

Contribución al estudio de las revistas de América Latina y el Caribe mediante el mapeo autoorganizado*

Lic. Oscar Saavedra Fernández, 1 Lic. Gilberto Sotolongo Aguilar 2 y M.Sc María V. Guzmán Sánchez 3

Resumen

A partir del uso de las redes neuronales artificiales (RNA), y más concretamente de la técnica de mapas autoorganizados (SOM), conocidos también como algoritmo de Kohonen, se intenta identificar algunos elementos que influyen la posición que ocupan las revistas de América Latina y el Caribe (AL y C), y el efecto que determinadas acciones pudieran tener sobre ella. En su calidad de técnicas exploratorias, aquí sólo se revelan patrones de comportamiento (signos) de la realidad que necesariamente deben valorar los expertos. Al abordar este tema, se han considerado los resultados de esfuerzos regionales recientes (resultados parciales de proyectos en curso como LATINDEX y RICYT) que permitirán en el futuro enfrentar tareas de evaluación más completas.

DeCS: PUBLICACIONES SERIADAS; AMERICA LATINA; REDES NEURALES (COMPUTACIÓN); REGION DEL CARIBE

Las publicaciones científicas y técnicas constituyen una de las formas de existencia de la ciencia y la tecnología, resultado del devenir histórico de esas dos actividades; en tanto los artículos aparecidos en revistas y las patentes se erigen como las publicaciones más paradigmáticas.

Las publicaciones científicas son depositarias de los conocimientos documentados que la humanidad acumula en cualquier campo del saber; son la vía fundamental para transmitir dichos conocimientos, debido a que la transmisión directa de sus propietarios a quienes los necesitan es prácticamente imposible. Al propio tiempo, el cúmulo de materiales publicados crece en forma de avalancha y no es posible acceder directamente a ellos, mucho menos asimilar de forma directa su contenido. La propia ciencia ha permitido el desarrollo de técnicas que pueden contribuir a solucionar en parte este problema, como es el caso de los estudios cuantitativos de la información que, para diferentes fines, se han desarrollado en el transcurso de los últimos 80 años.¹

* Basado en la ponencia titulada “Mapeo autoorganizado de las revistas de América Latina y el Caribe” preparada originalmente para la Reunión de Editores de Revistas Académicas, Científicas y Tecnológicas. 10-11 de octubre de 2000, ICFES, Bogotá, Colombia.

La bibliometría, en su sentido más amplio, ha seguido dos caminos, paralelos y relacionados. En un caso, la bibliometría se ha desarrollado en el ámbito de la gestión de información de las instituciones bibliotecarias e informativas. Por otra parte, se hallan los avances de la bibliometría evaluativa cuyo propósito es la aplicación en el campo de la política científica, mediante la evaluación de la producción científica vista a la luz de las publicaciones.² Más recientemente estas aplicaciones se han extendido a las empresas productivas y de servicios como parte de las herramientas empleadas en la vigilancia científica y tecnológica, así como

en los llamados observatorios de ciencia y tecnología.

Ya en los años 60, JD de Solla Price consideraba a la ciencia como todo lo publicado en revistas y un científico como aquel que publica un artículo. Lo anterior, reconocido como “reduccionismo bibliométrico” ha conferido a las revistas y a los artículos un lugar preponderante en el desarrollo científico y tecnológico de la humanidad. [Grivel L. L'hypertexte, comme mode d'exploitation des resultats d'utiles et methodes d'analyse de l'information scientifique et technique. These PHD, CRRM, France, 2000]. No existen dudas sobre la importancia que han tenido y tienen las revistas como objeto de estudio, ellas son centro de atención en todos los órdenes en los cuatro puntos cardinales; sin embargo, en ocasiones la importancia de su existencia no se ha valorado correctamente con respecto a su función en la práctica histórico-social.

Dos reflexiones sobre el lugar de las revistas y las publicaciones permiten precisar esta idea:

- El modo social para validar el conocimiento y que permite, al mismo tiempo, su difusión pública, implica un proceso riguroso que lleva a la publicación de un artículo en una revista científica. La revista acredita el aporte cuya validez persiste, siempre que sea de su competencia.³
- La investigación científica no publicada, no existe. La publicación en una revista de prestigio reconocido asegura la prioridad en la producción de un resultado, acrecienta el crédito académico de un científico, legitima su actividad y permite la existencia de sistemas de comunicación científica ligados a procesos activos de persuasión, negociación, refutación y modificación, mediante los cuales el significado de las observaciones científicas, al igual que las interpretaciones teóricas, tienden a ser selectivamente construido y reconstruido en el campo científico.⁴

A pesar de esto, la situación que presenta la gran mayoría de las revistas científicas y técnicas editadas en América Latina y el Caribe no permite que estas puedan cumplir cabalmente con las funciones referidas con anterioridad.

Se presenta a continuación una breve caracterización de esta problemática.⁵ Se puede afirmar que las revistas científicas y técnicas de América Latina y el Caribe son poco conocidas, difícilmente forman parte de las colecciones de las bibliotecas, están representadas marginalmente en los grandes servicios de indización o resúmenes (GSIR), carecen de visibilidad y además reciben contribuciones de bajo nivel.

Otros autores han abordado la problemática de las ciencias agrícolas en la región, precisamente desde la óptica de sus publicaciones, con el objetivo de contribuir al mejor conocimiento de la realidad que presentan las publicaciones de nuestra área geográfica en este campo.¹ Ahora, a partir del uso de las redes neuronales artificiales (RNA), y más concretamente de la técnica de mapas autoorganizados (SOM), conocidos también como algoritmo de Kohonen, se intenta identificar algunos elementos que influyen en esta situación, y el efecto que determinadas acciones pudieran tener sobre ella. En modo alguno se pretende ser concluyente, porque las técnicas empleadas son exploratorias y revelan sólo patrones de comportamiento (signos) de la realidad que necesariamente deben valorar los expertos. Al abordar este tema, se han considerado los resultados de esfuerzos regionales recientes (resultados parciales de proyectos en curso como LATINDEX_{xxx} y RICYT_{xxx}) que permitirán en el futuro enfrentar tareas de evaluación más completas.

Métodos

Fuentes de datos

Se utilizaron los datos de la última actualización del directorio titulado LATINDEX <http://www.LATINDEX.unam.mx>. Su consulta, en agosto de 2000, se realizó mediante el acceso a dicho sitio como a partir del trabajo de Alonso Gamboa mencionado anteriormente.⁵

Con ello, de forma muy general, se caracterizó cualitativa y cuantitativamente la situación actual de las revistas, en ella se hace referencia fundamentalmente a su distribución por rangos⁶ en la región. Así aparecen las revistas editadas en América Latina y el Caribe. También, se utilizaron algunos datos procedentes del directorio y se complementaron con indicadores tomados de la Red Iberoamericana de Indicadores de la Ciencia y la Tecnología (RICYT) <http://www.ricyt.edu.ar/> accedida en igual fecha,⁶ se construyeron mapas autoorganizados (SOM) de América Latina y el Caribe. Ellos posibilitan la realización de diferentes análisis para determinar las variables con la influencia necesaria de manera que se mejore la visibilidad de las publicaciones del área.

Indicadores

A partir de los datos del directorio LATINDEX se estableció la distribución por rangos de las revistas que se publican en 22 países (tabla 1, variable: r-revistas (2000)). En esta tabla, se muestra la posición que cada país ocupa en relación con la cantidad de revistas editadas. Se hizo lo mismo con las revistas procesadas por GSIR (tabla 1, variables: rev-scisearch, rev-ca, rev-biosis, rev-medline, rev-cab). Con ello, es posible determinar la posición que ocupa cada país, según su aporte a estos servicios de información.

Por otra parte, a partir de datos ofrecidos por la RICYT, se obtuvieron los valores promedios de las contribuciones que durante 8 años (1990-1997), cada uno de los países realizó en artículos a GSIR (tabla 1, variables: r-scisearch, r-pascal, r-inspec, r-compindex, r-ca, r-biosis, r-medline, r-cab, r-icyt, r-ime). En este caso, se contaron los artículos indizados en dichos servicios, para incluirse en el total, la dirección del primer autor debió pertenecer a alguno de los países considerados. A partir de los datos obtenidos, se jerarquizaron por rangos, así se obtuvo la distribución por rangos de los datos, una vía para brindar una imagen de la visibilidad de la producción científica. Adicionalmente, se hallaron cuatro indicadores de la distribución por rangos de las tasas de referencias por países en las bases de datos SciSearch y Pascal, por cada 100 000 habitantes, también se determinaron tasas similares de referencias en ambas bases de datos, referidas al producto bruto interno (PBI) de cada país en miles de millones de dólares estadounidenses (USD), (tabla 1, variables: r-scisearch-hab, r-pascal-hab, r-sci-pbi, r-pascal-pbi). En total se construyeron 20 indicadores de 22 países.

** LATINDEX es un proyecto en curso que desde 1995, tiene como finalidad, la creación de un sistema de información que cubra las revistas científicas y técnicas publicadas en AL y C.

*** La RICYT es un proyecto que desde 1995 surge con el propósito de promover el desarrollo de instrumentos para la medición y el análisis de la ciencia y la tecnología en Iberoamérica, en un marco de cooperación internacional, con el fin de profundizar en su conocimiento y utilización como instrumento político para la toma de decisiones.

Para obtener un cuadro más acabado de la situación de las revistas en ALyC, se tomó de RICYT, los datos de la producción científica de ALyC correspondiente a 10 bases de datos internacionales correspondientes a 1990-1997 y se calculó su promedio***. Posteriormente se halló la distribución por rangos de dichos datos según se puede observar en la tabla 1 (variables: r-scisearch, r-pascal, r-inspec, r-compindex, r-ca, r-biosis, r-medline, r-cab, r-icyt, r-ime). Se añadieron dos indicadores: la tasa de la producción científica según SCI y PASCAL por cada 100 mil habitantes y del PBI en millones de USD (tabla 1, variables: r-scisearch-hab, r-pascal-hab, r-sci-pbi, r-pascal-pbi). Con ello se completan los datos de la tabla 1 en la que se recoge la distribución por rangos de 20 indicadores de 22 países de ALyC.

Mapeo autoorganizado

Con el propósito de explorar el comportamiento de los datos y de obtener una representación gráfica integrada de estos indicadores, los datos obtenidos se utilizaron para entrenar una red neuronal artificial (RNA) mediante el algoritmo de mapeo autoorganizado (algoritmo de Kohonen). Con ello fue posible modelar parcialmente la problemática bajo estudio y conocer la influencia de algunas de las variables (e.g. cantidad de revistas en el directorio LATINDEX, cantidad de revistas procesadas en GSIR, etc.).

Se empleó el software Viscovery SOMine® 3.0 Enterprise para obtener los SOM****; por la novedad del uso de estas técnicas, se hace necesario una pausa para ofrecer algunas explicaciones.

Marco teórico

Inspiración biológica

La corteza cerebral es posiblemente la estructura más fascinante que existe en la fisiología humana. A pesar de su enorme complejidad desde el punto de vista microscópico, la corteza revela una estructura uniforme a escala macroscópica, incluso al pasar de un cerebro a otro. Los centros correspondientes a actividades tan diversas como el pensamiento, la visión, oído, y las funciones motoras, están situados en zonas concretas de

la corteza, las cuales se hallan ubicadas, de cierta forma, unas con respecto a las otras. Además, las zonas individuales muestran una ordenación lógica de su funcionalidad. Un ejemplo (es el denominado mapa tonotópico) de las frecuencias auditivas, en el cual las neuronas próximas entre sí responden a frecuencias de sonido similares según una sucesión ordenada, desde los tonos más altos a los más bajos. Otro ejemplo (es el mapa somatotópico) de nervios motores, que se representa artísticamente mediante el homúnculo. Las regiones tales como el mapa tonotópico y el mapa somatotópico suelen recibir el nombre de correspondencias ordenadas de características. El interés por descubrir esta organización es el que llevó a Teuvo Kohonen (Finlandia) a desarrollar el algoritmo al que se refiere este apartado.⁶

**** Tomado de: Red Iberoamericana de Indicadores de Ciencia y Tecnología (RICYT). Indicadores de ciencia y tecnología. Iberoamericanos/Interamericanos (1990-1997). Buenos Aires: REDES A.C; 1999.

Tabla 1. Distribución por rangos de todos los indicadores (22 países de ALyC y 20 indicadores de ciencia y tecnología)

Pais	r-revistas (2000)	rev-	rev-ca	rev-biosis	rev-medicine	rev-cab	r-scisearch	r-pascal	r-inspec	r-compindex	r-ca	r-biosis	r-medicine	r-cab	r-icyt	r-ime	r-scisearch	r-pascal-hab	r-sci-pbi	r-pascal-pbi
Argentina	1	3	4	3	6	3	2	3	3	3	2	2	3	2	1	1	4	2	9	8
Barbados	16	10	9	12	6	7	17	17	15	14	17	17	16	19	20	21	14	14	12	11
Bolivia	15	10	1	12	6	7	16	15	17	16	15	16	17	16	11	15	9	8	13	14
Brasil	3	3	9	1	1	1	1	1	1	1	1	1	1	1	5	5	13	13	18	15
Colombia	6	6	4	6	6	7	7	4	4	5	4	4	4	5	4	4	7	6	3	4
Costa Rica	9	6	4	5	2	7	9	7	7	7	7	7	7	7	7	8	12	9	4	2
Cuba	4	6	1	3	6	5	8	8	11	11	10	11	11	8	10	13	3	3	2	3
Chile	5	2	4	6	2	7	4	6	6	6	6	6	6	4	2	2	17	16	19	16
Ecuador	11	10	9	12	6	7	14	13	13	12	14	15	15	15	12	9	22	22	22	22
El Salvador	14	10	9	12	6	7	22	22	22	22	22	22	22	22	13	10	16	18	15	19
Guatemala	12	10	9	12	6	7	15	16	17	17	16	14	14	13	18	12	20	20	16	17
Honduras	18	10	9	6	6	7	19	20	19	19	21	18	20	17	15	19	11	11	14	13
Jamaica	20	10	9	12	6	7	6	11	12	13	11	10	8	12	20	21	18	17	11	7
México	2	1	4	2	2	2	3	2	2	2	3	3	2	3	3	3	8	12	6	11
Nicaragua	22	10	9	12	6	7	20	19	21	17	18	19	18	18	17	18	19	19	21	20
Panamá	13	10	9	12	6	7	12	14	16	14	13	12	13	14	15	16	15	15	17	18
Paraguay	20	10	9	12	6	7	21	18	20	19	19	13	19	20	18	17	21	21	20	21
Perú	10	10	9	6	6	7	11	9	10	10	9	9	9	9	8	11	6	6	8	9
R. Domini- cana	16	10	9	12	6	7	18	21	14	21	20	21	21	21	14	14	10	10	10	10
Trinidad y Tobago	19	6	9	6	6	5	13	12	9	8	12	20	12	11	20	19	1	1	5	6
Uruguay	8	10	9	12	6	7	10	10	8	9	8	8	10	10	9	7	2	5	1	1
Venezuela	7	3	1	6	2	3	5	5	5	4	5	5	5	6	6	6	5	4	7	5

La corteza es, en esencia, una capa extensa (aproximadamente de 1m2 , en humanos adultos) y fina (entre 2 y 4 mm de grosor) que consta de seis capas de neuronas (con un gran nivel de interconexión entre ellas) de distintos tipos y densidades (unos diez mil millones de neuronas) que interaccionan mediante los procesos físico-químicos. La corteza está plegada de forma tal que maximiza la densidad de empaquetado en el cráneo. Una hoja bidimensional de elementos de procesamiento es un buen modelo de la corteza cerebral.

Lo artificial

Las redes neuronales artificiales, o simplemente redes neuronales, intentan emular lo antes descrito

mediante procesos computacionales. Se puede pensar que las redes neuronales son técnicas de regresión, no lineal, en múltiples capas y de forma paralela. Existen dos tipos de redes neuronales: las supervisadas y las no supervisadas.

Las redes neuronales supervisadas son técnicas para extraer datos a partir de las relaciones de entrada-salida y para almacenar tales relaciones en ecuaciones matemáticas que pueden utilizarse en actividades de pronóstico o en la toma de decisiones. Requieren de que el usuario especifique la salida deseada. La red aprende a detectar la relación entre las entradas y las salidas suministradas, mediante un proceso iterativo y adaptativo. Una vez que la red se ha entrenado, se puede utilizar con datos que nunca haya visto o puede ser embebida en un programa para el apoyo a las decisiones.

Las redes neuronales no supervisadas son técnicas para clasificar, organizar y visualizar grandes conjuntos de datos. Los mapas autoorganizados (SOM) son un ejemplo del enfoque de las redes neuronales no supervisadas. Este enfoque, con origen a comienzos de los años 80, se ha utilizado ampliamente en ingeniería y en muchos otros campos. Muchas aplicaciones de redes neuronales no supervisadas y de SOM pueden encontrarse en la obra de *Teuvo Kohonen*. 7

Los mapas autoorganizados son un enfoque de red neuronal con alimentación directa que utiliza un algoritmo de entrenamiento no supervisado, que mediante un proceso conocido como autoorganización, configura las unidades de salida en una representación topológica de los datos originales. Los SOM, son una técnica de red neuronal que aprende sin supervisión. En contraste con las técnicas de las redes neuronales supervisadas que requieren que se especifique una o más salidas, conjuntamente con una o más entradas a fin de encontrar patrones o relaciones entre los datos, los SOM reducen datos multidimensionales a un mapa de menores dimensiones o una rejilla de neuronas.

El algoritmo de los SOM se basa en el aprendizaje competitivo. Proporciona un mapeo de preservación topológica a partir de un espacio multidimensional para mapear las unidades. Las unidades de mapeo o neuronas, usualmente forman una rejilla de dos dimensiones, el mapeo es de un espacio multidimensional en un plano o cubo. La propiedad de preservación topológica significa que en los SOM se agrupan vectores de datos de entradas similares en neuronas: los puntos que se encuentran cercanos unos de otros en el espacio de entrada son mapeados en el SOM, en unidades del mapa que son cercanas; por ello el SOM puede servir como una herramienta de agrupamiento (clustering) así como para visualizar datos multidimensionales.

Los mapas autoorganizados consisten de dos capas de unidades de procesamiento: la primera es una capa de entrada que contiene unidades de procesamiento para cada elemento en el vector de entrada; en segundo lugar está la capa de salida o rejilla de unidades de procesamiento, que está completamente conectada con la capa de entrada. El número de unidades de procesamiento en la capa de salida es determinado por el usuario, a partir de la forma y el tamaño inicial del mapa deseado. A diferencia de otras redes neuronales, no existe ninguna capa ni unidad de procesamiento oculto.

En contraste con los métodos clásicos, los mapas autoorganizados ofrecen una fácil visualización, imponen pocos presupuestos y restricciones, y son capaces de manipular grandes conjuntos de datos para detectar en estos, estructuras y patrones aislados. Por ello, los mapas autoorganizados han cobrado creciente interés para el análisis exploratorio de datos y la minería de datos en diferentes campos entre los que se destacan las finanzas y la economía. Su aplicación en el campo de la bibliometría es reciente.

Proceso de autoorganización

Cuando se presenta un patrón de entrada a una red SOM, las unidades en la capa de salida compiten entre sí por el derecho de ser declarada la ganadora. La unidad de salida ganadora será la unidad cuya ponderación de la conexión de entrada está más cerca del patrón de entrada en términos de distancia euclidiana. Entonces, la entrada se presenta y cada unidad de salida compete para cotejar con el patrón de entrada. La salida que está más cerca al patrón de entrada es declarada la ganadora. Entonces, se ajusta la ponderación de la conexión de la unidad ganadora (i.e), se mueve en la dirección del patrón de entrada por un factor determinado por la tasa de aprendizaje; lo anterior constituye la naturaleza básica de las redes neuronales competitivas.

Los SOM realizan el mapeo topológico no sólo al ajustar las ponderaciones de la unidad ganadora, también ajustan las ponderaciones de las unidades de salida que son adyacentes en la vecindad más próxima de la unidad ganadora. Por lo tanto, no sólo la unidad ganadora es ajustada, sino también que toda la vecindad de unidades de salida se mueve más cerca del patrón de entrada. Con frecuencia, comenzando con valores aleatorios de las ponderaciones, las unidades de salida se alinean solas, de forma tal, que cuando se presenta un patrón de entrada, entonces una vecindad de unidades responde a los patrones de entrada. Según el entrenamiento progresa, decrece el tamaño de la vecindad alrededor de la unidad ganadora. Inicialmente, se actualiza una gran cantidad de unidades de salida, pero en la medida en que avanza el entrenamiento sólo se ajusta la unidad ganadora. De forma similar, las tasas de aprendizaje disminuirán según avanza el entrenamiento, y en algunas aplicaciones, la tasa de aprendizaje se reducirá con la distancia de las unidades de salida ganadoras.

El resultado es un grupo de conexiones ponderadas que representa un patrón de entrada prototipo para el subconjunto de entrada que corresponde a un grupo (cluster) en particular. El proceso de reducir datos multidimensionales a un conjunto de grupos (clusters) se llama segmentación. El espacio multidimensional de entrada se reduce a un mapa de dos dimensiones. Si se utiliza el índice de la unidad de salida ganadora, en esencia lo que produce es la partición de los patrones de entrada en un conjunto de categorías o grupos (clusters). Los SOM también tienen la capacidad de generalizar. Esto significa que la red puede reconocer o caracterizar entradas que nunca antes haya recibido. Una nueva entrada se asimila por la unidad del mapa con la cual se mapea. Aún más, los vectores de entrada con falta de datos pueden utilizarse para buscar o pronosticar los valores de los datos que faltan, basados en el mapa entrenado.

Los resultados obtenidos de la exploración de datos al utilizar esta técnica de los SOM, sugieren ideas de cómo mejorar la visibilidad de la ciencia y la tecnología de ALyC a diferentes niveles.

Resultados y discusión

La actualización del directorio LATINDEX, correspondiente al año 2000,5 recoge unas 7 000 revistas de cerca de 30 países de Iberoamérica (se ha incluido a España, Portugal y organizaciones internacionales radicadas en la región), así como otras islas del Caribe no incluidas anteriormente en el Directorio LATINDEX. El 61 % de las revistas corresponde a países de ALyC y también del total, se ha identificado como activas el 60 %. Más de la mitad de la colección corresponde a revistas en ciencias sociales (39,8 %) y humanidades (13 %). El resto se distribuye entre 6 agrupaciones temáticas: ciencias de la salud (15,1 %), ingenierías (9,7 %), ciencias de la tierra (3,7 %), ciencias naturales (11,4 %), ciencias exactas (4,7 %) y revistas multidis-ciplinarias (2,6 %). La mayoría de las revistas surgieron en el transcurso de las últimas tres décadas. Más de la mitad de las revistas tiene una frecuencia anual-trimestral-irregular-semestral; y sólo el 2 % aparece en alguna modalidad electrónica (i.e. en línea, CD ROM o disco flexible). Finalmente, el 62 % se edita en español, mientras que el 22 % en portugués, el 15 % son multidisciplinarias y el 1 % aparece en inglés.

En la tabla 1 (variable: r-revistas (2000)), se puede observar la posición de cada uno de los países, según el rango que ocupan en el directorio LATINDEX. Argentina ocupa la primera posición, seguida de México, Brasil, Cuba, Chile y Colombia. Colombia a su vez supera entre otros a Venezuela, Uruguay, Costa Rica y Perú.

Mediante al acceso al directorio LATINDEX, se pudo identificar la presencia de revistas de ALyC procesadas en algunos de los GSIR. De esta forma, se pudo saber que en SciSearch, que incluye los tres índices de citas editados por el ISI (Institute for Scientific Information, Filadelfia), está representada de la siguiente forma SCI (Science Citation Index) = 14 (Colombia=1), SSCI (Social Science Citation Index)=8 (Colombia=1), AHCI (Arts & Humanities Citation Index) = 6. En otros GSIR, la presencia es la siguiente: PASCAL = 1; INSPEC = 1; Chemical Abstract (CA) = 10 (Colombia = 1); Medline = 9; Biological Abstracts (BIOSIS) = 46 (Colombia = 1), Commonwealth Agricultural Bureau (CAB) = 20. Resulta curioso que COMPENDEX no aparece en la lista de servicios de resúmenes reportados por LATINDEX (ni tampoco Engineering Index). La temática de las revistas de ALyC procesadas en GSI denota el peso específico de las ciencias sociales y humanidades, así como de las biociencias en la región.

En la tabla 1 (variables: rev-scisearch, rev-ca, rev-biosis, rev-medline, rev-cab) aparece la distribución por rangos de la presencia de revistas de ALyC en 5 de los GSIR. Así por ejemplo, si se compara Cuba y

Colombia se observa que ambos países ocupan la misma posición (6) por la presencia de sus revistas en SciSearch y Medline; al tiempo que Cuba supera a Colombia respecto a los restantes GSIR, esto es, en Chemical Abstract, Cuba 1, Colombia 4, en BIOSIS, Cuba 3, Colombia 6 y en CAB, Cuba 5 y Colombia 7.

La visibilidad de la producción científica de ALyC, no se logra sólo por medio de la indización de sus revistas en los grandes servicios de indización o resumen, sino también mediante la publicación de los trabajos de los autores de la región en otras revistas fuera de esta área geográfica. Se observa que las ciencias agrícolas (CAB) constituyen el mayor aporte a dichos servicios, ICYT e IME tienen carácter local al reflejar la producción científica de España (tabla 2). Por otra parte en la tabla 1 (variables: r-scisearch, r-pascal, r-inspec, r-compindex, r-ca, r-biosis, r-medline, r-cab, r-icyt, r-ime) aparecen las aportaciones de referencias independientemente del lugar de publicación de cada país a cada uno de las GSIR. Estos indicadores están liderados por Brasil y Argentina. También se puede observar, al comparar de nuevo a Cuba con Colombia, que la primera ocupa posiciones que oscilan entre 8 y 11, mientras que las de Colombia oscilan entre 4 y 7. Otros análisis son posibles, sin embargo resulta difícil manejar dichas tablas sin el auxilio de herramientas para realizar valoraciones multidimensionales. En la práctica se dispone de los valores que caracterizan hasta 20 dimensiones (variables) de 22 países de ALyC.

Tabla 2. Producción científica de ALyC según su aporte a las bases de datos internacionales (datos de 1997) rango Bases de datos %

Rango	Bases de datos	%
1	ICYT	5,3
2	CAB	5
3	IME	3,8
4	BIOSIS	2,3
5	SciSearch	2,1
6	PASCAL	1,8
7	INSPEC	1,6
8	COMPENDEX	1,4
8	CA	1,4
8	MEDLINE	1,4

Estos datos permiten alimentar una red neuronal no supervisada de forma que se obtiene el mapeo autoorganizado (SOM) de los datos conforme al algoritmo de Kohonen. De esta manera, se puede disponer de mapas de ALyC en dos dimensiones a partir de los datos referidos anteriormente y examinarse la influencia que ejercen las distintas variables en la conformación del mapa de ALyC (22 países), conformado a partir de los indicadores bajo estudio (20 variables).

SOM de ALyC vs. todos los indicadores

La figura 1 muestra el mapa autoorganizado de ALyC considerando las 20 variables simultáneamente. Como se puede observar, se forman 21 clusters o agrupamientos coincidiendo en todo los casos con cada uno de los países con la sola excepción de Nicaragua y Paraguay los que se agrupan en un solo cluster. Este resultado sugiere la idea de que todas las variables permiten un mapeo de ALyC donde cada uno de los países, con las excepciones antes señaladas, muestran su plena identidad. La situación de Nicaragua y Paraguay se debe a los bajos rangos que tienen en todas las variables, que se pueden constatar en la tabla 1.

SOM de ALyC vs. directorio LATINDEX

La figura 2 muestra el agrupamiento de los países, si se considera sólo la dimensión correspondiente al aporte que hace el Directorio LATINDEX al mapa de ALyC. Como se puede observar, sólo se forman 3

clusters, los que de abajo hacia arriba están conformados por:

1. Chile, Brasil, México, Venezuela, Argentina, Colombia y Cuba.
2. Ecuador, Guatemala, Panamá, Costa Rica, Perú y Uruguay.
3. Paraguay, Nicaragua, República Dominicana, Bolivia, Barbados, El Salvador, Honduras, Jamaica y Trinidad y Tobago.

De nuevo, el agrupamiento se debe a la semejanza de las posiciones (rangos de los países debido a su aporte de revistas registradas en el Directorio LATINDEX). Como se puede corroborar, la mera presencia del directorio ejerce muy poca influencia en la conformación del mapa de ALyC, presentado en la figura 1.

SOM de ALyC vs. revistas de ALyC procesadas por GSIR

La figura 3 muestra el agrupamiento de los países atendiendo a la presencia de algunas revistas de ALyC en GSIR. Como se puede apreciar, esta variable muestra signos de influencia en el mapa completo de ALyC, donde se incluyen todas las variables (figura 1). Asimismo, aparecen claras definiciones de las fronteras de los países: Honduras, Trinidad y Tobago, Bolivia, Perú, Costa Rica, Cuba, Colombia, Chile, Brasil y Argentina. Venezuela y México forman un cluster; también, se forma otro cluster con los restantes países. Estos resultados sugieren la idea de que en general, aunque la presencia en los GSIR de las revistas de ALyC es baja, los dos cluster formados por múltiples países presentan la misma situación; por este concepto no se logra ni una visibilidad mínima.



Fig. 1. Mapa auto-organizado de ALyC.



Fig. 2. Agrupamiento de los países según el aporte que hace el Directorio Latindex al mapa de ALyC.

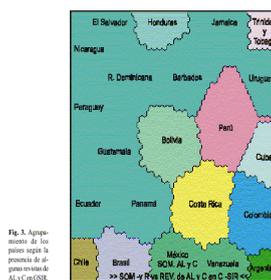
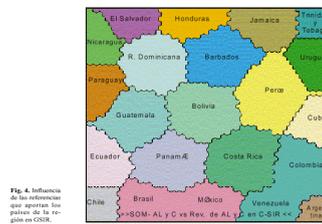


Fig. 3. Agrupamiento de los países según la presencia de algunas revistas de ALyC en GSIR.



SOM de ALyC vs. las referencias de ALyC presentes en GSIR

La figura 4 muestra la influencia de las referencias que aportan los países de la región en GSIR con independencia de las revistas que en ellos se procesan. Su influencia en la conformación del mapa de ALyC es indudable (figura 1). Todos los países se agrupan en sí mismo, incluso, Paraguay y Nicaragua; sin embargo, los dos clusters formados por Brasil y México de una parte, y Venezuela y Colombia de la otra, sugieren la idea de la insuficiencia de esos países por sí mismos de lograr la visibilidad adecuada mediante este indicador. En particular es necesario destacar la situación de México y de Venezuela, que aunque agrupados de forma diferente, si se observan las figuras 3 y 4, no es posible realizar su plena identificación.

SOM de ALyC vs. las tasas de referencias en SCI y Pascal (por cien mil habitantes y en millones de U\$D del PBI)

La influencia de los indicadores relacionados con la cantidad de referencias aportadas por países, es mínimo, tanto al SCI como a las bases de datos PASCAL, en relación, por una parte, con cada cien mil habitantes y, por otra, con el PBI en millones de U\$D (figura 5).

Como puede observarse, los 22 países de la región se agrupan en 3 clusters que denotan la semejanza de ellos en estos cuatro indicadores. No se observa prácticamente ninguna influencia de estos indicadores sobre la estructuración del mapa en la figura 1.

Otras consideraciones

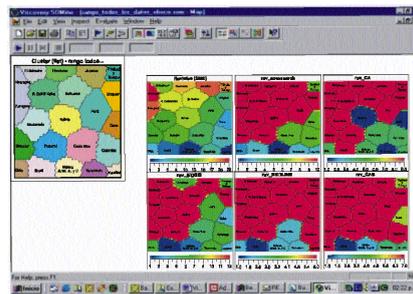
Por las características del método empleado en este tipo de análisis, es posible realizar una mayor cantidad de análisis y con más profundidad. Considere-se que de cada uno de los mapas mostrados, es posible obtener hasta otros 25 mapas en los que puede examinarse el comportamiento de cada variable. Al propio tiempo es posible revelar la influencia de cada variable, una por una, como sucede en el caso de la figura 2, que muestra la influencia de la variable correspondiente al directorio LATINDEX, o de una combinación de estas, como es en el caso de las restantes figuras en las cuales se muestra la influencia de diversas variables de forma simultánea. La figura 6 es un ejemplo de lo antes planteado.

La figura 6 presenta una pantalla del software Viscosity SOMine® después de realizar el entrenamiento de la RNA (con los datos de los 20 indicadores de los 22 países de ALyC) conforme al algoritmo de Kohonen. El gráfico que aparece en la parte superior izquierda, muestra el mapa completo de ALyC según se presentó en la figura 1. A continuación se presentan otros siete gráficos. Los primeros tres, situados en la parte superior derecha así como los primeros tres en la parte inferior izquierda, corresponden a la tabla 1 (variables: r-revistas (2000), rev-scisearch, rev-ca, rev-biosis, rev-medline, rev-cab). Los colores que presenta cada país se corresponden con la escala de colores que aparece debajo de cada gráfico, así los tonos de azul corresponden a valores de cada variable denotando un alto rango, en tanto hacia la derecha los tonos de rojos denotan un nivel más bajo en la escala. Los colores intermedios denotan rangos intermedios. El último gráfico en el extremo inferior derecho representa la curvatura del mapa. Las zonas con azul más intenso muestran la delimitación de los clusters, esta es la zona donde el plano flexible se cambia formando una suerte de accidente geográfico montañoso (ver la escala en el borde inferior del gráfico). Las zonas de una azul menos intenso, pueden llegar a ser completamente blancas, ahí están los llanos o sabanas donde se ubican los países según los datos utilizados.



Conclusiones y recomendaciones

La simulación y modelación realizada de la situación de algunos de los aspectos que caracterizan la problemática de las revistas científicas y técnicas de ALyC, muestra que la visibilidad de la producción científica de los países de la región es un hecho multidimensional y que cada una de dichas dimensiones juega su papel, unas con más fuerza que otras. La necesidad de publicación que tienen nuestros científicos, tecnólogos y especialistas se satisface con publicaciones, tanto en las revistas del área como de otras latitudes; en estas últimas sucede con mayor frecuencia. Incrementar, en todos los órdenes, la calidad de nuestras publicaciones es una tarea impostergable, como lo es también, lograr que exista una mayor cantidad de las revistas de ciencia y técnica de ALyC, procesadas por GSIR. Las nuevas y variadas formas de presentación de las publicaciones (resultantes de la aparición de las nuevas tecnologías de información) por parte de los GSIR, posibilita que estas tareas se faciliten; al tiempo que permiten una presencia más efectiva de las publicaciones y su consecuente visibilidad.



Agradecimientos

Al colega Octavio Alonso Gamboa que propició el acceso a su reciente artículo sobre Latindex del cual se utilizaron parte de sus datos para realizar el presente trabajo. Al Dr. Gerhard Kranner, presidente de la firma Eudaptics Software GmbH de Viena, Austria, por su licencia para emplear el software Viscovery SOMine® 3.0 Enterprise y que se usó en este trabajo para obtener los SOM.

Abstract

Contribution to the study of the journals from Latin America and the Caribbean by self-organized mapping

Starting from the use of neural artificial networks (NAN), and more concretely from the technique of self-organized maps (SOM), also known as Kohonen's algorithm, the authors try to identify some elements influencing on the position occupied by journals from Latin America and the Caribbean and the effect that certain actions may have on it. As exploratory techniques, they only reveal patterns of behavior (signs) of reality that have to be assessed by experts. On dealing with this topic, there have been considered the results of recent regional efforts (partial results of ongoing projects as LATINDEX and RICYT) that will allow to face in the future more complete tasks.

Subject headings: SERIALS; LATIN AMERICA; NEURAL NETWORKS(COMPUTERS); CARIBBEAN REGION

Referencias bibliográficas

1. Saavedra O, Sotolongo G. Medición de la producción científica en América Latina en el campo agrícola y afines: un estudio bibliométrico. XII RIBDA, Coronado, Costa Rica, mayo-junio 1999.
2. Narin F. Evaluative bibliometrics: the use of publications and citation analysis in the evaluation of scientific activity. Washington, D. C: Computer Horizons; 1976.
3. Krauskopf M, Vera MI. Las revistas latinoamericanas de corriente principal: indicadores y estrategias para su consolidación. *Interciencia* 1995;20(3):144-8.
4. Vessuri Hebe MV. La revista científica periférica. El caso de *Acta Científica Venezolana*. *Interciencia* 1987;12(3):124-34.
5. Cetto AM, Alonso Gamboa O. Latindex como fuente de información bibliográfica: resultados y productos iniciales. *Indicios* 2001;(1):33-7.
6. Freeman JA, Skapura DM. *Redes neuronales: algoritmos, aplicaciones y técnicas de programación*. Nueva York: Addison Wesley; 1993.
7. Kohonen T. *Self-organization and associative memory*. Nueva York: Springer Verlag; 1984. (Springer Series in Information Science; 8).

Recibido: 14 de mayo de 2002

Aprobado: 20 de junio de 2002

Lic. *Oscar Saavedra Fernández*. Gerente General de EBSCO México, INC. S.A. DE C.V. Ensenada 91. Col. Hipódromo Condesa, Del. Chauhtemoc 06170 México, D.F. México. E.mail: osaavedra@exchange.ebsco.com

1 Gerente General de EBSCO México, INC. S.A. DE C.V.

2 Licenciado en Información Científico-Técnica y Bibliotecología. Director de Información Científica del Instituto Finlay. Cuba.

3 Máster en Ciencias de la Información. J del grupo de Gestión de Información.

© 2004 2000, *Editorial Ciencias Médicas*

Calle E No. 452 e/ 19 y 21, El Vedado, La Habana, 10400, Cuba.

e-Mail

acimed@infomed.sld.cu