

Vigilancia y evaluación de la actividad científico-tecnológica

Gilberto Sotolongo Aguilar
María Victoria Guzmán Sánchez
Ignacio García Díaz
Elias Sanz Casado

Instituto Finlay, La Habana, Cuba.
Instituto Finlay, La Habana, Cuba.
Instituto de Documentación e Información Científica, La Habana, Cuba.
Universidad Carlos III de Madrid, Madrid, España.

Resumen

El gerente de un sistema de información identifica fuentes y recopila un importante volumen de datos, su tarea consiste en organizarlos de un modo intelegible. La bibliometría es uno de los modelos de análisis que nos ayudarían a solucionar la problemática anterior. En este trabajo se examinan métodos y herramientas de la bibliometría y disciplinas afines. Se argumenta sobre las diferentes aplicaciones de sus indicadores, como los empleados para evaluar programas de investigación y desarrollo, el comportamiento de la innovación tecnológica, la capacidad de una compañía para determinar la estructura y movimiento de los mercados, entre otros.

Abstract

The manager of an information system has identified sources and collected an important data volume. His task consists of organizing the data in an intelligible way. Bibliometry is one of the analysis models which would help him solve the problem. In this work the authors examine bibliometric methods and tools. Different uses of indicators such as the ones used to evaluate research and development programs, technological innovation behaviour and a firm's ability, or to assess a market's structure and movement, among others, are discussed here.

■ Introducción

La bibliometría, en su desarrollo, ha devenido en disciplina teórica e instrumental. Esta última cualidad obedece a las posibilidades de las herramientas con ella asociadas. Sin embargo, pese al desarrollo que ha tenido la bibliometría aún hoy en día, ésta no se utiliza en toda su magnitud; probablemente debido al desconocimiento que se tiene sobre todas sus probabilidades. Reducir las posibilidades de la bibliometría a la modelación bibliográfica como es el análisis de la dispersión de las fuentes, la productividad de los autores o la obsolescencia de la literatura, por sólo citar algunas de sus aplicaciones más conocidas, es condenarla a una suerte de subdesarrollo disciplinar. Al propio tiempo, aplicar sus modelos exclusivamente a la selección y adquisición de la literatura es limitar sus potencialidades. Sin embargo, los límites de esta ciencia aún no han sido encontrados.

Los índices asociados con la bibliometría devienen en indicadores del desarrollo científico y tecnológico pasado, presente y futuro. Sin embargo, absolutizar su valía no es menos que un crimen de *lesa científicidad*. Estos indicadores han evolucionado desde los univariados hasta al menos dos generaciones de indicadores relacionados. Estos últimos, tales como los que se derivan del análisis de *cluster* y el escalado multidimensional, han enriquecido las posibilidades de esta disciplina.

De la sinergia entre la bibliometría y su prima-hermana, la cienciometría, ha surgido la informetría como disciplina



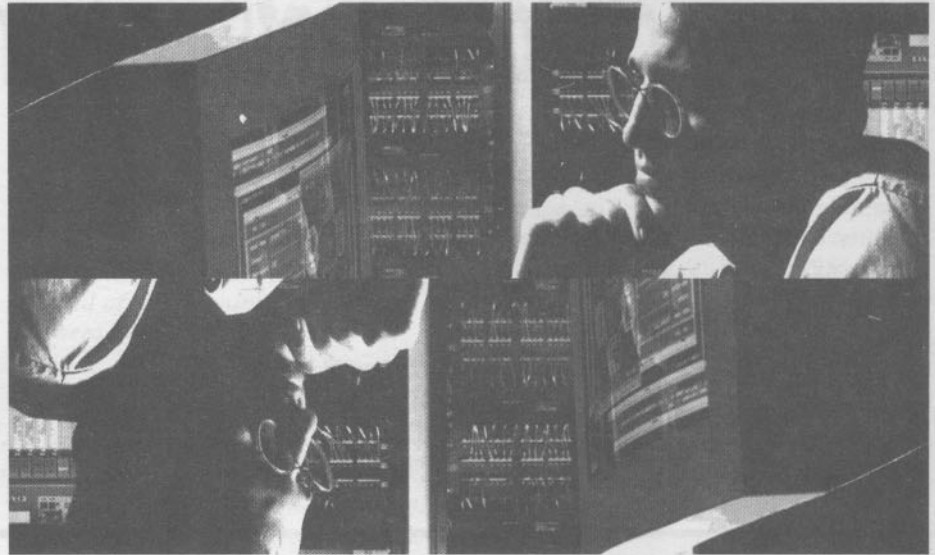
integradora donde se respetan mutuamente la individualidades de sus elementos integrantes y cuyo amalgamiento permite abordar nuevas y más complejas tareas, en los tiempos en que no es precisamente la complejidad de los problemas lo que falta.

En la praxis bibliométrica, la gestión de la información en las organizaciones (*the information management*), es uno de sus modelos de ejecutoría. Baste señalar, a guisa de ejemplo, el valor agregado de que es posible dotar a los servicios y productos informativos en el ambiente, de un sistema enfocado al monitoreo y análisis de la información científica y tecnológica.

Los especialistas en gestión de la información y en informetría han llegado por distintas vías a darle solución a la necesidad y disponer de la información requerida, en el momento oportuno y de manera adecuada. Buscan la posibilidad de obtener la ventaja competitiva que reclaman las empresas (no importa el fin de éstas) abordando la problemática de lo que se ha dado en llamar inteligencia competitiva o incluso iteligencia económica.

Presumiblemente lo más atractivo de este tema es precisamente el hecho de que la realidad que nos rodea es al mismo tiempo infinita y diversa (parafraceando a Dostoiévski), e igual infinitud y diversidad tiene el reflejo de una parte nada despreciable de esta diversidad constituida por las publicaciones científicas, en particular los dos tipos más paradigmáticos: las publicaciones seriadas y las patentes. Entidades éstas que surgen constantemente, de su carácter discreto se deriva, la cualidad de su continuidad.

Los retos que atribuimos a la bibliometría: la vigilancia científico-tecnológica, por un lado, y por el otro, la evaluación de este tipo de actividad (también la científico-tecnológica), al parecer no son más que dos tipos de actividades similares cuya diferencia ra-



dica en la periodicidad de su realización. La evaluación de la actividad científico-tecnológica se realiza puntualmente para un periodo de tiempo determinado.

Si esta evaluación llega a ser sistemática entonces se convierte en vigilancia tecnológica. Claro está que esta simplificación puede llevar a confusión si no se predefine con claridad el propósito que se persigue; no obstante, en esencia su similitud y sus diferencias son claros, en tanto su comunidad de proceder, herramientas y tecnología de instrumentación no difieren substancialmente. Al mismo tiempo, virtualmente los índices necesarios para determinar los indicadores de desempeño de uno y otro caso son prácticamente idénticos, sólo variarán los tipos de documentos bajo estudio, y los umbrales de estos indicadores, ya que obviamente el lapso de tiempo y las temáticas estudiadas variarán según sea el caso.

■ Sistema estructurado para los estudios bibliométricos

En los últimos cuatro años hemos venido trabajando en el ámbito del análi-

sis cuantitativo de la literatura y hemos logrado un sistema estructurado para estos fines. La idea que subyace consiste en conseguir un sistema modular abierto que permita incorporar nuevos desarrollos sin desarticular la esencia misma del sistema.

Desde el punto de vista metodológico, este sistema estructurado puede instrumentarse en las etapas siguientes:

1. Búsquedas bibliográficas.
2. Tratamiento (conversión) del fichero resultante.
3. Creación de la base de datos (B-B) bibliográfica (Bibliométrica).
4. Normalización de la base de datos B-B.
5. Creación de los ficheros de los campos bajo estudio (o de una combinación de estos).
6. Tratamiento estadístico (bibliométrico).
7. Creación de la bibliografía y sus índices (opcional).

Las siete etapas presentadas anteriormente, constituyen una visión simplificada del procedimiento para la implantación del sistema estructurado al que se ha hecho referencia, no obstante, a continuación hacemos algunos comentarios y premisas necesarios.



■ Algunas premisas y condiciones necesarias

El fin que persiguen los estudios bibliométricos no es otro que el de la elaboración de hipótesis de trabajo científicamente fundamentadas, que constituyan a su vez puntos de partida de otros especialistas que, como pares, tendrán ante sí la tarea de corroborarlos, rechazarlos o tomarlos en consideración para nuevos y más detallados análisis.

Dos de las premisas que permiten abordar empeños como los descritos anteriormente, son los conceptos de Proceso de Producción de Información (PPI) y el concepto de Orden Jerárquico (rangos).

En el concepto de Proceso de Producción de Información (PPI) intervienen dos tipos de entidades: las fuentes y los *ítems* que son producidos por esas fuentes. Como ejemplos tenemos a los autores que producen publicaciones; revistas que producen artículos; artículos que producen referencias, etcétera. El otro concepto de Orden Jerárquico tiene que ver con el rango de las fuentes en un PPI: la fuente más productiva ocupará el primer lugar en la jerarquía (rango=1); la segunda ocupará el rango=2, y así sucesivamente, hasta que a la última fuente más productiva le corresponderá el rango=T.

Las condiciones necesarias generalmente aceptadas para la realización de análisis cuantitativos de la literatura, son las siguientes:

- El tema de la bibliografía en estudio debe estar definido con claridad.

- La bibliografía debe ser lo más completa posible.
- La bibliografía debe limitarse a un periodo de tiempo en que todas las fuentes tengan la misma oportunidad de contribuir *ítems*.

■ Reflexiones

Partimos del supuesto de que este análisis se realiza sobre la base de búsquedas bibliográficas, con medios automatizados, en multimedia o en línea. No se descarta la idea de que la información original aparezca en otros soportes, lo que haría necesario llevarla a soportes magnéticos para su ulterior tratamiento.

La cadena de tratamiento que se seleccionó es de vital relevancia; de ahí la importancia de una adecuada selección del *software* a emplear. Como se verá posteriormente, debido a la necesidad de un máximo de normalización de los campos y de los elementos de cada registro, resulta de suma trascendencia la posibilidad de tratar con mucha flexibilidad tales elementos, con vistas a modificarlos a conveniencia.

Nuestra práctica nos ha llevado a utilizar como *software* fundamentales, para convertir y crear bases de datos, el *BiblioLink II* y el *ProCit* (en sus versiones para MS-DOS y Windows). Ambos permiten un tratamiento prácticamente universal a partir de búsquedas realizadas en cualquier servicio de información en línea o CD-ROM. Otra opción sería la combinación de *COMVI* con otros sistemas de gestión de bases de datos bibliográficas tales como *MICRO-ISIS*, *EndNotes*, *Papirus*, *Reference manager*, etcétera, por citar algunos.

No existe prácticamente ninguna base de datos perfecta, por lo que se hace necesario el proceso de normalización de ésta. Se trata, pues, por ejemplo, de eliminar espacios inconvenientes, como los existentes al inicio de una palabra o un descriptor, o en medio de un índice de clasificación. También ortografías diferentes para las mismas fuentes o autores. Con frecuencia resulta necesario agrupar algún otro elemento a los registros o alguna selección de éstos para lograr distinguirlos de alguna forma, como es el caso de la determinación de la pertenencia por un *cluster* u otro.





Un instrumento imprescindible para lograr lo anterior, es que los sistemas de gestión de información bibliográfica que se utilicen (e.g. Pro-Cite) permiten obtener *listas autorizadas* de los elementos en los campos, así como posibilidades de realizar cambios globales en los diferentes campos mediante la sustitución o eliminación de elementos.

Una vez realizado lo anterior, la base de datos normalizada está lista para su tratamiento estadístico (bibliométrico).

Obviamente, de nuevo la selección del *software* a utilizar resulta importante. En general, un buen aliado de este trabajo y de otras aplicaciones asociadas con la preparación de los ficheros de trabajo, lo constituye la familia de programas *Microsoft Office*. En particular *Word*, para el tratamiento de textos, y *Excel* para el procesamiento estadístico preliminar. Estos programas son prácticamente insustituibles.

La exportación de ficheros del tipo texto, con el contenido de los diferentes campos para su tratamiento estadístico, es una función que realiza muy eficientemente Pro-Cite. Con Excel y más particularmente con las herramientas de tablas dinámicas que éste posee, así como las herramientas de análisis de datos, se pueden realizar muy rápidamente estudios trascendentes desde el punto de vista bibliométrico.

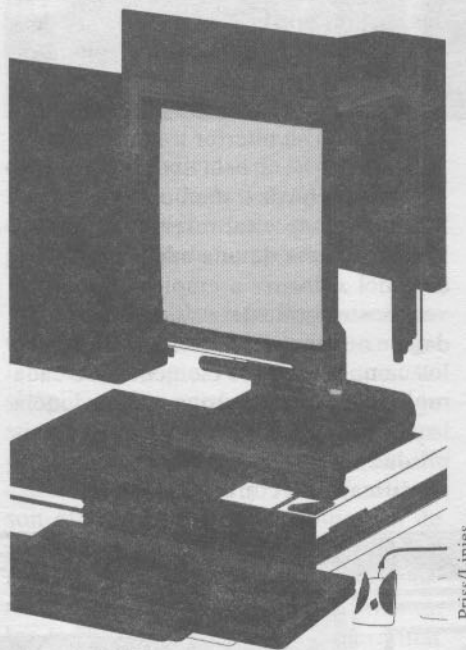
Sin embargo, existen herramientas específicas que pueden por sí mismas calcular las distribuciones bibliométricas fundamentales (Bradford, Lotka, Zipf) sin mucho esfuerzo y con ello completar o corroborar los hallazgos que se obtengan con *Excel*. Nuestra práctica la basamos para estos fines, en el *software The Bibliometrics Tool-box*, un *free-ware* cuyo desempeño está ampliamente probado.

Algunos de los índices que se pueden obtener con este programa son los siguientes:

1. Distribución por rangos según Bradford.

Esto incluye además, el *bibliógrafo* correspondiente, así como los coeficientes de regresión del *bibliógrafo* (el de los datos primarios y el de los datos ajustados). También se ofrece el coeficiente de determinación de ambas regresiones como indicativo del grado de ajuste de los modelos de regresión lineal.

2. El listado completo por rangos.



3. Diferentes particiones cohortes.

Se obtienen las cohortes de artículos que producen multiplicadores perfectos y sus residuos como guías para realizar las peticiones de la literatura. También se obtiene la cohorte mínima según Goffman/Warren; el mínimo de zonas estimado según Egghe. Todo ello para estudiar las diferentes particiones. Al propio tiempo, al seleccionar la partición deseada se obtiene el *índice de agrupamiento (Clustering Index)* que expresa la intensidad de agrupamiento de la literatura; la *productividad recu-*

rrente promedio (Average Recurring Productivity) que es la producción promedio de las revistas recurrentes. Se obtienen además dos índices de concentración, el de Pratt y el de Egghe C prima, que expresan la distribución equitativa de los artículos en las zonas creadas. Finalmente, es posible obtener la lista completa de las revistas en el núcleo.

4. Estimado del tamaño de la literatura.

Calcula el número de revistas y de artículos en la bibliografía completa ya que se parte del hecho de estar trabajando con una muestra de la literatura. Aquí se incluye la distribución de rangos y predicciones según los cálculos de la fórmula de Leimkhuler.

5. Distribución Lotka. Incluye el cálculo de la mayor diferencia entre el número de autores y el estimado, para calcular la prueba de bondad de ajuste según Kolmogorov-Smirnov.

6. Distribución de Zipf.

Presenta la distribución de rangos según Zipf, e incluye la curva correspondiente mostrando la relación entre el logaritmo de las frecuencias y el logaritmo de los rangos, así como el coeficiente de regresión correspondiente y el coeficiente de determinación.

7. Coeficiente de dispersión de referencias.

Éste es calculado como una alternativa al propuesto por Bradford. Fue sugerido por P. F. Cole. También ofrece la curva correspondiente, el coeficiente de correlación y el coeficiente de determinación asociado.

Mediante los índices referidos anteriormente es posible obtener una caracterización preliminar, pero bien acabada, de la bibliografía en estudio. Las limitaciones en tamaño sólo esta-



rían dadas por la memoria de la computadora disponible.

Existen otros muchos indicadores univariados de uso intensivo y extensivo mediante el simple cómputo de las publicaciones, sin embargo, no quisiéramos pasar por alto el conocido como índice de actividad, que permite caracterizar el esfuerzo relativo de un país en cierto campo de investigación, según se refleja en las publicaciones que produce (cociente que resulta del porcentaje de publicaciones del país en el campo de investigación, dividido entre el porcentaje mundial de publicaciones del país en todas las disciplinas). Es posible encontrar otras variantes de este índice al considerar campos y sub-campos de investigación.

Otro grupo de índices, los bivariados, ocupan un lugar muy importante y su utilización ofrece resultados muy interesantes. En particular mencionaremos aquellos en los que interviene el tiempo como una variable. Estos índices permiten explorar el proceso de producción de información de diferentes elementos bibliográficos de forma sincrónica y nos muestra, con relativa facilidad, tendencias en el tiempo.

Para concluir con los univariados y bivariados, sólo mencionaremos los índices que se derivan de tan controvertido análisis de citas, entre los que se destacan, el factor del impacto y el índice de inmediatez.

Concebidos para seguir la dinámica de las interacciones múltiples que se desarrollan en el seno de la investigación científica y técnica, surgen los índices relacionales. Éstos se agrupan en

dos familias: una primera generación, que no aborda directamente el contenido del documento bajo análisis. Mientras que el objetivo de los indicadores de segunda generación es precisamente el análisis del contenido de los documentos. Entre los índices relacionales de la primera generación, tenemos, entre otros: las firmas conjuntas de artículos, las redes de citas, las citas de una revista en otra y las co-citaciones o citas conjuntas.

El análisis de co-citas y las posibilidades de obtener *clusters* de espe-

En el análisis multivariado, la familia de herramientas estadísticas que prevalece en la determinación de los desarrollos que se dieron a partir del análisis de la aparición conjunta de palabras (*co-word analysis*), como se le conoce, o simplemente co-ocurrencia de palabras, son el análisis de *cluster* y el escalado multidimensional. A ello habría que agregar el análisis de factores, el análisis del componente principal, etcétera, sin embargo los dos primeros son los más generalizados.

El principio, al igual que el de las co-citas (en nuestro modo de ver es un caso particular del de la aparición conjunta de palabras, frases, etcétera) es sencillo. Parte del conteo de la frecuencia en que dos elementos bibliográficos coinciden en un registro. Al analizar todas las posibilidades de que ello suceda $\frac{n(n-1)}{2}$ veces, es posible estudiar las posibles asociaciones que tienen lugar (similitudes) entre los documentos. Esto da lugar a grupos (*clusters*) de documentos asociados naturalmente.

Recientemente se ha podido comprobar que más del 70% de los artículos

citados en la primera página de las patentes industriales estadounidenses, corresponden a trabajos producidos en instituciones públicas (universidades, laboratorios gubernamentales y otras agencias públicas). Esto se origina precisamente cuando la industria privada gasta mucho más que el gobierno federal en investigación. Obviamente, la alta innovación tecnológica en ese país depende, en mayor grado, del desarrollo científico público, lo cual constituye un indicador de la vinculación estrecha que existe entre la ciencia y la tecnología.



Priss/Linies

cialidades, así como la visualización mediante mapas de las interacciones de las especialidades fue ideado en 1973. Y desde entonces mucho se ha hecho en ese sentido, sin embargo, a final de los años setentas surgió la variante de hacer análisis similares no sólo de las co-citas, sino de la co-ocurrencia de diferentes elementos bibliográficos (autores, descriptores, clasificaciones etcétera), con base en el análisis de la aparición conjunta de palabras. Estos desarrollos dieron lugar a la familia de los índices relacionales de segunda generación que coexisten y/o se complementan con los de la primera.



Finalmente, la bibliografía que integró el corpus inicial objeto de estudio se enriquece con el valor añadido del análisis realizado. Segmentos de esta bibliografía, a su vez, pueden ser objeto de análisis más detallados y a los cuales se les puede aplicar de nuevo índices univariados para profundizar en aspectos de interés.

El sistema que se ha esbozado en este trabajo no quedaría completo sin referirse, aunque sea brevemente, al resultado final que se puede obtener mediante su aplicación. El efecto de aplicar sistemas de este tipo se materializa mediante un reporte de información estratégica; ésta será enriquecida y evaluada y permitirá develar los actores de las redes de investigación e innovación. Estos datos finales son necesarios para la toma de decisiones, desarrollo de políticas científicas, etcétera, que en forma de hipótesis de trabajo alimentarían a especialistas en su quehacer profesional.

■ Bibliografía

-1. E. Spinak, *Diccionario Enciclopédico de Bibliometría, Cienciometría e Informetría*, Caracas: UNESCO-CII/II, 1996, pp. 245.

-2. M. Callon, J. P. Courtial, H. Pénen, *Cienciometría. La medición de la actividad científica: de la bibliometría a la vigilancia tecnológica*, Madrid: Ediciones TREA, S.L., 1995, pp. 110.

-3. B. C. Peritz, L. Eaghe, eds., *Procedings of the sixth conference of the International Society for Scientometrics and Informetrics* (Jerusalem, Israel, June 16-19, 1997).

-Jerusalem: The Hebrew University of Jerusalem, School of Library, Archive and Information Studies, 1997, pp. 593.

-4. T. A. Brooks, *The Bibliometrics Toolbox Reference Manual*, Versión 2.8, Washington D.C.: Nort City Bibliometrics, 1987, pp. 21.

-5. L. Egghe, R. Rousseau, "Introduction to informetrics. Quantitative Methods in Library", *Documentation and Information Science*, Amsterdam: Elsevier Science Publishers B.V., 1990, pp. 450.

6. W. J. Broad, "Study Finds Public Science Is Pillar of Industry", en: *The New York Time (Science Time)*, Tuesday, May 13, 1997, pp. c1.

7. *Biblio Link II 1.01*, (1994), Personal Bibliographic Systems, Inc.

8. *ProCite for Windows*, (1986-1996), Personal Bibliographic systems, Inc.

9. *CSS.STATISTICA 4.0*, 1994 StatSoft, Inc.

