez,^{II} Lic. Niurka Labañino Mulet^I

^I Centro Provincial de Información de Ciencias Médicas. Universidad de Ciencias Médicas de Holguín. Holguín, Cuba.

^{II} Centro Nacional de Información de Ciencias Médicas-Infomed. La Habana, Cuba.

RESUMEN

La búsqueda y recuperación de información es una etapa fundamental del quehacer científico. La búsqueda de información en bases de datos bibliográficas internacionales prominentes es una práctica común en este contexto. Pero, antes de realizar una búsqueda de información en una de estas, es necesario conocer qué contienen y cuán similares y diferentes son entre ellas, sus facilidades, herramientas, alcance y tipología documental procesada, entre otros aspectos. Obtener esta información es ciertamente difícil, porque existen múltiples fuentes de información con datos dispersos, parciales y faltos de exactitud. Por eso, se realizó una profunda investigación documental y se consultó a los expertos que laboran en el desarrollo de bases de datos como *PubMed-Medline*, *Embase* y *Scopus*, con vistas a ofrecer a los profesionales de la salud del país, y en especial a los investigadores, una caracterización precisa de ellas.

Palabras clave: servicios de información, base de datos, salud.

ABSTRACT

Information search and retrieval is a fundamental stage in scientific activity. The search for information in prominent international bibliographic databases is common practice in this context. However, before conducting an information search in databases, it is necessary to be aware of their contents, as well as the similarities and differences between them, their features, tools, scope and document typology, among other aspects. Obtaining such information is certainly difficult, due to the large number of information sources containing partial, dispersed and inaccurate data. This is the reason why a profound document research was conducted and consultations were held with experts engaged in the development of databases like PubMed-Medline, Embase and Scopus, with a view to providing health professionals and particularly researchers throughout the country with a precise characterization of such databases.

Key words: information services, database, health.

INTRODUCCIÓN

Un conocimiento profundo de las características de las bases de datos más prominentes en un sector o área de la actividad científica no es solo un requisito indispensable para los profesionales de la información, sino también de científicos, académicos y tecnólogos, porque ese es un prerrequisito importante para la obtención de la información especializada que cada uno de ellos requiere para su quehacer profesional, sus proyectos e investigaciones, así como para su actualización sistemática.

Saber dónde buscar qué forma parte del conjunto de competencias en información que debe poseer un profesional. Sin embargo, esta es una de las competencias más difíciles de adquirir, ya que solo un examen profundo de los principales recursos de información disponibles en un área del conocimiento, su uso extenso e intenso y la comparación de sus resultados obtenidos, permite desarrollar el conocimiento necesario sobre las fortalezas y debilidades de cada uno de ellos. El logro de este conocimiento a partir de la experiencia personal es un proceso complejo y costoso en tiempo y esfuerzo.

A causa de esta necesidad, se decidió estudiar con profundidad las características y fortalezas, así como las principales similitudes y diferencias entre tres bases de datos utilizadas con gran frecuencia a escala internacional para la búsqueda de información en salud: *PubMed*, *Embase* y *Scopus*, con vistas a proveer a los profesionales de la salud con un conocimiento comparado sobre sus principales características y fortalezas.

MÉTODOS

ESTUDIO DE LAS CARACTERÍSTICAS Y FORTALEZAS DE LAS BASES DE DATOS *PUBMED, EMBASE Y SCOPUS*

PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>), desarrollado por el Centro Nacional para la Información Biotecnológica (NCBI por sus siglas en inglés), de la Biblioteca Nacional de Medicina de los Estados Unidos, una entidad con rango de instituto nacional de salud, es el recurso bibliográfico gratuito más utilizado en el área de la salud en Internet.

Embase (<http://www.elsevier.com/online-tools/embase/>), producido por Elsevier, la casa editorial líder en Ciencia, Tecnología y Medicina a escala mundial, es la mayor base de datos de resúmenes con información biomédica y el recurso líder en la búsqueda de información en salud en Europa.

Scopus (<http://www.elsevier.com/online-tools/scopus>), creada en 2004 por Elsevier B. V., es la mayor base de datos multidisciplinaria de citas y resúmenes de literatura arbitrada y de fuentes de alta calidad en el Web. A pesar de su reciente aparición en el mercado, posee una creciente popularidad en la comunidad académica y científica mundial. Aproximadamente la mitad de los títulos procesados por *Scopus* clasifican en las áreas de ciencias de la vida y medicina.

PROCEDIMIENTO

Se realizaron múltiples búsquedas, tanto en bases de datos estudiadas como en las plataformas de productos del *Nacional Council for Biotechnology Information* de la *Nacional Library of Medicine of United States of America*; y de Elsevier, esta última productora de *Embase* y *Scopus*. Se revisaron numerosos catálogos promocionales y documentos no publicados sobre el tema. Se comparó la información suministrada por cada fuente y se sintetizó. Los aspectos contradictorios se aclararon mediante nuevas búsquedas de información y, en no pocos casos, por medio de la consulta a los expertos que atienden los productos analizados.

RESULTADOS

PUBMED

Cubre los campos de la medicina, la enfermería, la estomatología, la veterinaria, la gestión de salud, las ciencias preclínicas y algunas áreas de las ciencias de la vida ([tabla](#)). La retrospectividad de su colección alcanza el año 1809. *Medline*, la subbase más importante de *PubMed*, comprende registros de artículos publicados entre 1966 y el presente. *OldMedline* abarca el periodo 1946-1965. Ambas atesoran más de 20 millones y medio de registros (aproximadamente el 89 % del total reunido por *PubMed*). En *Medline* se procesan más de 5 600 revistas, seleccionadas mediante un riguroso proceso de evaluación.¹

Tabla. Principales características de *PubMed*, *Embase* y *Scopus*

Aspectos	PubMed	Embase	Scopus
Cobertura temática	Biomedicina, medicina, ciencias de la salud	Biomedicina, medicina, ciencias de la salud	Multidisciplinaria
Total de registros	≈ 24 millones	+ 28 millones	+ 50 millones
Cobertura de títulos	+ 25 000	+ 8 300	≈ 22 000
Vocabulario controlado	Sí	Sí	No
Interfaz	Especializada	Especializada	No especializada

Fuente: *PubMed*, *Embase* y *Scopus*, 12 de septiembre de 2013.

En su colección, *PubMed* posee referencias de unas 220 000 revisiones sistemáticas (más de 8 000 realizadas por los grupos de revisión de la Colaboración Cochrane), más de 18 000 guías para la práctica clínica, más de 700 000 informes de ensayos clínicos y más de 1,8 millones de artículos de revisión (12 de septiembre de 2013). Presenta una tendencia marcada a procesar materiales publicados por revistas de los Estados Unidos, así como en idioma inglés. Entre 2005 y 2009, casi el 45 % de los registros que ingresaron a esta base procedieron de revistas estadounidenses, y el 91 % se encontraba en inglés.²

Una búsqueda en *PubMed* comprende: referencias bibliográficas (en proceso de inclusión en *Medline*); referencias a fuentes (revistas esencialmente) que preceden a su fecha de ingreso a la base de datos; referencias a trabajos no cubiertos en *Medline*; referencias a manuscritos de autores pertenecientes a los Institutos Nacionales de Salud de los Estados Unidos, publicados en revistas no procesadas por *Medline*, así como de algunas revistas en ciencias de la vida, etcétera.

Los artículos procesados se indizan mediante el tesoro conocido como MeSH, una lista de alrededor de 29 000 encabezamientos de materias y unas 213 000 entradas temáticas que facilitan encontrar los términos autorizados en el tesoro, además de 43 calificadores para aumentar la precisión de la búsqueda.^{1,3}

EMBASE

Su colección incluye *Embase* (1974 hasta el presente) y *Embase Classic* (1947-1973). Su retrospectividad alcanza el año 1947. Contiene los más de 20 millones de los registros reunidos por *Medline* y unos 6 millones de registros únicos, publicados por unas 2 600 revistas (31 %), que no procesa la base de datos estadounidense, donde abundan fuentes europeas, asiáticas, latinoamericanas y otras, así como en idiomas diferentes al inglés. *Medline* y *Embase* comparten aproximadamente el 71 % de referencias comunes, aunque existen diferencias importantes entre temáticas y regiones. Su cobertura alcanza más de 90 países y 40 idiomas, y su crecimiento anual es de alrededor de 1,3 millón de registros. A diferencia de *PubMed*, que se concentra preferentemente en títulos procedentes de los Estados Unidos y otros países anglófonos con un alto nivel de desarrollo y *Embase*, presenta una presencia

mayor de títulos publicados en América Latina, Europa, África y Asia. Es particularmente fuerte en información sobre medicamentos y equipos médicos.⁴

Para la realización de revisiones sistemáticas de tratamientos con medicamentos es recomendable su uso junto a bases de datos como *Medline*, *Embase* y *CENTRAL*, esta última de la Colaboración Cochrane. Es una fuente importante de evidencias clínicas. A partir del año 2009, *Embase* comenzó a incorporar a su colección, materiales de reuniones científicas hasta alcanzar un registro anual de unos 300 000 resúmenes procedentes de unas 1 000 actas de eventos, algo que la diferencia notablemente de otras bases de datos.

La información que se procesa en *Embase* se indiza por medio de su tesoro denominado EMTREE, una lista de más de 65 000 términos de materia (más de 30 000 son nombres de medicamentos y compuestos químicos) y 280 000 sinónimos, además de 78 calificadores para aumentar la precisión de la búsqueda. El tesoro de *Embase* proporciona facilidades particulares para la indización profunda de literatura sobre medicamentos y equipos médicos (existen más de 3 000 términos de indización para equipos y dispositivos médicos). Las referencias importadas de *Medline* que forman parte de las fuentes que procesa *Embase* adquieren automáticamente la indización realizada por los expertos de esta última base de datos; los materiales únicos, es decir, que no proceden de fuentes procesadas por *Embase*, no se indizan manualmente sino que obtienen sus términos índices a partir de un sistema de equivalencias temáticas inter-tesoros MeSH-EMTREE.^{5,6}

En el año 2008, Elsevier se asoció con QUOSA para potenciar la gestión de los textos completos de las referencias recuperadas en esta base de datos como resultado de una exploración bibliográfica. QUOSA *Information Manager* empodera la gestión de la información existente en los artículos recuperados a través de la posbúsqueda automatizada de alto rendimiento en los textos completos de estos. Se distribuye mediante suscripción.⁷

SCOPUS

Cubre miles de títulos de publicaciones seriadas activas de más de 5 000 casas editoras; más de 20 000 son revistas arbitradas y más de 2 600 del total presentan acceso abierto. Contiene 29 millones de registros desde el presente hasta 1995 (el 84 % de los artículos poseen resumen; hasta 1996 presentan también sus referencias bibliográficas, procedentes de publicaciones académicas —revistas y series monográficas y comerciales—) y 21 millones de registros corresponden a contribuciones publicadas antes de 1996. Presenta, además, una extensa cobertura de materiales de conferencias (más de 5,5 millones), páginas científicas en Internet (unos 376 millones) y patentes (más de 25,2 millones) y procesa los artículos "en prensa" de más de 3 850 revistas.⁸

A pesar del carácter multidisciplinario de su colección, el 48 % de sus títulos clasifican en las categorías ciencias de la vida y de la salud. Los artículos publicados desde 1996 hasta el presente incluyen sus referencias bibliográficas, necesarias para los análisis de citación, aunque existe una gran cantidad de artículos sin sus referencias de fechas anteriores hasta el 1823. Su actualización es diaria; cada día ingresan unos 5 500 registros nuevos, aproximadamente 2 millones al año. *Scopus* es la base de datos internacional con mayor cobertura de revistas y otras fuentes de información científica de América Latina. En el año 2013 inició un esfuerzo particular por extender el número de libros disponibles en la base de datos que se extenderá hasta 2015.⁸

Scopus es la alternativa europea al monopolio que durante más de 40 años ejercieron las bases de datos del antiguo *Institute for Scientific Information* (ISI actual Thomson Reuters) en el área de los estudios de citación en el contexto científico internacional. Posee herramientas inteligentes para seguir, analizar y representar el comportamiento de la actividad en la ciencia, en especial con respecto a su consumo, a partir del empleo de los datos de citación de las obras y los autores. Además de la búsqueda bibliográfica, *Scopus* permite la obtención de una serie de indicadores métricos que permiten evaluar el comportamiento de las fuentes de información procesadas por la base de datos. Entre los abundantes indicadores métricos utilizados para la evaluación de las fuentes procesadas por esta base de datos se encuentran el *Source Normalized Impact per Paper* (SNIP) que mide el impacto de citación de una revista en su contexto, y el *SCImago Journal Rank* (SJR), desarrollado por el Grupo SCImago, y que es una medida del prestigio científico de una revista que se basa en la ponderación del valor de las citas recibidas.⁹

Algo más del 50 % de los títulos de revistas y actas de conferencias que procesa *Scopus* son únicos; es decir, no aparecen en el *Web of Science*, el porcentaje estimado de títulos únicos en *Web of Science* es del 25 %, aunque existen diferencias según temáticas y regiones que se acentúan en particular en los territorios ubicados fuera de Norteamérica y Europa occidental. Las revistas solapadas, es decir, aquellas que se procesan por ambos recursos tienden a ser las más citadas. *Scopus* abarca el 80 % de las citas emitidas por las revistas procesadas en su colección. Aunque posibilita la búsqueda por materias según términos comprendidos en el MeSH y el DeCS, siempre que los registros procesados los posean, no permite el uso de las facilidades de búsqueda propias de cada uno de los tesauros en sus respectivas bases de datos, como la construcción de la estrategia de búsqueda desde ellos o la llamada explosión de términos subordinados. *Scopus* dispone de las facilidades del sistema para gestión documental y de la información recuperada denominado QUOSA. Se distribuye mediante suscripción.¹⁰

CONSIDERACIONES FINALES

Aunque *PubMed* y *Embase* son bases de datos especializadas en biomedicina y salud, existen diferencias cuantitativas importantes en relación con el volumen de registros que procesa cada base de datos en las diferentes áreas del conocimiento, su cobertura geográfica, idiomática y según tipos de documentos.

Scopus, por su parte, posee la mayor cantidad de artículos en salud al procesar la totalidad de las contribuciones registradas en *Medline* —que constituye aproximadamente el 90 % de los artículos procesados por *PubMed*— y más del 97 % del total de títulos procesados por *Embase*, e incorporar a sus colecciones un número alto de artículos afines o de interés relacionados con los campos de las ciencias de la salud en general.

A pesar de la superioridad de *Embase* en relación con su cobertura con respecto a *PubMed* en los aspectos referidos, la distribución gratuita de *PubMed* aconseja el inicio de la búsqueda de información por esta última. Las interfaces de búsqueda de *Embase* y *PubMed* presentan facilidades específicas para la exploración bibliográfica por medio de lenguajes controlados en sus colecciones, a diferencia de *Scopus* que, a causa de su carácter multidisciplinario, carece de una interfaz de búsqueda apropiada para las búsquedas principalmente de información de carácter clínico.

Como ventaja sobre las bases de datos referidas, presenta herramientas particularmente útiles para la evaluación cuantitativa de artículos, autores, organizaciones, países y regiones.

Es muy importante comprender que la cobertura de títulos y la retrospectividad del procesamiento de la literatura en una base de datos no son los únicos aspectos de importancia para su evaluación, porque algunos elementos —como la profundidad de la indexación mediante vocabularios controlados y la existencia de interfaces de búsqueda diseñadas de acuerdo con las particularidades del contenido temático de la base y las necesidades de sus usuarios, como sucede con *PubMed* y *Embase*, que presentan interfaces elaboradas "a la medida" de las necesidades de la búsqueda de la información biomédica, clínica y de salud— son factores importantes para aumentar la precisión de los resultados de una exploración bibliográfica.

REFERENCIAS BIBLIOGRÁFICAS

1. National Center for Biotechnology Information (NCBI). National Library of Medicine. United States. PubMed. 2013 [citado 14 de noviembre de 2013]. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/>
2. National Library of Medicine. United States. Medline. 2013 [citado 14 de noviembre de 2013]. Disponible en: <http://www.nlm.nih.gov/pubs/factsheets/medline.html>
3. National Library of Medicine. United States. MeSH. 2013 [citado 14 de noviembre de 2013]. Disponible en: <http://www.nlm.nih.gov/pubs/factsheets/mesh.html>
4. Elsevier. Embase about. 2013 [citado 14 de noviembre de 2013]. Disponible en: <http://www.elsevier.com/online-tools/embase/about>
5. Elsevier. Embase: Indexing with Life Science Thesaurus Emtree. 2013 [citado 14 de noviembre de 2013]. Disponible en: <http://www.elsevier.com/online-tools/embase/emtree>
6. Rochie AM. Embase: Tips and tricks. 2013 [citado 14 de noviembre de 2013]. Disponible en: <http://www.slideshare.net/rocheam/embase-tips-and-tricks-for-trainers-webinar-26-june-2013>
7. Elsevier. Elsevier's EMBASE.com Partners with QUOSA. 2013 [citado 23 de noviembre de 2013]. Disponible en: <http://www.elsevier.com/about/press-releases/science-and-technology/elseviers-embase.com-partners-with-quosa>
8. Elsevier. Scopus. Facts and figures. 2013 [citado 14 de noviembre de 2013]. Disponible en: http://cdn.elsevier.com/assets/pdf_file/0007/148714/scopus_facts_and_figures.pdf

9. Cañedo Andalia R, Cruz Font J. Tendencias, limitaciones y perspectivas en la evaluación de las publicaciones científicas y académicas mediante indicadores cuantitativos. En: Cañedo Andalia R, Rodríguez Labrada R, Fernández Valdés MM, Zayas Mujica R, Nodarse Rodríguez M, Sánchez Tarragó N, et al. Lecturas avanzadas para la alfabetización informacional en salud. Holguín: Centro Provincial de Información de Ciencias Médicas. Universidad de Ciencias Médicas de Holguín; 2011 [citado 26 de noviembre de 2013]. Disponible en: http://www.hlg.sld.cu/alfin/download/lecturas_avanzadas/5.1.factor_de_impacto_lista.pdf
10. Center for Research Libraries. Global Resources Network. Academic Database Assessment Tool. About Scopus. 2013 [citado 14 de noviembre de 2013]. Disponible en: <http://adat.crl.edu/databases/about/scopus>

Recibido: 9 de septiembre de 2014.
Aprobado: 6 de diciembre de 2014.

Lic. *Rubén Cañedo Andalia*. Centro Provincial de Información de Ciencias Médicas. Universidad de Ciencias Médicas de Holguín. Holguín, Cuba. Correo electrónico: ruben@infomed.sld.cu