

Los límites de Pubmed como herramientas para la investigación métrica

The PubMed Limits as tools for metric research

Rubén Cañedo Andalia

Licenciado en Información Científico-Técnica y Bibliotecología. Departamento Fuentes y Servicios de Información. Centro Nacional de Información de Ciencias Médicas-Infomed. La Habana. Cuba

PubMed (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?linkbar=jsmenu2>), accesible por medio de Entrez -el sistema integrado de búsqueda y recuperación de información textual, desarrollado por el Centro Nacional para la Información Biotecnológica (NCBI por sus siglas en inglés)- en la Biblioteca Nacional de Medicina de los Estados Unidos -una entidad con rango de instituto nacional de salud-, es el resultado de un proyecto conjunto de estas instituciones en colaboración con un grupo de casas editoras en el campo de la biomedicina para proporcionar a sus usuarios una herramienta de búsqueda capaz ofrecer acceso, tanto a las referencias bibliográficas como a los textos completos de los materiales recuperados.¹

El anuncio oficial de la apertura de PubMed-MedLine se realizó el 26 de junio de 1997.

PubMed es un "recurso" de recursos de información en el área de la biomedicina y el más utilizado en Internet. Su acceso es gratuito. PubMed-Medline -denominación que utilizaremos en lo adelante para diferenciar esta base de datos de la que constituye realmente Medline, su base de datos principal-, es el recurso insigne de la Biblioteca y la más popular en el campo de la salud desde hace varias décadas.

PubMed-Medline procesa más de 5 200 revistas de unos 80 países y comprende 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.¹ Sus archivos contienen más de 18 millones y medio de referencias desde el año 1865 hasta la fecha.

PubMed recibe docenas de millones de visitas anuales de usuarios de todo el mundo debido a su reconocida calidad y acceso gratuito. Con PubMed es posible acceder a múltiples bases de datos en el área de las ciencias de la vida, en particular en los campos de la genética y la farmacología, a partir de las facilidades que ofrece la plataforma del NCBI

PubMed-MedLine contiene muchos más registros que los procesados para el Medline tradicional. Una búsqueda en PubMed-Medline 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.¹

PubMed-Medline comprende, por tanto, una magnífica colección de publicaciones seriadas en el mundo de la biomedicina. Ella representa un importante subconjunto del total de ediciones en esta esfera, escogido como sumo rigor para servir a los intereses, tanto de la investigación como de la práctica médico-clínica de alto nivel. Todo ello la convierte en un excelente fuente para la obtención de valores métricos a los efectos de caracterizar, tanto la investigación como la literatura en salud y de establecer puntos o valores de referencia para realizar comparaciones con respecto a ciertos subconjuntos en temas de interés y responder interrogantes como: ¿Es actualmente el tratamiento del cáncer de mama un tema de investigación más maduro que otros estudiados en el sector de la salud?

Aunque aparentemente es difícil responder esta pregunta, en realidad no lo es, si se conoce la estructura el conocimiento biomédico. Hasta que no existen suficientes investigaciones originales en un tema es prácticamente imposible realizar revisiones sobre este. La existencia de abundantes revisiones sobre cierto tema es un buen predictor de su grado de madurez como objeto de investigación. Y aún más si estas revisiones son sistemáticas o metanalíticas y existen guías para la práctica clínica en el conjunto recuperado.

Si conocemos el valor medio, es decir, el número de revisiones realizadas y el por ciento que representa con respecto al total de documentos existentes en la base en general en el período comprendido entre el año 2000 y el presente, podemos hallar si el número de revisiones desarrolladas sobre el tratamiento del cáncer de mama (y su por ciento) es superior, igual o menor a la media de la literatura procesada por la base.

Primero, observemos el total de documentos y de revisiones incorporados en general a PubMed-Medline desde el 2000 hasta el presente. Si abrimos su página de límites y sin colocar nada en la ventana de búsqueda, seleccionamos la opción de período de tiempo, Publisher in the Last: Specify date range (YYYY/MM/DD); tecleamos "2000" en la caja correspondiente al año de inicio y oprimimos el botón Go; el sistema nos devuelve la cantidad de documentos procesados durante ese período: 5 734 040 y el total de artículos de revisión: 672 340. Si repetimos la búsqueda colocando en la ventana de búsqueda, la estrategia "Breast Neoplasms/therapy"[Mesh] con igual límite tenemos entonces:

| Alcance | Total de documentos | Total de revisiones | Por ciento |
|-----------------|---------------------|---------------------|------------|
| General | 5 734 040 | 672 340 | 11,72 |
| Tratamiento del | 30 715 | 6 163 | 20,06 |

| | | | |
|----------------|--|--|--|
| cáncer de mama | | | |
|----------------|--|--|--|

PubMed-Medline: 9 de marzo de 2009, 11:30 AM

Si comparamos el por ciento general que representa el total de revisiones existentes en la base de datos desde el año 2000 hasta la fecha (11.72) con respecto al grupo de interés (20.06), a simple vista, puede apreciarse la superioridad de este último. El por ciento general, hallado para el total de temas tratados en la base de datos, se tomó como valor de referencia para determinar si el grado de madurez, medido según el por ciento de revisiones existentes en el tema de interés era superior, igual o inferior al de la media de la base de datos. En este caso, el grado de madurez del tema objeto de estudio superó la media general del total de temas tratados en la base y por tanto presenta una madurez mayor a la media en PubMed-Medline.

Este parámetro es bastante confiable, sobre todo, si se considera el cubrimiento "tapa a tapa" de la inmensa mayoría de las fuentes procesadas por la base de datos y el alcance general de muchas de las revistas clínicas procesadas; así como los pocos cambios que experimenta esta base en su orientación temática más allá de la propia evolución del conocimiento.

Otro aspecto de interés puede ser el comportamiento del nivel de acceso al texto completo de los documentos referidos por la base de datos. Si realizamos una búsqueda marcando sólo la opción Limit to free full text, el sistema nos devuelve un total de 2 603 537 documentos (a las 14:00 horas de igual día) que cumplen con esta condición, es decir, del total de 18 702 298 de referencias, existe un 13.92 % (casi un 14 %) o poco más de 2 millones y medio de documentos que es posible acceder a texto completo.

Si quisiéramos saber cómo se comporta este parámetro en relación con el tiempo, podríamos realizar una exploración, donde se combinara fecha y nivel de acceso: desde 1980 a la fecha, existen 2 128 647 documentos que cumplen esta condición; desde 1990, 1 793 214 y desde el 2000, 1 160 383 (44.56 % del total con acceso libre a texto completo). Ello quiere decir que en la medida que el documento posee una fecha de edición más reciente es más probable que se encuentre su texto completo en forma libre y eso, es una buena noticia para todos. También es una evidencia de que el movimiento pro acceso abierto avanza. Pero, existen desigualdades con respecto a este parámetro, según casas editoras, temáticas y países. Y ellas se revelan cuando se comparan los valores medios generales en la base con los de ciertos conjuntos de interés.

Otro tanto sucede, si deseamos averiguar el nivel científico de la investigación realizada. Los ensayos clínicos controlados y controlados aleatorios se encuentran entre los diseños más sólidos de la investigación médica. Desde 1980 a la fecha, la base de datos ha procesado 311 900 de ellos; desde 1990, 262 663 y desde el 2000, 147 177. Y eso quiere decir que el 47.18 % (prácticamente la mitad) de los ensayos de este tipo procesados por PubMed-Medline desde 1980, lo fueron en los 8 años. Ello indica un avance en general de la experimentación, la profundidad y la calidad de la investigación médica. Pero esos valores sirven además como puntos de referencia para realizar comparaciones con los resultados de búsquedas en temas particulares como el expuesto al principio de nuestra contribución.

Son muchas las opciones y combinaciones probables para estos propósitos - PubMed-Medline permite limitar los documentos, según su nivel de acceso (al resumen o al texto completo), su fecha de publicación, el sexo y la edad de los

individuos estudiados, el idioma del documento, el tipo de artículo, etcétera- pero no es nuestro propósito analizarlas todas en esta carta sino sólo estimular el uso de las posibilidades que nos ofrece directamente -sin referirnos a herramientas como GoPubMed y PubReMiner- esta base de datos para responder interrogantes cualitativas en los campos de la investigación y la producción científica, a partir de exploraciones métricas, en el área de la biomedicina.

Pruebe usted mismo. Determine si la producción científica sobre el tratamiento quirúrgico del cáncer de mama es superior a la media de otros temas presentes en PubMed-Medline en el período 1990-2009.

Queda en sus manos explotar estas posibilidades con profundidad.

REFERENCIAS BIBLIOGRÁFICAS

1. NCBI. PubMed. PubMed Overview. Disponible en: <http://www.ncbi.nlm.nih.gov/entrez/query/static/overview.html> [Consultado: 6 de febrero de 2009].

Recibido: 9 de marzo de 2009.

Aprobado: 15 de marzo de 2009.

Lic. Rubén Cañedo Andalia. Departamento Fuentes y Servicios de Información. Centro Nacional de Información de Ciencias Médicas-Infomed. Calle 27 No. 110 e/ N y M, El Vedado. Plaza de la Revolución. Ciudad de La Habana. Cuba. Correo electrónico: ruben@infomed.sld.cu

Ficha de procesamiento

Términos sugeridos para la indización

Según DeCS¹

BIBLIOMETRÍA; BASES DE DATOS BIBLIOGRÁFICAS; PUBMED.

BIBLIOMETRICS; DATABASES, BIBLIOGRAPHIC; PUBMED.

Según DeCI²

INFORMETRÍA; BASES DE DATOS BIBLIOGRÁFICAS; MEDLINE.

INFORMETRICS; DATABASES, BIBLIOGRAPHIC; MEDLINE.

¹BIREME. Descriptores en Ciencias de la Salud (DeCS). Sao Paulo: BIREME, 2004.

Disponible en: <http://decs.bvs.br/E/homepagee.htm>

²Díaz del Campo S. Propuesta de términos para la indización en Ciencias de la Información. Descriptores en Ciencias de la Información (DeCI). Disponible en: <http://cis.sld.cu/E/tesauro.pdf>

Copyright: © ECIMED. Contribución de acceso abierto, distribuida bajo los términos de la Licencia Creative Commons Reconocimiento-No Comercial-Compartir Igual 2.0, que permite consultar, reproducir, distribuir, comunicar públicamente y utilizar los resultados del trabajo en la práctica, así como todos sus derivados, sin propósitos comerciales y con licencia idéntica, siempre que se cite adecuadamente el autor o los autores y su fuente original.

Cita (Vancouver): Cañedo Andalia R. Los límites de Pubmed como herramientas para la investigación métrica. Acimed 2009; 19(3). Disponible en: Dirección electrónica de la contribución. [Consultado: día/mes/año].