
ANÁLISIS BIBLIOMÉTRICO DE LA PRODUCCIÓN SISTEMÁTICA EN AMÉRICA LATINA

Bibliometric Analysis of Systematics Production in Latin America

LAYLA MICHÁN¹, Ph. D.

¹ Departamento de Biología Evolutiva, Facultad de Ciencias UNAM.
México, D.F. laylamichan@ciencias.unam.mx

Presentado el 13 de mayo de 2010, aceptado 9 de marzo de 2011, correcciones 26 de mayo de 2011.
Trabajo presentado en la Cátedra José Celestino Mutis de Biodiversidad, Bogotá, 2010.

RESUMEN

Presento la investigación sobre las bases de datos bibliográficas Biosis, CAB, Periódica, SCI, *Scopus* y *Zoological Record*. Se estudian características para identificar cuáles y cuántas es necesario utilizar para tener la información más representativa. De 1.363 revistas analizadas, el 36,5% se encuentra en más de una base de datos y el 63,46% se registra solo en una. *Zoological Record* contiene 63.421 documentos, Biosis 19.079, CAB 14.363, Periódica 11.185, SCI 3.257 y *Scopus* 916. CAB y ZR son las más antiguas, el promedio de artículos publicados por año para Biosis fue de 6.417, ZR 433, Periódica 361, CAB 145, SCI 91 y *Scopus* 30. El análisis de redes muestra mayor relación entre Biosis y SCI, que comparten 16% de títulos, CAB y Biosis comparten 15%, Biosis y ZR 14%, y CAB y ZR 11%. Un análisis de parsimonia muestra que Biosis y SCI comparten más revistas y Periódica tiene más revistas exclusivas (285). Periódica publicó el 24,6% de artículos con descripciones de nuevos taxones y un promedio de 90 al año, CAB 54,6% con promedio de 82 y ZR 72,9% con 322. Ninguna de las bases de datos cumple con las características suficientes empleadas como fuente de información única, es conveniente utilizar distintas combinaciones de acuerdo con las necesidades informativas. Biosis tiene la información más exhaustiva sobre cualquier taxón actual, *Zoological Record* y CAB tienen una cobertura temporal amplia e incluyen principalmente animales o plantas, respectivamente. SCI tiene el mayor número de documentos en revistas de corriente principal, con referencias, citas e indicadores bibliométricos. Periódica cuenta con mayor cantidad de documentos y cobertura temporal más amplia sobre lo publicado en América Latina. El cladograma obtenido por el análisis de parsimonia resultó ser una herramienta de visualización ideal para representar las características principales de las bases de datos.

Palabras clave: Bibliometría, revistas, artículos, análisis de parsimonia, análisis de redes, sistemática.

ABSTRACT

I compare six bibliographic databases with information on Latin American systematics: Biosis, CAB, Periódica, SCI, *Scopus* and *Zoological Record*. The databases are

characterized and compared considering their content, temporal, typological, geographical, thematical coverage, kind of access and new taxon description, to identify which and how many should be used to be more representative. Of the 1363 journals analyzed, 36.5% are found in more than one database and 63.46% are recorded in a single one. Zoological Record contains 63421 documents, Biosis 19079, CAB 14363, Periódica 11185, SCI 3257 and Scopus 916. CAB and ZR are the oldest databases, the average number of articles published per year was 6417 for Biosis, 433 for ZR, 361 for Periódica, 145 for CAB, 91 for SCI and 30 for Scopus. According to the network analysis, there are stronger relations between SCI and Biosis, which share at least 16% of titles, CAB and Biosis share 15%, Biosis and ZR 14%, and CAB and ZR 11%. Based on the cladogram obtained from a parsimony analysis on the shared journals, the strongest relation is between Biosis and SCI; Periódica has the largest number of exclusive journals with 285. ZR has 72.9% of published articles with descriptions of new taxa and an average of 322 a year, CAB 54.6% with 82, and Periódica 24.6% with 90. None of databases meets the characteristics to be used as a single source of information, therefore it would be appropriate to use different combinations according to the aim of the analysis. Biosis has the most comprehensive information on any current taxon, Zoological Record and CAB have a broader temporal coverage and include mainly animals or plants, respectively. SCI has the largest number of documents in mainstream journals, with references, citations and bibliometric indicators. Periódica has the largest number of documents and temporal coverage published in the area. The cladogram proved to be an optimal visualization tool to represent the main features of each data base.

Key words: Bibliometry, journals, parsimony analysis, network analysis, systematics.

INTRODUCCIÓN

La investigación sistemática se basa en información biológica y literatura sobre las especies y las dos se encuentran dispersas en el tiempo y el espacio, por lo tanto es necesario reunir las y organizarlas para su uso. El producto de la investigación sistemática es fundamentalmente la literatura, por medio de la cual se evalúa, difunde y discute el nuevo conocimiento producido. Por lo tanto, la bibliografía es materia prima y producto de la investigación taxonómica. De hecho, esta afirmación podría generalizarse a cualquier tipo de investigación científica, con la diferencia que en sistemática la literatura presenta singularidades interesantes respecto al valor, vigencia, vida media e impacto (Krell, 2002), propias de la dinámica de esta disciplina.

Para llevar a cabo un estudio sistemático, a diferencia de otros, es necesario hacer una revisión exhaustiva de literatura, tanto antigua como contemporánea, pues los hallazgos del pasado no pierden vigencia, importancia y significado. Muchos datos antiguos siguen siendo útiles, por ejemplo la información de nuevos registros, los estudios florísticos o faunísticos y la descripción de nuevas especies en la práctica científica nunca caducan, siempre son fuente de referencia geográfica, taxonómica, nomenclatural y cronológica sobre las especies. Uno de los principales problemas que enfrentan los sistemáticos es la escases de información bibliográfica, además la que hay en muchos casos está dispersa, poco actualizada y mal sistematizada (Godfray *et al.*, 2007).

Actualmente las nuevas tecnologías electrónicas y de comunicación han reducido trabajo, costo y tiempo requeridos para el análisis de información. Se producen a diario nuevas herramientas para realizar búsquedas más eficientes y precisas, y para hacer análisis mejores y más extensos. El acceso inmediato a datos de una variedad extensa de fuentes especializadas de diversas procedencias y temas ha fomentado la colaboración, integración e interdisciplinariedad (NSF, 2005). La herramienta por excelencia para sistematizar la información es la base de datos, un sistema informático de registros almacenado en tablas con un orden establecido que permite guardar, ordenar, mantener, procesar, recuperar, presentar y generar información.

El diseño y difusión de gran cantidad de bases de datos que hay en la actualidad, particularmente las que se pueden consultar en línea, fue sin duda, el resultado de un fenómeno representativo surgido en la industria de la información de finales del siglo XX. En 1951 el *U.S. Bureau of Census* diseñó la primera base de datos formal. Diez años después se crearon las primeras con bibliografía, que en un principio contenían resúmenes e índices, posteriormente incluyeron fotocomposiciones y citas que se almacenaban en cintas magnéticas. Entre las más destacadas se encuentran: 1) *Medical Literature Analysis and Retrieval System* (MEDLARS), producido en 1960 por *The National Library of Medicine* (ahora MedLine, proyecto en el cual se inserta PubMed); 2) *Chemical Registry System and Chemical and Biological Activities* (CBAS), publicados en 1965 por el *Chemical Abstracts Service* en forma impresa y cinta magnética para computadora; 3) *Machine Readable Catalog* (MARC), emitida en 1969 por *The Library of Congress*; 4) *Engineering Information* en 1967 y 5) *BioSciences Information Service* (BIOSIS) en 1969 (Neufeld y Cornog, 1986). La última será analizada en este trabajo. Se calcula que en 1965 existían 12-20 bases de datos bibliográficas, en 1968 había 25, para 1970 en promedio había 75 (Neufeld y Cornog, 1986), en 1986 la cuenta fue de 3.337 (Hawkins, 1984) y para el 2008 aparecen cerca de 15.600 (Thomson Reuters, 2008c).

En general todas las bases de datos bibliográficas actuales son relacionales, se caracterizan por contener registros con información básica sobre: 1) documentos (generalmente artículos) como título, tipo de documento, idioma, palabra clave, descriptores, volumen, número y páginas; 2) fuente de donde provienen (principalmente revistas) como título, año de publicación y tema; y 3) autoría, el o los nombres de los autores, institución de inscripción y país. Para cada documento (o la fuente en algunos casos) se asignan descriptores para clasificarlos y utilizarlos posteriormente como referencia para recuperarlos por medio de búsquedas. Los índices más completos también contienen resúmenes, referencias, citas, enlaces al documento en texto completo (libre o restringido a pago), documentos relacionados, análisis bibliométricos sencillos (cuentas de autores, temas, países, tipo de documento, idioma y descriptores) o complejos (citación, vida media e índice h), catálogos y tesauros, y herramientas electrónicas para guardar, etiquetar (tagging), almacenar, sistematizar, analizar y manejar las referencias recuperadas. Entre las características más importantes a considerar en una base de datos documentales están la cantidad de registros y tipo de campos que capturan, criterios de inclusión, sistematización, relación y clasificación de la información, tipo de acceso, características de la aplicación, herramientas de búsqueda, manejo y análisis de registros, cobertura (representación) tipológica, idiomática, temática, geográfica y temporal.

Cada base de datos tiene características diferentes. Para elegir cuáles y cuántas es necesario reconocer las siguientes particularidades de cada una. La primera es el tipo de acceso: libre o restringido, se refiere a que solo se puede acceder previa contratación y pago, que generalmente se hacen de forma institucional, pues el precio es demasiado alto (unos miles de dólares) para hacerlo de forma individual. Estos servicios se pueden consultar, generalmente, en bibliotecas de instituciones de enseñanza, investigación o en algunas instancias gubernamentales. La segunda es la cobertura temática, que se obtiene estudiando la distribución de registros para algún tema en distintos índices (Hood y Wilson, 2001). La tercera se refiere a la cobertura regional que se identifica a partir del lugar de edición de los documentos que se incluyen. Y por último, los criterios de inclusión y clasificación de los documentos, para lo cual se utilizan catálogos y tesauros diseñados por especialistas en ciencias de la información (documentalistas) y constituyen una herramienta importante para buscar y analizar la información de manera completa, consistente y eficaz. La búsqueda, acceso, análisis y actualización de literatura en bases de datos se ha vuelto una tarea cotidiana. Generalmente es necesario consultar varios índices para tener la más completa representación de literatura sobre el tema de interés. Pero es tal la cantidad y diversidad de documentos sobre taxonomía, son tantos, tan diferentes y complejos los recursos electrónicos (en especial las bases de datos bibliográficas), por medio de los cuales se puede acceder a dicha información, y no solo eso, sino que cambian, progresan y se actualizan constantemente, lo que hace difícil estar al tanto de todos ellos e identificar cuáles y cuántos se pueden y deben utilizar.

Para responder estas preguntas es preciso realizar estudios comparativos entre las distintas bases de datos que abordan el mismo tema, para identificar la distribución de documentos entre ellas, y establecer sus similitudes y diferencias (Hood y Wilson, 2001; Hood y Wilson, 2003a). Con este objetivo realicé un análisis comparativo de las seis bases de datos bibliográficas especializadas con mayor cantidad de información sobre sistemática de América Latina, caracterizándolas y comparando su contenido respecto a las revistas que comparten y las que les son exclusivas, su cobertura temporal, tipológica, geográfica, temática, idiomática, tipo de acceso y la descripción de nuevos taxones.

MATERIALES Y MÉTODOS

La investigación consistió en la búsqueda, recuperación, migración, homogeneización, normalización, validación, análisis y visualización de la información. Primero, se realizó una búsqueda de las bases de datos disponibles con información bibliográfica (especialmente artículos) digital sistematizada sobre sistemática en América Latina. Para cada caso se determinaron de manera general los criterios de inclusión de información, contenido (tipo de registros), tipo de acceso, cobertura (representación) tipológica, temática, geográfica, temporal y disponibilidad.

Con base en los resultados se eligieron las seis bases de datos bibliográficas con mayor representación de artículos en sistemática sobre América Latina que fueran susceptibles a realizar análisis bibliométricos, es decir, que las aplicaciones permitieran recuperar, extraer y analizar la información cuantitativamente. Estas seis fueron: Periódica (DGB-UNAM, 2008), *Science Citation Index Expanded* (SCI) desde 1900 (Thomson Reuters, 2008b),

Biosis Previews (Thomson Reuters, 2008a), *Zoological Record* (ZR) (Thompson Reuters, 2008), CAB (CABI, 2008b) y *Scopus* (Elsevier, 2008b). Se hicieron varias pruebas para determinar los campos, términos y operadores que se aplicarían en la búsqueda para cada base de datos, seleccionando documentos sobre sistemática y la restricción a los países de América Latina. Se extrajeron los documentos recuperados y se diseñó una base de datos relacional de acuerdo con la estructura lógica de toda la información extraída, se migró la información, depuró, normalizó y validó. Posteriormente se realizaron las consultas y construcción de matrices de datos que se utilizaron para hacer los análisis estadísticos y figuras correspondientes en Excel de *Microsoft Office*®.

Para comparar el contenido de las bases de datos se utilizaron todas las revistas que publicaron sobre sistemática de América Latina contenidas en Periódica (411), SCI (460) y *Scopus* (160), para el caso de *Biosis*, ZR y CAB se eligieron las 500 revistas que publicaron más artículos, de las cuales el 91, 58 y 79% correspondieron al total de documentos registrados para cada una. La visualización del análisis de redes se hizo con el programa Pajek (Batagelj y Mrvar, 2008), cada nodo corresponde a una base de datos y el tamaño es proporcional a la cantidad de revistas totales que publicaron sobre sistemática de América Latina, las líneas representan la cantidad de revistas que comparten entre ellas. Se aplicó el algoritmo de Kamada-Kawai, 1989.

Se llevó a cabo un análisis de parsimonia para las seis bases de datos con una matriz de 1.363 caracteres (títulos de revistas), donde para cada celda se identificó la ausencia (0) o presencia (1) de cada revista (se puede solicitar a la autora). La matriz se analizó con los programas Nona 2.0 y WinClada 1.00.08, empleando una búsqueda heurística con multiple TBR+TBR.

RESULTADOS Y DISCUSIÓN

En cinco de las seis bases de datos analizadas existe una categoría correspondiente a sistemática. En *Biosis* se refiere al campo *major concepts*: “*Systematics and Taxonomy*”, en CAB es el campo cabicode “ZZ380: *Taxonomy and Evolution*” (CABI, 2008a), en ZR aparece como campo *subject descriptors*: “*Systematic*”, y en Periódica es el campo tema: “Taxonomía y Sistemática” (DGB-UNAM, 2008). Para SCI la única forma de identificar los temas es a través del campo *subject*, que es asignado por igual a todos los documentos de cada revista. En la versión en línea no es posible identificar las palabras clave o algún otro descriptor para cada documento, en este rubro no existe alguna categoría correspondiente a sistemática, de tal manera que la búsqueda se realizó a partir de título y *abstract*. En *Scopus* sucedió algo parecido, pero en este caso se utilizaron título, *abstract* y *keywords*. Por estas razones, las búsquedas en SCI y *Scopus* fueron las más complejas. En la tabla 1 se presentan las características relevantes de cada base de datos analizada. Se pueden consultar los documentos y las revistas totales registrados para cada una de ellas, así como los obtenidos para las referentes a sistemática sobre América Latina. Todas las bases de datos revisadas contienen al menos una proporción del 85% de artículos de investigación incluidos en las bases. Las revisiones, memorias de congresos, libros y capítulos estuvieron pobremente representados. En cuanto a la cobertura de cada base de datos, entendida como la proporción de documentos sobre sistemática encontrados respecto al total fueron, en orden descendente: Periódica 4,2%, ZR 3,5%,

Biosis 0,63%, CAB 0,28%, SCI y *Scopus* tuvieron una representación de milésimas. En cuanto a la cantidad neta de registros en el área, ZR contiene 63.421 registros, tres veces más que Biosis con 19.079, CAB 14.363, Periodica 11.185, SCI 3.257 y *Scopus* con 916 documentos.

Base de Datos	Biosis Previews	Zoological Record	Science Citation Index	Scopus	CAB	Periódica
Información	Registro bibliográfico	Registro bibliográfico	Registro bibliográfico, referencias, citas e indicadores bibliométricos	Registro bibliográfico	Registro bibliográfico, referencias, citas e indicadores bibliométricos	Registro bibliográfico
Cobertura geográfica	Internacional	Internacional	Internacional	Internacional	Internacional	América Latina
Cobertura temática	Ciencias biológicas y biomédicas	Zoología	Ciencia	Ciencia, tecnología, medicina y ciencias sociales	Agricultura y medio ambiente	Ciencia y tecnología
Cobertura temporal	1969- actualidad	1864- actualidad	1900- actualidad	1996- actualidad	1910- actualidad	1975- actualidad
Acceso	Restringido	Restringido	Restringido	Restringido	Restringido	Libre
Registros totales	33 millones	1.8 millones	33 millones	3.4 millones	5 millones	265 mil
Documentos taxonomía	19.079 (91,5% artículos)	63.421 (97,7% artículos)	3.257 (93,6% artículos)	916 (98,8% artículos)	14.326 (88,5% artículos)	11.185 (84,5% artículos)
Revistas totales	5.500	5.000	6.600	15.000	9.000	1.500
Autores	22.225	33.177	6.073	159	13.500	11.510

Tabla 1. Características relevantes de las seis bases de datos bibliográficas (principalmente revistas) sobre sistemática de América Latina.

* Se refiere a títulos de revistas, series, memorias, etc.

Respecto a las bases de datos internacionales, la compañía Thomson Reuters mantiene tres de las investigadas en este trabajo; 1) *Zoological Record*, especializada en animales, que tiene la mayor cantidad de registros y la mayor cobertura temporal; 2) Biosis, la de mayor información en la actualidad y 3) *Science Citation Index (SCI)*, la base de datos multidisciplinaria más utilizada, porque presenta herramientas sofisticadas de recuperación y análisis de información, cubre revistas de corriente principal (alto impacto) y tiene criterios estrictos de selección de los títulos (Thomson Reuters, 2008d). *Scopus* es un proyecto reciente y comparte las características mencionadas para SCI, pero incluye mayor cantidad de revistas (Elsevier, 2008a; Thomson Reuters, 2008d). SCI y *Scopus* son bases de datos bibliográficas multidisciplinarias, internacionales, incluyen literatura de corriente principal (representativa, con alto índice de impacto, publicada principalmente en inglés), registran las referencias de los documentos e incluyen análisis de citas (De Moya-Anegón *et al.*, 2007). Estas bases de datos registraron pocos artículos sobre sistemática latinoamericana, lo que indica el bajo impacto de este tema en las tendencias científicas contemporáneas.

En la figura 1 se muestra la cobertura temporal de documentos sobre sistemática de América Latina para cada una de las bases analizadas. CAB y ZR son las bases de datos más antiguas y con mayor cobertura temporal, representan la tendencia de publicación sobre el tema durante casi un siglo. En la figura se nota un aumento en la producción, aunque no es significativo, los cambios drásticos se deben al incremento de la cantidad de revistas incluidas en los sistemas de información y coinciden con su origen o desarrollo. El promedio de artículos publicados por año para Biosis fue de 6.417, para ZR 433, para Periódica 361, para CAB 145, para SCI 91 y para *Scopus* 30. La primera base incorpora un orden de magnitud mayor de registros sistemáticos por año que las demás, entre el 2000 y el 2004 registró la mayor cantidad de documentos. El máximo fue en el último año, con 11924, después la cifra disminuyó en más de 1.000. Esta disminución coincide con el año en el que Biosis fue adquirida por Thompson Reuters. Por el contrario ZR, incluida en el mismo contrato con Biosis tuvo el máximo de documentos indizados en el 2004 con 1595.

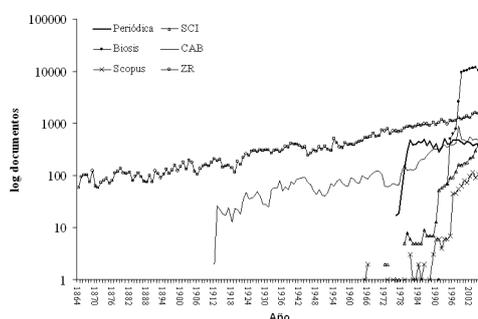


Figura 1. Cobertura temporal de los documentos sobre sistemática de América Latina para las seis bases de datos analizadas.

La identificación y comparación de las revistas de cada base de datos fue un proceso complicado, en especial los títulos latinoamericanos. Esto podría deberse a que en

muchos casos las revistas de esta región tienen nombres largos. Además al estar escritos en un idioma extraño para los referencistas, no reconocen las revistas similares y les asignan abreviaturas distintas. En los registros actuales los nombres de las revistas fueron más homogéneos, en muchos casos estaban escritos de formas distintas, lo cual fue frecuente en ZR y CAB, las más antiguas, y al parecer, fueron las menos normalizadas. La falta de homogeneidad de la información es una de las dificultades más frecuentes en los análisis de información de bases de datos (Hood y Wilson, 2003b). En total se identificaron 1.363 revistas distintas en las seis bases de datos, 865 (63,5%) estuvieron presentes en una sola base, del resto, 238 (17,5%) aparecieron en dos, 131 (9,6%) en tres, 80 (5,9%) en cuatro, 47 (3,4%) en cinco y solo dos, la Revista de Biología Tropical y Neotropical Entomology estuvieron incluidas en las seis bases de datos. La cantidad de revistas sobre sistemática de América Latina contenidas en cada base de datos y la cifra de revistas que comparten entre cada una de ellas (líneas) se pueden apreciar en el análisis de redes de la figura 2. El tamaño de los nodos es proporcional al número de revistas encontradas. En el diagrama se nota la similitud entre ellas con base en la cantidad de revistas que comparten. Como era de esperarse, SCI, Biosis y ZR aparecen juntas, porque comparten muchas revistas, pues son mantenidas por la misma compañía. CAB, Biosis y ZR también tienen una correlación fuerte por su analogía en la cobertura temática afín a las ciencias biológicas. *Scopus* y SCI se reúnen porque son multidisciplinarias, registran las revistas más representativas y relevantes (de corriente principal) e incluyen el registro de las referencias y citas. Periódica, como es natural, fue el conjunto con mayor número de revistas distintas, debido a su cobertura geográfica restringida a América Latina. Las relaciones más estrechas se dan entre Biosis y SCI, que comparten al menos 16% de títulos; CAB y Biosis comparten el 15%; Biosis y ZR el 14% y CAB y ZR el 11%.

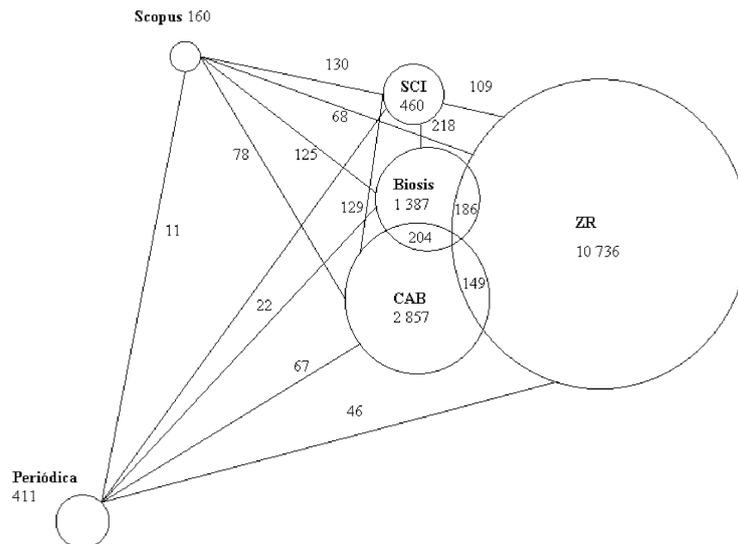


Figura 2. Análisis de redes de las seis bases de datos. Los nodos son las bases de datos y los vértices las revistas que comparten.

Del análisis de parsimonia se obtuvo un solo cladograma (Fig. 3), con 896 pasos, índice de consistencia de 0,55 e índice de retención de 0,43. Este análisis permitió reconocer a las revistas compartidas por las bases de datos, los números corresponden a las sinapomorfias (revistas que comparten las bases de datos) y las autapomorfias (revistas que se encuentran únicamente en cada base de datos). Periódica tiene la cifra mayor exclusividad, con 285 revistas. En el cladograma obtenido se identifican los grupos principales a los que corresponde cada base de datos de acuerdo con 1) su cobertura geográfica: regionales (una base de datos), con preponderancia del idioma español, e internacionales (cinco bases de datos) principalmente en inglés (Fig. 4), donde la mayor proporción de artículos en inglés la tuvo SCI; 2) el tipo de revistas que consideran: las de corriente principal, multidisciplinarias con altos índices de impacto y estándares editoriales estrictos, y el resto, cuya finalidad es tener una representación temática específica y detallada (Fig. 5); 3) el acceso libre o restringido previo pago, donde la única base de datos con consulta gratuita fue Periódica; y 4) las tres bases de datos mantenidas por Thomson Reuters (desde 1958) no conforman un grupo natural, pues esta compañía adquirió en el 2004 a Biosis (instituida desde 1926), que en 1980 se había fusionado con ZR (fundada en 1861 por la Zoological Society of London; Neave, 1950), lo cual explica que en la actualidad compartan muchas revistas.

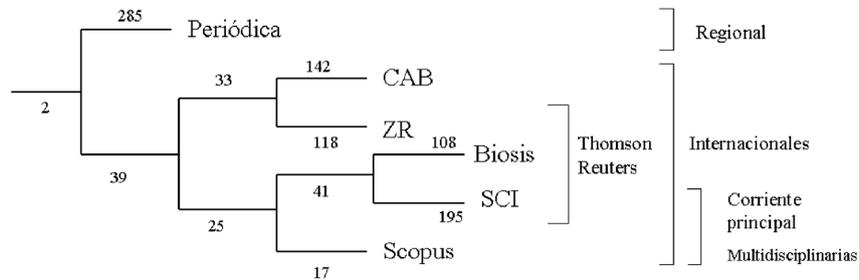


Figura 3. Cladograma de las seis bases de datos.

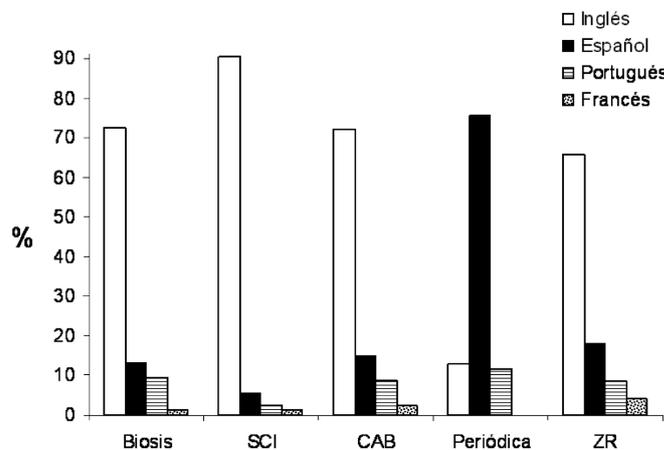


Figura 4. Cobertura idiomática de las bases de datos.

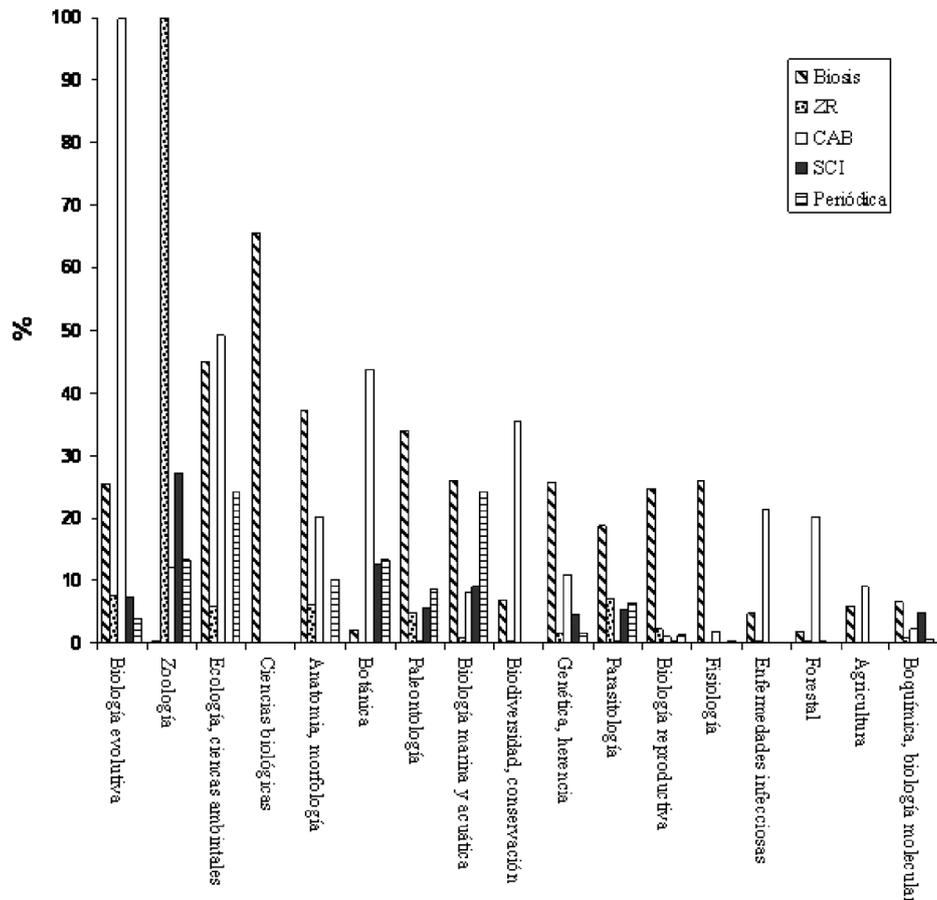


Figura 5. Cobertura temática de las bases de datos con base en las revistas. El tema se refiere al tema (*subject*) asignado a las revistas (Thomson Reuters 2008e). No incluye en el análisis Periódica porque los temas se asignan a los documentos, no a las revistas, además las categorías son distintas.

Entre los artículos taxonómicos, son de especial interés aquellos que presentan el descubrimiento y las descripción de especies nuevas, una de las tareas básicas de la sistemática (Mayr, 1969). Las bases de datos que registraron mayor proporción de artículos con descripciones de especies nuevas sobre América Latina fueron Periódica con 24,6%, CAB con 54,6% y ZR con 72,9%. ZR publicó en promedio 322 documentos al año con descripciones de especies nuevas, Periódica 90 y CAB 82, la primera contiene más del doble de documentos que las otras dos (Fig. 6).

De acuerdo con los resultados obtenidos, ninguna de las bases de datos cumple con las características suficientes para ser utilizada como única fuente de información. Por lo tanto, lo conveniente será utilizar distintas combinaciones de acuerdo con las necesidades informativas. Por ejemplo, Biosis tiene la información más exhaustiva sobre cualquier taxón actual, con el inconveniente de que su cobertura temporal es pequeña. ZR y CAB tienen una cobertura temporal amplia e incluyen principalmente animales o

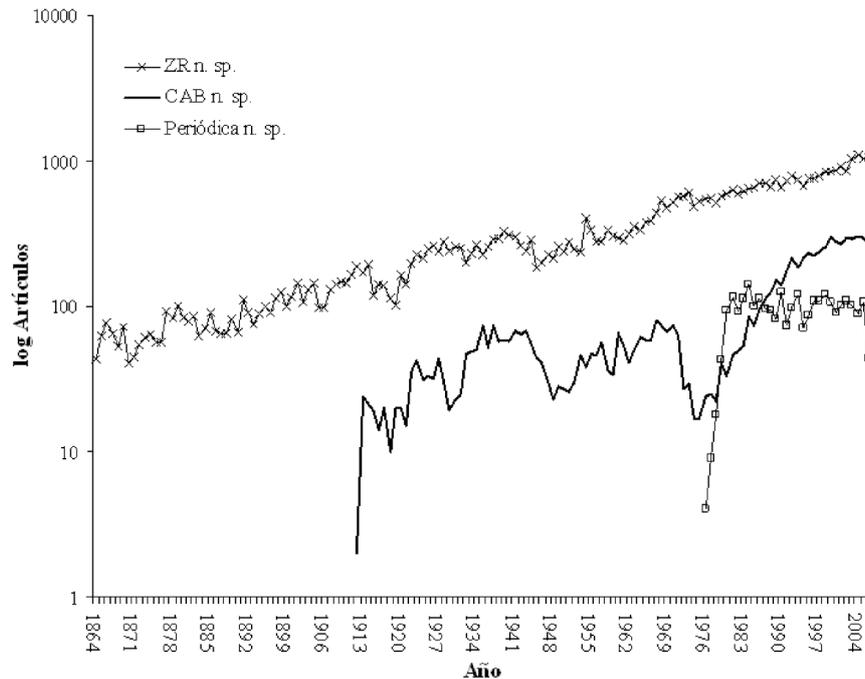


Figura 6. Tendencia temporal de documentos con descripciones de nuevas especies.

plantas, respectivamente. SCI es la base de datos con mayor número de documentos en revistas de corriente principal, con referencias, citas e indicadores bibliométricos. Periódica es la base de datos con mayor cantidad de documentos y cobertura temporal más amplia sobre lo publicado en la zona. Con excepción de Periódica, las demás bases de datos permiten recuperar, manejar y analizar grandes cantidades de registros. Para tener un universo representativo de artículos se debería utilizar al menos dos bases de datos, una internacional especializada en el taxón de interés como CAB (para botánica), ZR (para zoología) o Biosis (general) y la de cobertura latinoamericana, Periódica. SCI o *Scopus* servirían para identificar las tendencias actuales e ideas innovadoras en el tema. Por ejemplo, para hacer una búsqueda sobre sistemática de angiospermas latinoamericanas en los últimos 20 años sería preciso usar al menos dos bases de datos como Biosis y Periódica. Si se quisiera estudiar un periodo anterior, entonces sería más conveniente usar CAB y Periódica.

El cladograma resultó una herramienta de visualización óptima para representar las características principales de las bases de datos, utilizada con la figura 5 proporcionan una guía fácil y rápida para consultar en qué casos se puede utilizar determinada base de datos. Los resultados de este análisis corroboran que la literatura sobre sistemática biológica está ampliamente dispersa y poco sistematizada. Esta es, en general, una limitante importante para la investigación en esta área, pues es difícil que una sola institución pague el acceso a todas las bases de datos, siendo estas elegidas generalmente de acuerdo con el presupuesto y el perfil académico requerido. Por ejemplo, la Universidad Nacional Autó-

noma de México cuenta actualmente con suscripciones a *Biological Abstracts* (un subconjunto de Biosis) desde 1992, a ZR de 1978 en adelante, a CAB desde 1973 a la actualidad, a *Science Citation Index* a partir de 1995 y a *Scopus*. Esto permite acceder a una cantidad y diversidad de bases de datos con información bibliográfica sobre ciencias biológicas suficiente desde el punto de vista bibliotecológico. Sin embargo, para las investigaciones sistemáticas cuya literatura de referencia es extensa, no caduca y puede estar publicada en cualquier formato, como en boletines o reportes de escasa distribución, la representación es claramente insuficiente. Además, hay que considerar que la literatura analizada en este trabajo representa solo una pequeña proporción de la existente, pues en estas bases de datos se consideran principalmente revistas, el resto de los documentos tiene escasa presencia. En la actualidad estoy realizando una investigación sobre las bases de datos disponibles con literatura sobre taxonomía que incluya cualquier tipo de documento.

El desarrollo de las bases de datos ha fomentado la evolución de la manera cómo se sistematiza, analiza y comparte la información, lo cual ha impulsado la evolución de disciplinas tradicionales como la taxonomía en campos emergentes como la cibertaxonomía (Hine, 2006; Hine, 2008; Wheeler, 2007), también denominada Informática de la Biodiversidad (Soberón y Peterson, 2004; Johnson, 2007; Sarkar, 2007). Este nuevo enfoque propone métodos innovadores para compartir, usar e incluso publicar información taxonómica (Godfray *et al.*, 2007; Zhang, 2008), pero la aplicación de estas innovaciones dependerá, en gran medida, de que se cuente con literatura disponible en formato digital, sistematizada, completa, consistente, actualizada, inmediata y pública, que pueda manejarse, recuperarse y analizarse a gran escala y de manera sistémica. Una de las tendencias actuales en la informática biológica es la obtención de nuevo conocimiento a partir del análisis de documentos aplicando técnicas como *text mining* o el análisis de redes que permiten correlacionar genes o proteínas, en boga actualmente en la bioinformática e informática médica. En un futuro también se podrían aplicar estos enfoques para el estudio sistemático a partir de textos completos. Este es, sin dudas, un reto de la sistemática actual.

Sería importante acelerar, en forma simultánea a la sistematización y digitalización de información biológica latinoamericana (especímenes y nombres), la indización de documentos taxonómicos, en especial los asientos bibliográficos (información básica sobre el documento y lugar en el que está archivado), y con base en esta información, decidir cual tiene prioridad para digitalizarse. Al respecto no hay programada iniciativa alguna a gran escala. Están en marcha proyectos interesantes, a nivel internacional, uno es *Biodiversity Heritage Library*, que tiene registrados y digitalizados únicamente 5.146 documentos (BHL, 2008), pero tiene la desventaja que la aplicación dificulta la recuperación de registros y el análisis de grandes cantidades de información, característica indispensable en la investigación bibliográfica del siglo XXI. Sería interesante considerar la realización de una base de datos de literatura taxonómica sobre América Latina de acceso libre, que incluya la literatura internacional y regional, un buen principio podría ser integrar la información existente. Este proyecto sería de impacto en la práctica sistemática en la región. No existe iniciativa alguna sobre sistematizar en una base de datos bibliográfica la taxonomía de alguna región, aunque existen varios intentos interesantes por abarcar la literatura de determinados grupos.

Los resultados obtenidos en este artículo también proporcionan información interesante

para realizar análisis bibliométricos (Frandsen y Nicolaisen, 2008). Por ejemplo, para estudiar la sistemática realizada en América Latina durante los últimos 30 años, se puede utilizar la información de las seis bases de datos aquí presentadas para tener un panorama general del desarrollo de esta disciplina (Michán y Llorente-Bousquets, 2010).

AGRADECIMIENTOS

A Juan José Morrone y Tania Escalante por la asesoría para la realización del cladograma, por la discusión de algunas ideas y por la revisión del manuscrito. A Judit Aguirre por la lectura y sugerencias al manuscrito. A Antonio Sánchez Pereyra y Antonia Llorens Cruset, del Departamento de Bibliografía Latinoamericana, Dirección General de Bibliotecas-UNAM, por la extracción de la información de Periódica Taxonomía-sistemática. A Minerva García Palacios por su asesoría en cómputo. A Jorge Llorente-Bousquets, Armando Luis, Isabel Vargas, Programa PROFIP- DGAPA UNAM, a la Library of Congress de los Estados Unidos de América y a Shaday Michán por el apoyo que me brindaron para la realización de la investigación. Este trabajo se realizó con fondos de los proyectos PAPIME PE 201509 y CONACYT, Ciencia Básica Proyecto 13276.

BIBLIOGRAFÍA

- BATAGELJ V, A MRVAR. Pajek, Program for large network analysis. 2008. (Consultada: <http://pajek.imfm.si/doku.php?id=download>).
- BHL. Biodiversity Heritage Library. 2008. (Consultada: 3 de junio de 2008, <http://www.biodiversitylibrary.org/>).
- CABI CAB. CAB codes. 2008a. (Consultada: 25 de mayo de 2008, <http://www.cabi.org/guides/cabicodes.htm>).
- CABI CAB. CAB Abstracts. 2008b. (Consultada: 15 de abril de 2008, <http://www.cabi.org/datapage.asp?iDocID=165>).
- DE MOYA ANEGÓN F, CHINCHILLA RODRÍGUEZ Z, VARGAS QUESADA B, CORERA ÁLVAREZ E, MUÑOZ FERNÁNDEZ FJ, *et al.* Coverage analysis of Scopus: A journal metric approach. *Scientometrics*. 2007;73(1):53-78.
- DGB-UNAM. Clasificación utilizada por las bases de datos CLASE y PERIÓDICA organizada por temas y subtemas. 2008. (Consultada: 27 de mayo de 2008, http://132.248.9.1:8991/F/T1PUR4JLCJ147F8T336EVRUE3ML3D9F6B49PVC4BPV3E GE21UD-00188?func=filefile_name=clasificacion-per01).
- ELSEVIER BV. Scopus in Detail: What does it cover? 2008a. (Consultada: <http://www.info.Scopus.com/detail/what/>).
- ELSEVIER BV. Scopus. 2008b. (Consultada: 3 de junio de 2008, <http://www.Scopus.com/Scopus/home.url>).
- FRANSEN TF, NICOLAISEN J. Intradisciplinary differences in database coverage and the consequences for bibliometric research. *J Am Soc Inform Sci*. 2008; 59(10):1570-1581.
- GODFRAY HCJ, CLARK BR, KITCHING IJ, MAYO SJ, SCOBLE MJ. The Web and the structure of taxonomy. *Syst Biol*. 2007;56 (6):943-955.
- HAWKINS DT. The literature of online information-retrieval - an update. Online

Rev. 1984;8(2):153-164.

HINE C. Databases as scientific instruments and their role in the ordering of scientific work. *Soc Stud Sci.* 2006;36(2):269-298.

HINE C. *Systematics as Cyberscience: Computers, Change, and Continuity in Science.* Boston, MIT; 2008.

HOOD WW, WILSON CS. The scatter of documents over databases in different subject domains: How many databases are needed? *J Am Soc Inform Sci.* 2001;52(14):1242-1254.

HOOD WW, WILSON CS. Overlap in bibliographic databases. *J Am Soc Inform Sci.* 2003a;54(12):1091-1103.

HOOD WW, WILSON CS. Infometric studies using databases: Opportunities and challenges. *Scientometrics.* 2003b;58(3):587-608.

JOHNSON NF. Biodiversity informatics. *Annu Rev Entomol.* 2007;52:421-438.

KAMADA T, KAWAI S. An algorithm for drawing general undirected graphs. *Inform Process Lett.* 1989;31:7-15.

KRELL FT. Why impact factors don't work for taxonomy. *Nature.* 2002;415:957.

MAYR E. *Principles of systematic zoology.* New York, McGraw-Hill; 1969.

MICHÁN L, LLORENTE-BOUSQUETS J. Bibliometría de la sistemática biológica sobre América Latina durante el siglo XX en tres bases de datos mundiales. *Rev Biol Trop.* 2010;58(2):531-545.

NEAVE SA. Concerning the Zoological Record. *Science.* 1950;112(2921):761-762.

NEUFELD ML, CORNOG M. Database history - from dinosaurs to compact disks. *J Am Soc Inform Sci.* 1986;37(4):183-190.

NSF. *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century.* National Science Foundation. 2005.

SARKAR IN. Biodiversity informatics: Organizing and linking information across the spectrum of life. *Briefings in Bioinformatics.* 2007;8:347-357.

SOBERÓN J, PETERSON AT. Biodiversity informatics: Managing and applying primary biodiversity data. *Philos T Roy Soc B.* 2004;359(1444):689-698.

THOMPSON REUTERS. *Zoological Record.* 2008. Disponible en: <http://scientific.thomson.com/products/zr/>

THOMPSON REUTERS. *Biosis Previews.* 2008a. (Consultada: 21 de abril de 2008, <http://www.Biosis.org/>).

THOMPSON REUTERS. *Dialog Bluesheets.* 2008b. (Consultada: 15 de mayo de 2008, <http://library.dialog.com/bluesheets/html/bli.html>).

THOMPSON REUTERS *Gale Directory of Online, Portable, and Internet Databases.* 2008c. (Consultada: 15 de mayo de 2008, <http://library.dialog.com/bluesheets/html/bi0230.html>).

THOMPSON REUTERS *The Thomson Scientific Journal Selection Process.* 2008d. (Consultada: <http://scientific.thomsonreuters.com/free/essays/selectionofmaterial/journalselection/>).

THOMPSON REUTERS. *Science Citation Index Expanded Subject Categories.* 2008e. (Consultada: <http://scientific.thomsonreuters.com/cgi-bin/jrnlst/jlsubcatg.cgi?PC=D>).

WHEELER QD. Invertebrate systematics or spineless taxonomy? *Zootaxa.* 2007;1668:11-18.

ZHANG ZQ. Zoological taxonomy at 250: Showcasing species descriptions in the cyber era. *Zootaxa.* 2008;(1671):1-2.