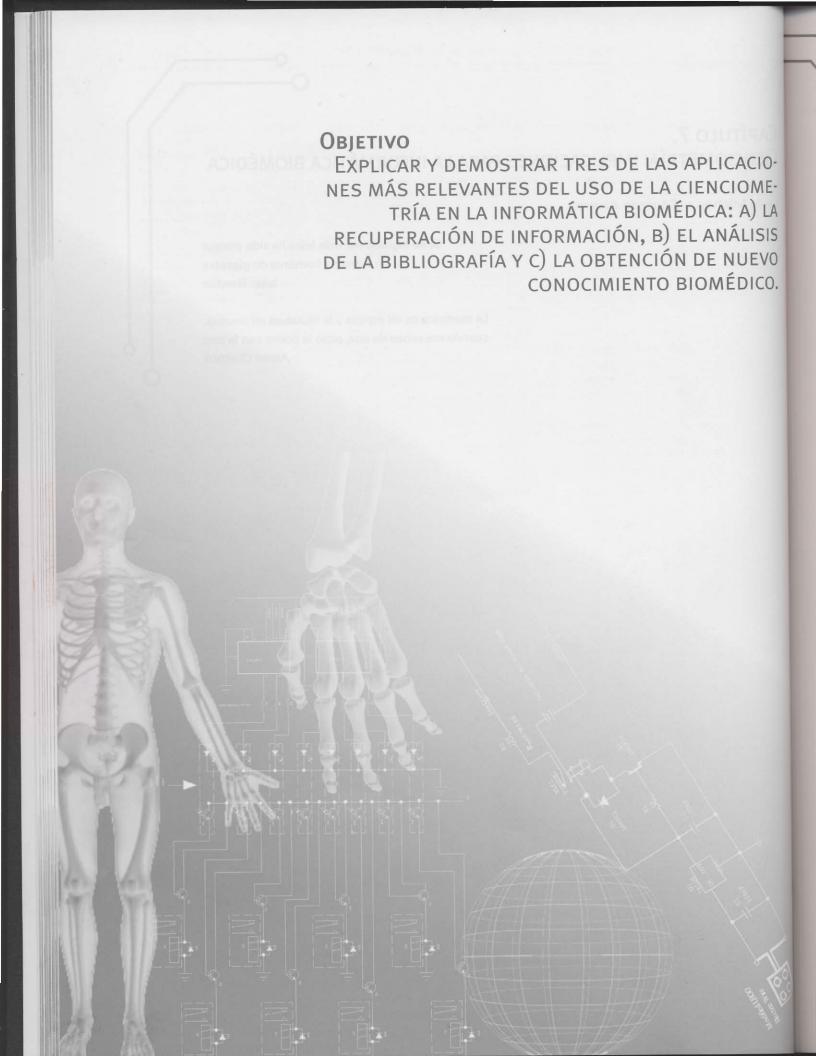
# CAPÍTULO 7. CIENCIOMETRÍA: IMPLICACIONES PARA LA INFORMÁTICA BIOMÉDICA

Layla Michán y Eduardo Alvarez

Si he logrado ver más lejos ha sido porque he subido a hombros de gigantes - Issac Newton

La medicina es mi esposa y la literatura mi amante, cuando me canso de una, paso la noche con la otra - Anton Chekhov



## Introducción

)

E-

IS

0

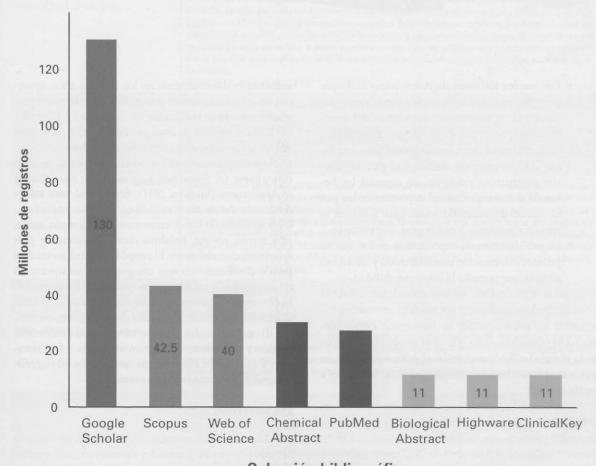
Э.

Una de las características distintivas de la informática biomédica es el uso de datos digitales tanto biomédicos como bibliográficos y su procesamiento *in silico*; sin embargo, los estudios de esta área que utilizan la bibliografía, sobre todo los artículos científicos, como objeto de análisis son escasos proporcionalmente; así, éste se ha convertido en uno de los rubros que resultan más novedosos e interesantes para la bioinformática, en especial por las siguientes razones:

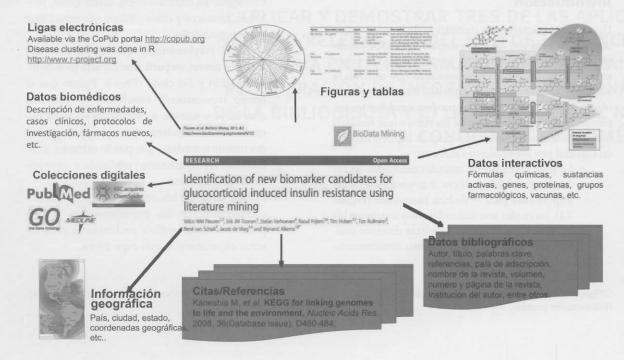
 La gran cantidad de artículos existentes en las colecciones bibliográficas internacionales más importantes para el ámbito biomédico (Figura 7.1), las cuales son utilizadas para recuperar bibliografía o incluso para realizar diversos análisis con el fin de generar nuevo conocimiento biomédico en distintos casos, como genes, proteínas, fármacos y casos clínicos, por mencionar los más frecuentes.

- El papel preponderante que los documentos médicos tienen, en particular los artículos de investigación y los casos clínicos. Puesto que el nuevo conocimiento biomédico se presenta, distribuye y somete a evaluación en la bibliografía, resulta evidente que es a partir de los documentos médicos que puede extraerse gran cantidad de información certificada y susceptible de corroborarse.
- El bajo costo de los análisis computacionales comparados con los experimentales, lo que permite realizar análisis preliminares para diseñar experimentos más específicos.

OFigura 7.1. Cantidad de registros bibliográficos en las colecciones digitales más relevantes a nivel internacional. (Elaboración propia).



OFigura 7.2. Ejemplos de metadatos relacionados con la bibliografía biomédica. Caso analizado en Fleuren et al., (2013). (Elaboración propia).



- Los nuevos formatos digitales, como RDF, que permiten realizar análisis más complejos a través de la representación semántica del conocimiento por medio de ontologías biomédicas.
- La factibilidad de procesar cantidades inmensas de información (terabytes) por la ciberinfraestructura existente, en especial las bases de datos especializadas y sofisticadas por el diseño de procedimientos que incluyen el uso de algoritmos y ontologías, por ejemplo.
- La proliferación de repositorios en los que se publica información estandarizada y de acceso abierto que permite la interoperabilidad.

El estudio cuantitativo de los metadatos sobre la bibliografía especializada, de manera particular los artículos de investigación, se denomina cienciometría (si se pone énfasis en que es el producto principal de la ciencia) o bibliometría (si el énfasis se pone en que son documentos); estos métodos se iniciaron a partir de 1960.

Una búsqueda en PubMed –la base de datos bibliográfica especializada en el área biomédica más grande de acceso público– con los términos scientometr\* OR bibliometr\* realizada el 5 de abril de 2013 arrojó más de 4 700 registros. Los tipos de trabajos que se obtienen han aumentado drásticamente en los últimos años, aparecen en revistas distintas y cumplen con diversos propósitos, desde la realización de revisiones bibliográficas hasta el análisis de citas y el factor de impacto para definir las revistas más sobresalientes en biomedicina.

Los detalles sobre el método bibliométrico pueden consultarse en *Biomedical Engineering – From Theory to Applications* (Michán, 2011). En general, este método consta del análisis cuantitativo de los metadatos bibliográficos de los documentos, por ejemplo, autor, año, temas, revista, palabras clave, institución y país de adscripción del autor. El propósito de estos estudios puede clasificarse en seis categorías: a) recuperación de información médica (principalmente bibliografía), b) obtención de nuevo conocimiento médico, c) revisiones bibliográficas y estados del arte sobre medicina, d) análisis de las ciencias médicas, e) evaluación, gestión y política científica para medicina, y f) generación y normalización de otras colecciones bibliográficas, por mencionar las más comunes.

### Desarrollo

Cuando se leen los artículos científicos, pocas veces se toma en cuenta la cantidad y diversidad de los metadatos que contienen y la información relacionada con ellos. Existen diferentes tipos de metadatos biomédicos vinculados de manera directa con la bibliografía susceptibles de ser analizados para obtener nuevo conocimiento biomédico (Figura 7.2). Cada una de estas identidades puede correlacionarse con otras a partir del uso de identificadores únicos como el de PubMed

(PMID) o DOI para bibliografía, el de GeneBank para

genes o el de UniProt para proteínas.

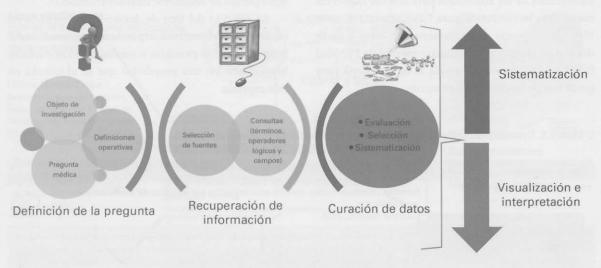
Entre todos los tipos de documentos biomédicos, los más utilizados son ocho (Tabla 7.1) y todos son susceptibles de analizarse bibliométricamente.

La elección del tipo de documento a emplear en un análisis bibliométrico depende en gran medida del propósito que se persigue; a continuación se explican brevemente los tres propósitos que se abarcarán en este capítulo.

OTabla 7.1. Tipos de documentos más utilizados en biomedicina. (Elaboración propia).

Tipo de documento	Descripción
Informe de casos (case report)	Presenta manifestaciones clínicas acompañadas por estudios de evaluación que dan lugar a un diagnóstico. Ejemplo: Negrini, 2008.
Artículo clásico (classical artícle)	Obra que consiste en una versión actual de un artículo impreso previamente que marcó un hito en la historia de la medicina o la ciencia. Por lo general va acompañado de comentarios introductorios que anuncian su reimpresión, como el aniversario de su publicación original o el aniversario del nacimiento o muerte del autor. Por lo general existe como reimpreso con la referencia bibliográfica completa a la fuente original.  Ejemplo: Glanville et al., 2011.
Ensayo clínico (clinical trial, phase I-IV)	Informe de un estudio clínico pre-planificado sobre la seguridad, eficacia o esquema de dosificación óptimo de uno o más fármacos de diagnóstico, terapéuticos o profilácticos, y de dispositivos o técnicas en humanos seleccionados de acuerdo con criterios predeterminados de elegibilidad y concluyendo los efectos favorables y desfavorables. Aunque la mayoría de los ensayos clínicos se refiere a seres humanos, este tipo de publicación puede usarse también en la práctica veterinaria, en la que los artículos satisfacen algunos requisitos para los seres humanos.  Ejemplo: Turroni et al., 2010.
Estudio comparativo (comparative study)	Comparación de los resultados, respuestas, etc., para diferentes técnicas, enfoques terapéuticos u otros insumos.  Ejemplo: Ponterotto et al., 2012respectively, as having either a profound, highly significant, significant, or important impact on the academic multicultural training of counseling graduate students. The top-cited scholars across the textbooks were also examined in relation to their scholarly productivity (number of publications.
Ensayo clínico controlado (controlled clinical trial)	Ensayo clínico que incluye uno o más tratamientos de prueba y por lo menos un tratamiento de control, y en el que se especifican las medidas de resultado para evaluar la intervención estudiada y un método libre de prejuicios para asignar a los pacientes el tratamiento de prueba. Los tratamientos pueden ser medicamentos, dispositivos o procedimientos estudiados en cuanto a eficacia diagnóstica, terapéutica o profiláctica. Las medidas de control incluyen placebos, medicina activa, sin tratamiento, formas de dosificación y regimenes, y comparaciones históricas, entre otros. Pueden agregarse métodos aleatorios para la asignación de pacientes. Ejemplo: Ortega et al., 2009.
Casos legales (legal cases)	Informes sobre leyes o casos resueltos y documentos o presentaciones relacionadas con esos casos.  Ejemplo: Birenbaum et al., 2006.
Meta-análisis (meta-analysis)	Estudios que utilizan un método cuantitativo de una combinación de los resultados provenientes de estudios independientes (por lo general extraídos de la bibliografía publicada) y la síntesis de resúmenes y conclusiones que pueden usarse para evaluar la eficacia terapéutica, planear nuevos estudios, etc. A menudo son una visión general de ensayos clínicos. Ejemplo: Taylor et al, 2012.
Revisión (review)	Trata sobre un artículo o libro publicado después del examen del material sobre un tema específico. Puede ser exhaustivo en distintos grados y el rango de tiempo del material analizado puede ser general o particular pero con mayor énfasis sobre la bibliografía actual. El material textual también puede ser amplio y abarcar, en la medicina específicamente, casos clínicos, así como informes de investigación experimental.  Ejemplo: Pennell et al., 2012.

OFigura 7.3. Esquema que resume los pasos a seguir en el proceso de recuperación de literatura. (Elaboración propia.)



# La recuperación de información

Recuperación de bibliografía se refiere a la búsqueda y obtención de información, comportamientos y procesos que la consulta conlleva; su función es recuperar bibliografía exhaustiva, precisa, relevante y pertinente de un tema específico para analizarla, difundirla, compartirla, sistematizarla, guardarla, exponerla, presentarla o compararla con el fin de resolver un problema médico determinado.

Para recuperar información es necesario tener una pregunta o problema claro al cual se enfocará la investigación. Este proceso va de lo general a lo particular y consiste en obtener la mayor cantidad de información disponible. Para realizar este procedimiento deben seguirse las siguientes estrategias: a) establecer claramente la pregunta y las disciplinas en las que se enmarca; b) identificar los términos de búsqueda (deseados, no deseados, sinónimos y homónimos) con los que se formulará la consulta (tomando como referencia datos como título, resumen, palabras clave, referencias e imágenes) para obtener un indicador de inclusión/exclusión de bibliografía buscada; c) diseñar las consultas con los términos elegidos, los operadores lógicos o booleanos; d) definir los campos de búsqueda de la colección; e) refinar la consulta tantas veces como sea necesario hasta encontrar resultados completos y exactos; f) producir definiciones operativas que permitan tomar decisiones respecto a la bibliografía que se seleccionará; g) llevar a cabo el proceso de búsqueda, selección y evaluación (este procedimiento varía según la base de datos

consultada) en distintas colecciones bibliográficas; y h) sistematizar los registros seleccionados, depurarlos y curarlos (Figura 7.3). A continuación presentamos ejemplos de artículos de este tipo.

Cabezas-Clavijo y Delgado-López-Cózar (2013) redactaron las características de las tres bases de datos bibliográficas más extensas y utilizadas, y compararon el procedimiento para extraer información entre ellas: Google Scholar, Web of Science y Scopus; también compararon el uso del índice h para calcular el rendimiento científico en el área de medicina intensiva.

Chen et al. (2013) describen un método semántico para el procesamiento de lenguaje natural, métodos de indexación y recuperación, análisis estructural e integración de bibliografía a través de los metadatos asociados con el cuerpo del documento digital con un método llamado Latent Semantic Indexing (LSI).

Li et al. (2012) evaluaron la productividad sobre revistas de medicina de cuidados intensivos, destacando las revistas más productivas según los datos del Journal Citation Reports® (JCR) y los países más productivos de 2006 a 2010.

En ese mismo año, Tobias Kuhn y colaboradores documentaron el procedimiento para buscar imágenes y relacionarlas con el texto biomédico utilizando PubMed e implementando la técnica de procesamiento digital de imágenes y algoritmos específicos para encontrar las imágenes más relevantes en una investigación completa, es decir, teórico-experimental.



## El análisis de la bibliografía

Este propósito se refiere a documentos que utilizan la bibliometría para presentar estados del arte o incluso obtener tendencias de investigación, estructuras y relaciones complejas a partir de los metadatos bibliográficos; entender las características de un conjunto de bibliografía y obtener el comportamiento de la publicación en el tiempo o por temas; y reconocer las revistas y los autores que más publican u obtener una red de colaboración, por ejemplo.

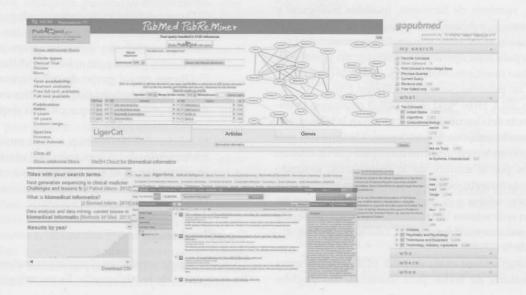
Estos análisis son factibles de realizar en la investigación sobre biomedicina a partir de los artículos de investigación o revisión. Se usan principalmente los datos bibliográficos que suelen estar disponibles en las colecciones bibliográficas, como autores, revistas, palabras clave, año de publicación, descriptores, citas bibliográficas, institución de adscripción del autor, país de adscripción y resumen, en la medicina basada en evidencia a partir de los ensayos clínicos. Las variables a considerar en este caso podían ser enfermedades, fármacos, genes, proteínas, estadísticos, reportes de hospitales, secuencias, cultivos celulares extraídos de la bibliografía, entre muchas otras. A continuación presentamos algunos ejemplos de estos estudios cienciométricos.

Falagas et al. (2008) se preguntaron a qué edad son más productivos los investigadores y analizando metadatos bibliográficos se dieron a la tarea de responder esta cuestión sobresaliente. Destacaron que el rango de edad en el que los investigadores son más productivos va de 30 a 50 años y que a partir de este último año comienza el declive en cuanto a su producción.

Ramos et al. (2013) realizaron un análisis bibliométrico sobre la investigación en leishmaniosis empleando los registros de Medline desde 1945 hasta 2010. Presentaron los resultados en un gráfico con la tendencia de publicaciones por año en el que se observa que la investigación aumentó al paso de los años; enlistaron las 50 revistas más importantes en los 60 años; presentaron también un análisis de las palabras indexadas en MeSH y su relación con el número de documentos procesados, y los 30 países con más artículos relacionados con la enfermedad estudiada; y finalmente, elaboraron una tabla en la que desglosan a los autores más productivos y una categorización por la forma de la enfermedad. Éste es un buen ejemplo de cómo se analiza la bibliografía de una base de datos especializada como Medline en el área biomédica.

También existen aplicaciones Web que nos permiten recuperar información y analizar tendencias de la bibliografía en el área biomédica en tiempo real de manera amigable, muchas de las cuales utilizan la bibliografía de PubMed. Tal es el caso de PubMed PubReminer (www.hgserver2.amc.nl/cgi-bin/miner/miner2.cgi), LigerCat (www.ligercat.ubio.org), PubMed con su menú para refinar consultas (www.ncbi.nlm.nih.gov/pubmed), ClinicalKey (www.clinicalkey.com) y GoPubMed (www.gopubmed.org/web/gopubmed) por mencionar algunos (Figura 7.4).

OFigura 7.4. Esquematización de aplicaciones Web amigables utilizadas para recuperar información, procesarla y visualizarla. (Elaboración de los autores).





#### La obtención de nuevo conocimiento

El proceso más complejo y ambicioso de un análisis bibliométrico consiste en la obtención de nuevo conocimiento generado a partir del procesamiento de grandes cantidades de bibliografía biomédica especializada. Descubrimiento basado en bibliografía (literature-based discovery en inglés), se refiere al uso de artículos de investigación para encontrar nuevas relaciones entre el conocimiento existente. Está técnica fue desarrollada por Don R. Swanson en la década de 1980 utilizando bibliografía médica y desde entonces su uso se ha generalizado y su impacto en la biomedicina ha sido muy importante. Sin lugar a dudas, es en esta disciplina donde se desarrollan los avances y las innovaciones más importantes de este método. El procedimiento incluye, además de la bibliometría, la implementación de métodos de vanguardia como la minería de textos; se analizan principalmente los metadatos bibliográficos, como palabras clave, descriptores, resumen, tablas y figuras, y en los casos más ambiciosos, el texto completo del documento.

El proceso abarca un gran número de pasos que dependen del tipo de información que se analice y los resultados se emplean en la generación de nuevas hipótesis para validar datos empíricos, curar información y generar correlaciones nuevas entre los datos ya existentes. Este método es fundamental para el campo de la medicina basada en evidencia porque la cienciometría aplicada a los casos clínicos recuperados para la obtención de nuevo conocimiento es un método eficiente y útil para lograr conclusiones más precisas.

En 2012, Younesi et al. (2012), a través de la minería de textos, clasificaron y marcaron diferentes componentes de los artículos biomédicos como proteínas, genes, nombres de fármacos, modelos biológicos y diagnósticos médicos, entre otros, para identificar la correlación entre ellos, desarrollar un motor de búsqueda para la recuperación de bibliografía y proponer y evaluar nuevos biomarcadores mediante el significado de las palabras (semántica) (Figura 7.5).

En 2013, Hristovski et al. (2013) recalcaron la importancia del descubrimiento basado en la bibliografía analizando artículos indexados en Medline con una hipótesis en mente: ligar tres conceptos que no estaban ligados por sí mismos en la bibliografía, tomando en cuenta tres categorías: a) enfermedades, b) patologías / función celular y c) drogas, genes, etc. Con ese fin usaron aplicaciones Web para relacionar los conceptos generando una red interactiva, con base en el significado de la palabra (ontología).

Hongyu Chen et al., en 2013, utilizaron la técnica de minería de textos para identificar las posibles relaciones entre los conceptos de los artículos biomédicos analizados con Latent Semantic Indexing (LSI).

## Actividades individuales y grupales

De manera individual ingresa a PubMed, localiza tres artículos del área biomédica que más te gusten y que utilicen la cienciometría [p. ej., (scientometr\* OR bibliometr\*) AND (cancer OR oncology)], y resuelve lo siguiente:

- a) Utiliza la Tabla 7.1 para clasificar los tipos de documentos analizados en el estudio.
- b) Con base en los tres objetivos del libro, ¿qué enfoque cienciométrico presenta el artículo que estás analizando?
- c) Usa la Figura 7.2 como referencia para enumerar el tipo de metadatos asociados que se utilizaron en el estudio.
- d) ¿Explica en qué sentidos se aplica la cienciometría en ese artículo biomédico?
- e) ¿De qué manera te ayuda este tema en tu formación académica?

De manera grupal discute y contesta lo siguiente; puedes ayudarte de las aplicaciones Web descritas en este capítulo.

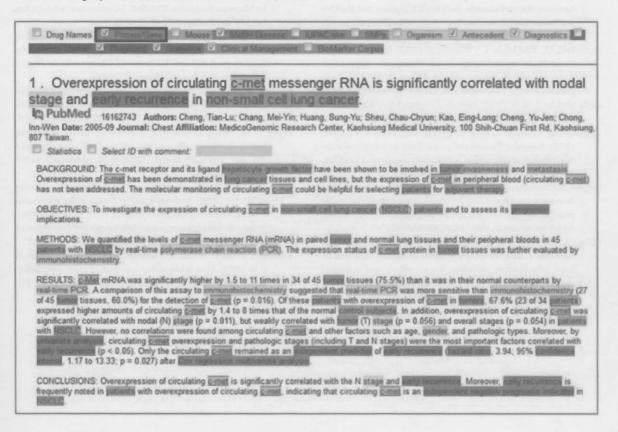
- 1) Utilizando LigerCat realiza la consulta gastroenterology y contesta:
- a) ¿Cómo refinas la consulta de tal manera que los resultados que deseas obtener sean estudios en hombre, mujer y neoplasia estomacal? b) ¿Cuál es el autor más productivo en esta
- área?
- c) ¿En qué años se publicaron más artículos?
- Realiza la misma consulta en ClinicalKey y discute las diferencias entre los resultados que arroja PubMed.
- 3) Da una definición de gastroenterology.
- 4) Si quisiéramos saber cuáles son los países más representativos y generar un mapa presentando estos países, ¿qué aplicación utilizarías?
- 5) Discute con el grupo de qué manera utilizarías la cienciometría en la práctica médica; diseñen una pregunta que puedan responder usando este método.
- 6) Adicionalmente podrían realizar un análisis bibliométrico de un área, enfermedad o técnica en particular en el que estudien una variable como autores, revistas, años, instituciones, etc. Discútanlo en grupo.

#### Conclusiones

La cienciometría consiste en el análisis cuantitativo de los metadatos de la bibliografía. Es un método poco



OFigura 7.5. Ejemplo de proceso de información y generación de nuevo conocimiento biomédico utilizando el procesamiento de lenguaje natural. Tomada de Younesi et al., 2012.



utilizado en la informática médica que representa un gran potencial de apoyo en la práctica médica para recuperar información, evaluarla, compartirla, sistematizarla, analizarla, procesarla, visualizarla e interpretarla, de manera que para cualquier estudiante, practicante, profesor y/o investigador en ciencias médicas siempre será útil conocerlo y aplicarlo con la finalidad de estar actualizado en cuanto a la información más relevante en su área a estudiar, generar una colección de bibliografía digital, analizarla y difundir los resultados o aun como un método alternativo para generar nuevo conocimiento biomédico. Así, las principales aplicaciones de la cienciometría en la informática biomédica son la recuperación de información, el análisis de la bibliografía y la obtención de nuevo conocimiento biomédico.

Incluso puede ser parte de una especialidad en bibliografía e informática médica que comprenda todos los aspectos relacionados con la generación, estructura, flujo, manejo, mantenimiento, curación, procesamiento y análisis de bibliografía, y aplique la cienciometría como método de análisis para la resolución de problemas médicos a través de la bibliografía especializada.

#### Referencias

- Birenbaum-Carmeli, D., Banerjee, A. y Taylor, S. (2006). All in the family: media presentations of family assisted suicide in Britain. Soc Sci Med, 63(8):2153-64. doi:10.1016/j.socscimed.2006.05.004.
- Cabezas-Clavijo, A. y Delgado-López-Cózar, E. (2013). Google Scholar and the h-index in biomedicine: The popularization of bibliometric assessment. Medicina intensiva. doi:10.1016/j.medin.2013.01.008.
- Chen, H., Martin, B., Daimon, C. M. y Maudsley, S. (2013). Effective use of latent semantic indexing and computational linguistics in biological and biomedical applications. Frontiers in Physiology, 4:8. doi:10.3389/fphys.2013.00008.
- Falagas, M. E., Ierodiakonou, V. y Alexiou, V. G. (2008). At what age do biomedical scientists do their best work? FASEB J, 22(12):4067-70. doi:10.1096/fj.08-117606.
- Fleuren, W. W., Toonen, E. J., Verhoeven, S., Frijters, R., Hulsen, T., Rullmann, T. et al. (2013). Identification of new biomarker candidates for glucocorticoid induced insulin resistance using literature mining. BioData mining, 6(1):2. doi:10.1186/1756-0381-6-2.

- Glanville, J., Kendrick, T., McNally, R., Campbell, J. y Hobbs, F. D. R. (2011). Research output on primary care in Australia, Canada, Germany, the Netherlands, the United Kingdom, and the United States: bibliometric analysis. BMJ, 342:d1028 Recuperado de http://www.bmj.com/content/342/bmj.d1028.
- Hristovski, D., Rindflesch, T. y Peterlin, B. (2013). Using Literature-based Discovery to Identify Novel Therapeutic Approaches. Cardiovascular Hematol Agents Med Chem, 11(1):14–24. Recuperado de http://www.ncbi.nlm.nih.gov/pubmed/22845900.
- Layla Michán, I. M. V. (2011). Biomedical Engineering From Theory to Applications. En: R. Fazel (ed.). InTech. doi:10.5772/2629
- Li, Z., Qiu, L. X., Wu, F. X., Yang, L. Q., Sun, Y.M., Lu, Z.-J. y Yu, W.F. (2012). Assessing the national productivity in subspecialty critical care medicine journals: A bibliometric analysis. J Crit Care, 27(6):747.e1-5. doi:10.1016/j.jcrc.2012.03.002.
- Negrini, S. (2008). Approach to scoliosis changed due to causes other than evidence: patients call for conservative (rehabilitation) experts to join in team orthopedic surgeons. Disability Rehab, 30(10):731-41. doi:10.1080/09638280801889485.
- Ortega, E., García, J. J., Bote, M. E., Martín-Cordero, L., Escalante, Y., Saavedra, J. M. et al. (2009). Exercise in fibromyalgia and related inflammatory disorders: known effects and unknown chances. Exerc Immunol Rev, 15:42–65. Recuperado de http:// www.ncbi.nlm.nih.gov/pubmed/19957871.

- Pennell, D. J., Carpenter, J. P., Firmin, D. N., Kilner, P. J., Mohiaddin, R. H. y Prasad, S. K. (2012). Review of Journal of Cardiovascular Magnetic Resonance 2011. BMC J Cardiovascular Magn Res, 14(1):78. doi:10.1186/1532-429X-14-78.
- Ponterotto, J. G., Fingerhut, E. C. y McGuinness, R. (2012). Legends of the field: influential scholars in multicultural counseling. Psychological Reports, 111(2):364-82. Recuperado de http:// www.ncbi.nlm.nih.gov/pubmed/23234083.
- Ramos, J. M., González-Alcaide, G. y Bolaños-Pizarro, M. (2013). Bibliometric analysis of leishmaniasis research in Medline (1945-2010). Parasites & Vectors, 6(1):55. doi:10.1186/1756-3305-6-55.
- Taylor, D. M., Smith, L., Gee, S. H. y Nielsen, J. (2012). Augmentation of clozapine with a second antipsychotic a meta-analysis. Act Psychiatr Scand, 125(1):15-24. doi:10.1111/j.1600-0447.2011.01792.x.
- Turroni, S., Vitali, B., Candela, M., Gionchetti, P., Rizzello, F., Campieri, M. et al. (2010). Antibiotics and probiotics in chronic pouchitis: a comparative proteomic approach. WJG, 16(1):30-41. Recuperado de http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2799914&tool=pmcentrez&rendertype=abstract.
- Younesi, E., Toldo, L., Müller, B., Friedrich, C. M., Novac, N., Scheer, A. et al. (2012). Mining biomarker information in biomedical literature. BMC Med Inform Dec Mak, 12(1):148. doi:10.1186/1472-6947-12-148.