

Enfoques metodológicos para identificar y caracterizar la investigación mexicana en química en bases de datos bibliográficas

Yoscelina Hernández García *
Mina Kleiche Dray **
Jane M. Russell ***

Artículo recibido:
4 de septiembre de 2012.
Artículo aceptado:
19 de octubre de 2012.

RESUMEN

Las distintas visiones que dan las diferentes fuentes de información con respecto a un campo de investigación en particular es un fenómeno conocido en los estudios bibliométricos y cuantitativos. El presente artículo tiene por propósito entender el impacto de las fuentes de información en la identificación y la caracterización de un campo científico en un contexto particular: la química en México. No se trata de comparar la producción científica en este campo extraída de varias bases de datos bibliográficas, sino de partir de una metodología original que consiste en construir estrategias de búsqueda

* Programa de Posgrado en Bibliotecología y Estudios de la Información, FFyL/UNAM, México. yosce2336@gmail.com

** Université Paris I-Sorbonne, Francia. Mina.Kleiche@ird.fr

*** Instituto de Investigaciones Bibliotecológicas y de la Información de la UNAM, México. jrussell@unam.mx

adecuadas para cada fuente y procesos de normalización, análisis y visualización de los resultados, mostrar que la definición del campo científico estudiado depende de las fuentes de información usadas. Se usaron fuentes nacionales e internacionales especializadas, o relacionadas con las ciencias químicas, así como recursos multidisciplinarios, a saber *Web of Science* (WoS), *Chemical Abstracts Service* (CAS), *PERIÓDICA* y *CAB Abstracts* (CAB) en el periodo de 2000-2004. Los resultados apuntan al impacto de las especificidades de la estructura de cada base de datos en los resultados de las búsquedas; en *PERIÓDICA*, la producción se concentra en 37 revistas (44%) mientras que los trabajos en revistas internacionales se encuentran repartidas en un total de 1559 publicaciones (WoS), 188 (CAB) y 1640 (CAS).

Palabras clave: Química; México; Bases de datos bibliográficas.

ABSTRACT

Methodological approaches to identify and describe Mexican chemistry research in bibliographic databases

Yoscelina Hernández-García, Mina Kleiche-Dray and Jane M.-Russell

The varying visions of a particular research field in diverse information sources are a recognized phenomenon in bibliometric and scientometric studies. The purpose of this paper is to provide an understanding of the impact of using distinct information sources for the identification and description of a scientific field in a specific context, in this instance the field of chemistry in Mexico. We use both international and national sources specializing in or related to chemistry, as well as suitable multidisciplinary services, including *Web of Science* (WoS), *Chemical Abstracts Service* (CAS), *PERIÓDICA* and *CAB Abstracts* (CAB) for 2000-2004. Using an original methodology involving the construction of appropriate search strategies for each database and processes for the normalization, the authors compare and analyze the scientific production extracted from the distinct bibliographic databases, showing that the definition of the scientific field under

study depends on the information source consulted. Our findings reveal the impact of a given database structure on search results. In the *PERIÓDICA* index, scientific articles are concentrated in 37 journals (44%), while in international journals it is distributed across a large number of titles, 1559 (WoS), 188 (CAB) and 1640 (CAS).

Keywords: Chemistry; Mexico; Bibliographic databases.

INTRODUCCIÓN

Las distintas representaciones de un tema de investigación en diferentes fuentes es un fenómeno reconocido en la literatura especializada, particularmente desde que Bradford la expresó en términos matemáticos a través de su ley de dispersión de la literatura científica. Esta ley tiene impacto tanto en la recuperación de información bibliográfica, como en el mundo de la bibliometría y de la cienciometría; la diferencia radica en los objetivos de las indagaciones. En el caso de la búsqueda de literatura sobre un tema en particular, la principal consideración es la proporción de duplicación de artículos y revistas entre el conjunto de fuentes seleccionadas. En cambio, la exploración bibliométrica busca crear diferentes visiones de un campo de estudio proporcionadas por cada una de las fuentes. No obstante, en los dos casos el especialista enfrenta el mismo problema: la selección de fuentes más representativas para después desarrollar estrategias de búsqueda de acuerdo con las características y posibilidades de cada una de ellas con distinta cobertura, frecuentemente montadas en diferentes plataformas tecnológicas bajo estructuras ontológicas distintas y expresadas en más de un solo idioma.

Varios son los autores que han estudiado esta dimensión de la literatura científica enfocándose a diferentes campos del conocimiento. Costas *et al.*, 2008, en ciencias de la salud, encontraron importantes diferencias en los resultados obtenidos de las diferentes fuentes seleccionadas, principalmente debido a la inclusión de un solo lugar de trabajo en *MEDLINE* en comparación con el multidisciplinario *Web of Science* (WoS), lo cual suscitó una mayor recuperación de trabajos de la producción científica de la región de Aragón en España en el WoS que en la fuente especializada. Más recientemente, Bolaños-Pizarro *et al.* (2009), en un análisis de las publicaciones sobre el uso de diferentes bases de datos, concluyeron que al tratarse de una disciplina emergente, la falta

de una terminología normalizada dificultó el desarrollo de las estrategias de búsqueda. La importancia de la recuperación sistemática de información para las profesiones relacionadas con la práctica basada en evidencias fue manifestada por Taylor *et al.*, en 2007. Ellos aplicaron una solicitud de información específica y claramente definida a siete bases de datos que proveen resúmenes de interés a trabajadores sociales, para determinar las ventajas y desventajas de cada una de ellas. También llegaron a la conclusión de lo esencial, que es un consenso internacional sobre la terminología profesional para lograr una óptima recuperación de documentos. Pulgarín Guerrero y Escalona Fernández (2007), al comparar tres bases de datos con contenidos similares sobre ingeniería para determinar el índice de duplicación con el fin de evitar gastos superfluos, identificaron una gran cantidad de fuentes únicas en cada una de ellas. Concluyeron que el uso de cualquiera de éstas en forma individual resulta en la pérdida de un número significativo de registros.

Entre las dificultades inherentes en los estudios de interrelación entre fuentes de información, Cañedo Andalia (1999) menciona la existencia de una serie de preferencias geográficas, idiomáticas, culturales, políticas y de otro tipo que caracterizan individualmente a cada base. En el caso de la literatura producida en las regiones del mundo en desarrollo, escasamente escrita en inglés o recogida por las grandes bases de datos internacionales, esta característica es de suma importancia.

Otros factores que influyen en cuestiones de la representación de información especializada en distintas fuentes que comentan diferentes autores son: la frecuencia de actualización de las bases de datos tomando el tema de la recuperación de información como caso de estudio (Salvador Oliván *et al.*, 1999) y la interdisciplinariedad de campos, como es el de la Bibliotecología y Ciencias de la Información que implica la distribución de información pertinente en una amplia gama de bases de datos, incluso en aquellas que tratan temas tradicionalmente poco asociados con el campo (Yerkey y Glogowski, 1990).

Debido a la proliferación en el número de bases de datos electrónicas, Hood y Wilson se preguntaron en 2001 sobre el número de éstas que era requerido para cubrir satisfactoriamente una necesidad de información. En particular, enfatizaron la falta de información que existe sobre la distribución de los documentos significativos entre un amplio rango de bases de datos. Los resultados confirmaron estudios anteriores y constataron que para alcanzar una recuperación adecuada de registros sobre un tema en particular, es necesario indagar en múltiples fuentes. En un estudio posterior estos mismos autores encontraron que cerca del 63% de registros son únicos para una base en particular y que el 37% restantes están duplicados en de dos a 12 diferentes bases de datos (Hood y Wilson, 2003).

Un estudio previo sobre la investigación química en México señaló las diferencias en los resultados que puede arrojar el uso de distintas bases de datos, como por ejemplo una mayor presencia en bases como el *Science Citation Index* que recoge las direcciones de todos los autores, que en el *Chemical Abstracts* que recoge únicamente la dirección del primer autor. Otra característica de las bases que resulta determinante fue el tipo de fuente considerada para inclusión en la base (Russell *et al.*, 1995).

En este trabajo presentamos la metodología de análisis bibliométrico construido en el marco del programa internacional de investigación (Institut de Recherche pour le Développement, IRD, Francia y Universidad Nacional Autónoma de México, UNAM, México) en torno de la evolución de la Química en México en el siglo XX (Kleiche-Dray, 2008; Kleiche-Dray y Casas Guerrero, 2008) donde se persigue un objetivo triple: a) Construir un tema de investigación *per se*, que proporcione un panorama en conjunto del campo de estudio –en este caso, la química– que nos permita conocer la composición institucional de este terreno de investigación, así como su volumen, el modo y la estructuración de su producción científica; b) Apoyar la metodología histórica del programa en curso, al proporcionar un panorama de la producción de la comunidad mexicana de químicos que nos permitirá situarla en relación con los temas científicos y en el contexto económico, político y social del México de hoy, y c) Darle mayor pertinencia al trabajo sociológico de campo: sus resultados, cruzados con los resultados de los censos de sitios y de actores, nos permitirán entonces seleccionar una muestra razonada de laboratorios y de actores por encuestar.

Nuestro trabajo no se proponía ser exhaustivo en la recuperación de la información, sino tratar de lograr precisión en el entorno local de la investigación química en México y, al estar inserto en un proyecto más general, complementar la visión que se estudia por otros medios.

Partimos de la necesidad de seleccionar un conjunto de bases de datos que respondieran a nuestros objetivos, que incluían tanto las especializadas en química, como las que cubren campos relacionados, y las multidisciplinarias, que cuentan con distintas coberturas geográficas, para luego elaborar una metodología que contemplara un estudio minucioso de la estructura general y la funcionalidad de cada base; se les da un especial énfasis a las estrategias de búsqueda que permiten la extracción de elementos bibliométricos y su normalización de manera retrospectiva, de 2000 a 2009. En una primera etapa y para tener un acercamiento a los instrumentos y la metodología, se trabajó con los datos hasta 2004, siendo el contenido del presente documento el resultado. En este sentido la actualización para las siguientes etapas del estudio se hizo complicada en el caso por ejemplo de *PERIÓDICA*

debido a problemas con la actualización de la base, por lo que la información que se obtuvo de 2000 a 2004 también era parte de una estrategia para darle continuidad a los análisis pertinentes, en función de lo que las bases de datos bibliográficas nos permiten hacer.

Este tratamiento de los datos nos permitirá alcanzar un análisis cuantitativo transversal que identifique los diferentes elementos que en una etapa posterior del estudio bibliométrico permitirán construir un panorama detallado de la investigación química en México durante la primera década del presente siglo.

En particular esta primera etapa se enfocó en reconocer las diferencias de cada una de las cuatro bases de datos, para a partir de la lógica imperante en el documento o la del instrumento de consulta, establecer luego las estrategias de búsqueda que permitan extraer la información que representan los trabajos en química con al menos una dirección de adscripción de México. De esta manera, se podrán confrontar los resultados de cada una de las cuatro bases de datos a través de las diferentes visiones que brinde cada una de ellas.

Un reto importante de este estudio fue poder definir la química de acuerdo con nuestros objetivos, fijar sus fronteras disciplinarias y traducir éstas en términos del lenguaje empleado por cada una de las bases. La estrategia habitual es apoyarse en los títulos de revistas clasificadas en las diferentes grandes áreas de la química.

En nuestro caso este enfoque nos situó ante dos grandes problemas, en primer lugar la única base seleccionada que presenta esta opción es el WoS, lo que imposibilitaba nuestra condicionante de aplicar estrategias lo más similares posible en las búsquedas de las diferentes bases. En segundo lugar, asignarle campos disciplinarios a artículos a través de una clasificación general de revistas produce una visión poco completa de la producción científica de un campo en particular (Neuhaus y Daniel, 2009). Frente a estas limitantes nos centramos en estrategias basadas en la dirección de la adscripción del autor; de esta forma la definición de la disciplina química era dada por la institución que acogía la investigación, y no por los descriptores temáticos asignados al texto indizado, o a la revista donde éste se publicaba. La dirección de adscripción es común a todos los instrumentos de consulta seleccionados, con la misma estructura y falta de normalización, aunque no proporcionada de la misma forma.

Las estrategias alternativas que usamos fijaron los límites de los registros que se recuperaron, a partir de la previa identificación de las instituciones que se reconocen como las que cultivan la investigación en las diferentes áreas químicas.

Se espera que los resultados de este trabajo servirán no únicamente para concebir diferentes visiones de la investigación química realizada en México sino que también contribuirán al avance de los aspectos metodológicos de la bibliometría, principalmente en lo que se refiere al conocimiento sobre la estructura de las cuatro bases de datos que emplean como fuentes de datos, las alternativas de búsqueda y las diferentes formas de recuperación y manejo de los registros.

MATERIAL Y MÉTODOS

Fuentes de datos

La originalidad del estudio reside en primer lugar en la selección de las bases de datos bibliográficas (BDb), a saber:

1. *Dos bases multidisciplinarias con distintos alcances geográficos*

Web of Science (WoS), norteamericana, creada en 1964 y actualmente propiedad de *Thomson Reuters*, a través de su plataforma en Internet, indexa 12,000 revistas de corriente principal; dentro de ellas la mayoría son anglófonas y más de 150,000 actas de congresos, que cubren información actual y retrospectiva en ciencias, ciencias sociales y artes y humanidades.¹ La información sobre la *mainstream science* o corriente principal permitirá situar, de manera comparativa, a la química en México en relación con criterios internacionales de excelencia. Otra base de datos bibliográfica que cumple con estas mismas características es *Scopus*, no se trabajó con ésta porque muchas de las revistas consideradas de corriente principal coinciden con las indizadas en WoS, por lo que se decidió por la de mayor tradición.

PERIÓDICA, base de datos regional desarrollada desde 1978 por la Universidad Nacional Autónoma de México (UNAM) a cargo de la Dirección General de Bibliotecas (DGB) (Alonso Gamboa y Reyna Espinosa, 2005), ofrece alrededor de 336,000 registros bibliográficos de artículos originales, informes técnicos, estudios de caso, estadísticas y otros documentos publicados en cerca de 1,500 revistas de América Latina y el Caribe, especializada en ciencias exactas y tecnología.² Esta base computa sólo las publicaciones científicas americano-latinas

1 Información extraída de: http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science/

2 Información extraída de: http://132.248.9.1:8991/F/3FR4FX7BT2FN7NNVCRAQIA746UMM9TFXL8QQJU57TDBF4DFERG-01337?func=file&file_name=base-info-per01

con el fin de dar a conocer la producción científica en las lenguas de la región (castellano y portugués).

2. *Una base disciplinaria*

Chemical Abstracts (CAS), base de datos internacional desarrollada desde 1947 por *Chemical Abstracts Service* de los Estados Unidos de Norteamérica, una división de la *American Chemical Society*. CAS se podría considerar como el líder mundial sobre la información en química porque provee la más completa base de datos en investigación química y ciencias relacionadas: monitorea, indexa y resume la literatura y patentes de la química a nivel mundial, incluye información sobre diversas disciplinas, como ciencias biomédicas, química, ingeniería, ciencias de los materiales, ciencias agrícolas, entre otras. Cuenta con ligas a referencias desde principios del siglo XIX, y a más de 280 editores y 6,700 revistas.³

3. *CAB Abstracts*

Desarrollada por *CAB International* del Reino Unido, proporciona desde 1972 acceso a más de cinco millones de registros seleccionados de más de 9,000 revistas, libros y memorias que cubren las áreas de las ciencias agrícolas y de la vida como: agricultura, ciencias veterinarias y animal, ciencias ambientales, salud humana, alimentación y nutrición, entre muchos otros. Esta base permitirá situar, en parte, la relación de la química en México con su entorno económico y social.⁴

Métodos

Se desarrolló una metodología ex profeso para este estudio que nos permitió construir cuatro bases de datos bibliométricas (BDB), con la información recopilada de las cuatro diferentes BDB correspondientes (WoS, CAS, *PERIÓDICA*, *CAB Abstracts*). El método está desglosado en siete pasos: 1. Reconocimiento de la base de datos bibliográfica. 2. Desarrollo y aplicación de las estrategias de búsqueda. 3. Tratamiento de los registros bibliográficos. 4. Diseño de la base de datos bibliométrica. 5. Migración de información. 6. Normalización y obtención de resultados. 7. Visualización interpretación y análisis de los resultados (*Figura 1*).

1. *Reconocimiento de la base de datos bibliográfica*

Con este primer paso fue posible determinar la estructura de las

3 Información extraída de: <http://www.cas.org/>

4 Información extraída de: <http://www.cabi.org/>

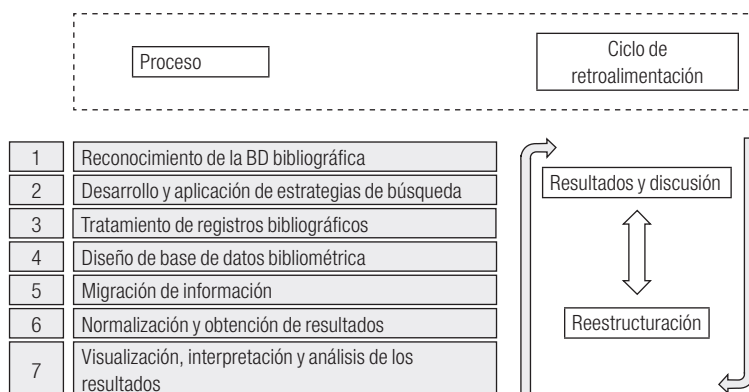


Figura 1. Esquema general de la metodología.

BDB, ¿qué campos bibliográficos se manejan en cada caso? y ¿cuáles son las opciones de búsqueda y recuperación que ofrecen?, y de esta forma tener una visión global de lo que se podía obtener en cada caso. La Bdb WoS⁵ basa su consulta en íconos y operadores booleanos, el reto en este caso se presentó en las estrategias de búsqueda, dado que a diferencia de la tendencia generalizada de usar categorías y títulos de revista para la selección de los registros que se iban a recuperar, decidimos centrarnos en la dirección de adscripción como se verá más adelante; CAS,⁶ por otro lado, tiene una estructura basada en pantallas de selección, descarte y filtros, lo que la hace una Bdb dirigida más a quien tiene bien definida su necesidad de información, y no a quien está explorando términos generales sobre un tema, pero a pesar de esta limitante nos fue posible desarrollar estrategias de búsqueda basadas en la dirección de adscripción del autor, con estructura diferente a la de WoS pero bajo la misma lógica. La base de datos *PERIÓDICA*⁷ tuvo la complicación de no poder ser explorada desde su sitio web, ya que no está diseñada para la recuperación ni de grandes bloques de información, ni para que una vez que ésta fue recuperada, esté completa en términos de los campos del registro bibliográfico, por lo que se pidió a la Dirección General de Bibliotecas de la UNAM, productor de la base, un bloque de información del que se limpió y recuperó la producción mexicana en química. Aunque esto fue de gran ayuda, el formato de los registros requirió un arduo trabajo de selección, limpieza y migración. La última Bdb con la que se trabajó fue *CAB Abstracts*,

5 Su acceso fue a través de Dirección General de Bibliotecas, UNAM

6 Su acceso fue a través de SciFinder

7 Su acceso fue a través de la Dirección General de Bibliotecas, UNAM

de estructura mucho más compleja, que se logró simplificar mediante el uso de un campo bibliográfico específico del instrumento; ahí las opciones fueron: operadores booleanos, índices de selección y filtro, así como lenguajes controlados, lo que nos permitió tanto recuperar información muy específica, como hacer exploración en general.

En la *Tabla I* se enumeran los campos bibliográficos que se extrajeron de cada BDB. Las dos intenciones por las que se muestra este cuadro son que identifica los campos disponibles por BDB, y que además permite conocer esos otros campos que pueden ser comparables (medibles), y los que pertenecen a una sola visión de la química. Como cada BDB estructura las relaciones entre los componentes de sus registros bibliográficos de forma particular, una parte importante de estas relaciones es la que se establece con los lenguajes libres o controlados que describen los contenidos disciplinares o temáticos. De las cuatro BDB usadas en este trabajo, WoS hace la descripción de los contenidos de las revistas indizadas con los *Journal Citation Reports* (JCR); CAS usa un lenguaje controlado desarrollado internamente, denominado *CAS sections* (secciones CAS), la sección se asigna al resumen del trabajo de acuerdo con la novedad del proceso o sustancia que está siendo reportado en la literatura;⁸ *PERIÓDICA* le asigna temas a cada uno de los trabajos, y en complemento a esta característica se emplearon los temas que usa el *Sistema Regional de Información en Línea para Revistas Científicas de América Latina, LATINDEX*.⁹ *CAB Abstracts* desarrolla y mantiene un listado de términos relacionados denominado *CABICODES*.¹⁰ Por último para una mejor selección de los registros que representan a la química en México es importante hacer mención de esta forma de caracterizar temáticamente el contenido de los registros indexados por las fuentes consultadas, dado que fue uno de los dos campos de los que no fue posible hacer un análisis transversal, el otro campo es la dirección de adscripción, como se explica más adelante.

8 Información extraída de: <http://www.cas.org/products/print/ca/casections.html>

9 *LATINDEX*. Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal: <http://www.latindex.org>

10 Los *CABICODES* son términos que permiten localizar categorías generales que el uso de descriptores no permitiría; otra ventaja de su uso es que ayuda a descartar información irrelevante. Los términos se derivan de los esquemas de clasificación usados en las bases de datos *AGRICOLA* y *AGRIS*, con cambios que permitan ampliar el espectro del término y su énfasis. Existen aproximadamente 250 *CABICODES*.

Tabla I. Tabla comparativa de campos de registros bibliográficos en las cuatro BD

WoS	CAS	PERIÓDICA	CAB <i>Abstracts</i>
Autor(es)	Autor(es)	Autor(es)	Autor(es)
Título	Título	Título	Título
Fuente	Fuente	Fuente	Fuente
Año	Año	Año	Año
Adscripción	Adscripción	Adscripción	Adscripción
Idioma	Idioma	Idioma	Idioma
Tipo de Documento	Tipo de Documento	Tipo de Documento	Tipo de Documento
Categorías	Secciones CAS	Tema <i>LATINDEX</i>	CABICODES
Palabras Clave	—	Tema	—
Resumen	Resumen	Resumen	

2. Desarrollo y aplicación de las estrategias de búsqueda

Cada BDb se trabajó por separado aplicando la regla de hacer por lo menos dos pruebas previas a la recuperación definitiva, asegurando contar con referencias numéricas y teniendo experiencia en el manejo de la herramienta. Las estrategias de búsqueda se pensaron y diseñaron para intentar tratar en un terreno equitativo las cuatro fuentes bibliográficas sin perder de vista sus particularidades y diferencias. Es por ello que se decidió usar la dirección de adscripción como el eje de las estrategias, y no los temas o las revistas, como el caso de otros trabajos.

- *Web of Science*

En 2006, Sampson y McGowan publican un trabajo donde se identifican 11 errores que con mayor frecuencia afectan el resultado de las estrategias de búsqueda; tuvimos cuidado de evitar los casos en los que se pudiera presentar alguno de estos errores. Por ejemplo, al desarrollar las estrategias en WoS se hizo una revisión de las variantes de los nombres de las instituciones mexicanas que cultivan la investigación química, de esta forma se evitó la pérdida de información a causa de las variantes ortográficas (Sampson y McGowan, 2006).

De los dos tipos de búsqueda, general o avanzada, a través de los que se ingresa a la información bibliográfica del WoS, se optó por la forma avanzada; ésta ofrece 18 campos del registro bibliográfico que combinados tienen el potencial de arrojar datos muy específicos o muy generales; en este caso en particular las estrategias tratan de mostrar la variedad de lugares donde se desarrolla la investigación química en México, y no la información asociada a instrumentos de control bibliográfico, como cuando se usan listados de revistas o categorías.

Las estrategias que con frecuencia se usan en estudios métricos recaen en dos campos: el de dirección y el de tópico. En el caso de la dirección (en algunos casos el país) con una delimitación temporal, por ejemplo: “ad=mexico not new mexico and py=2000-2004” o “cu=mexico and py=2000-2004”, la desventaja que identificamos al explorar esta opción es que se obtiene tanta información que se debe forzosamente usar algún instrumento confiable que permita seleccionar del amplio bloque de registros que se descargan, aquéllos que sean útiles en la delimitación de la disciplina.

El otro campo de amplio uso es el de tópico (ts=chemistry and py=2000-2004), nuestra oposición a su uso, como eje de la estrategia, residía en que se limita a buscar en el título del trabajo, palabras clave y el resumen (Thomson Reuters, 2006), y no coincide con el objetivo de nuestro trabajo. Pero no se descartó totalmente su uso, por lo que se complementaron las búsquedas basadas en la afiliación con el campo de tema (Ver *Figura 2*).

El camino que se trazó para afinar las estrategias con las que se trabajó se dividió en tres partes: pruebas con campos de búsqueda, refinamiento de búsquedas y pruebas de control.

En la etapa de las pruebas con campos de búsqueda se corrieron estrategias usando los campos “OG= Organización” y “SG= Suborganización”, que representaban segmentos dentro del campo de adscripción del autor(es), y se combinaron con otros campos como el de “tópico” (ts). Lo que se le pedía a la base de datos eran registros que junto con la delimitación temporal de nuestro trabajo, en el “ts”, tuvieran la raíz de la palabra “chemistry” o alguna variante, y además que en alguna parte de la organización o suborganización también se cumpliera con esa condición. De esta forma se complementa la condicionante de la dirección con las ocurrencias en el título.

La precisión que se alcanza usando los segmentos de la dirección de adscripción asegura la recuperación de las instituciones, lo que aunado al campo de “Country” identifica además la localización geográfica.

Se reconocen los sesgos que implica el uso de estas estrategias sobre otras formas más tradicionales, pero la “fotografía” que tratamos de obtener directamente de la BDb aporta información que de otra forma sería más complicado observar.

Una vez que se corrieron varias pruebas se empezó un proceso de refinamiento de búsquedas, y preferimos usar operadores de cercanía en lugar de los de combinación. Se encontraron casos en los que sabíamos que los trabajos que se publicaron eran de química, pero en el nombre de la adscripción no aparecía esto de forma explícita, para

cada caso en particular se desarrolló la estrategia que mejor cubriera las posibilidades de las variantes del nombre de la institución.

En la *Figura 2* se observan dos bloques de estrategias, las generales que recuperaron todos aquellos registros que en alguna parte del campo de dirección indicaran un interés por el cultivo de alguna disciplina de la química, a través del nombre de sus departamentos, laboratorios, o algún otro. Pero también estaba el bloque de las búsquedas específicas, donde se encontraban todas aquellas afiliaciones que no establecían de forma evidente su conexión con la química, pero que en la práctica sí la tienen. En todos los casos, y dado que se usó la versión más reciente del WoS se tuvo la función de “Lematización” encendida, que aseguraba la cercanía de palabras en la búsqueda. Esto no se limitó a ninguna base de datos en particular, es decir, se incluyó información del *Science Citation Index*, del *Social Sciences Citation Index*, y del *Arts and Humanities Citation Index*.

	Estrategia final	Total de registros recuperados
Búsquedas generales	1. ad=(mexico same(chem* or quim* or pharm* or farm*)) and py=2000-2004 and cu=mexico	1. 3634
	2. TS=(chem* or pharm*) and py=2000-2004 and cu=mexico	2. 2822
Búsquedas específicas	1. ad=(mexico same (ciad) or (ctr same invest* same aliment* same des*) or (Ctr same food same dev* same (invest* or res))) and (cu=mexico) and (py=2000-2004)	1. 272
	2. ad=(mexico same (imp) or (inst* mex* petr*) or (mex* petr* inst*)) and (py=2000-2004) and (cu=mexico)	2. 938
	3. ad=(mexico same (ciqa or ciq or ESQIE or esiquie or ESIQUE or EISQIE)) and py=2000-2004 and cu=mexico	3. 228
	4. ad=(mexico same (cideteq) or (ctr same invest* same desarr* same tecno* same electroqu*) or (Ctr same electroch* same dev* same (invest* or res))) and (cu=mexico) and (py=2000-2004)	4. 25

Figura 2. Estrategias de búsqueda para *Web of Science*

Dado que las estrategias que se usaron para WoS se centran en la lógica del documento, es decir, que se tomaron en cuenta elementos y características que se encuentran en el registro bibliográfico y a partir de ahí combinarlos con las formas de búsqueda de la BDb, se requirió una vía para corroborar si lo recuperado era representativo de la producción de

química en México; esto condujo a correr algunas pruebas de control. Estas pruebas de control consistieron en identificar los trabajos publicados en el mismo periodo de nuestro estudio de 20 químicos con dirección de adscripción “México” y pertenecientes al Sistema Nacional de Investigadores, para hacer la comprobación de haber recolectado los registros que contuvieran a éstos.

- *Chemical Abstracts*

Se introdujo la palabra “mexico” en la forma de búsqueda por exploración que indiza los “company names” (similar a la dirección de adscripción en WoS); posteriormente se filtraron las delimitaciones temporales (cada año por separado), por tipo de documento (sólo lo que CAS identifica como “journal”), por base de datos (CAS indiza registros bibliográficos de MEDLINE así que éstos se descartaron).

La Figura 3¹¹ desglosa el recorrido que se transita para aplicar la estrategia de búsqueda en CAS, desde la enorme cantidad que se obtiene al hacer la exploración por “company name”, hasta los registros que se recuperan pero que no corresponden al total con el que se trabaja, dado que aún se hace un proceso de limpieza y normalización posterior en campos como el de dirección, situación que es común a todos los instrumentos que se construyeron.¹² Cabe mencionar que la selección de “Company/Organization” permite descartar aquellas direcciones que corresponden a “New Mexico, EU”.

Al final la recuperación de los registros bibliográficos se da en archivos de texto plano individuales por cada sección CAS.

Filtrado por:	2000	2001	2002	2003	2004
Exploración por «company name»: mexico	117967	118135	118175	118187	118294
Delimitación temporal	4323	4706	5029	5365	5905
Filtrado por base de datos	2282	2401	2435	2576	2842
Filtrado por Tipo de documento	1951	1616	2105	2208	2388
Quitar duplicados	3693	2400	1588	—	—
Selección de «Company/Organization»	722 (candidatos)*	787	681	598	474
Total registros recuperados	1442	1616	1588	1703	1781

Figura 3. Fases de filtrado y selección en la estrategia de búsqueda en CAS.

11 Los “candidatos” se refieren a nombres de instituciones que pueden aparecer en más de un registro, por ejemplo, bajo el candidato “Universidad Nacional Autónoma de México, Mex” pueden recuperarse 150 registros bibliográficos. Al final se recuperan archivos por cada sección CAS. En 2003 y 2004, el historial de búsqueda no reportó la cantidad de registros que se removieron.

12 El orden en los filtros no afecta el resultado final, por lo que en algunos casos la reducción de los registros no es secuencial.

- *PERIÓDICA*

Se trabajó a partir de un archivo en hoja de cálculo, por lo que la estrategia para recuperar los trabajos con adscripción “México” fue usando las herramientas propias de este manejador de datos, en nuestro caso filtros y migración de bloques de información en diferentes hojas de trabajo, lo que dependió de dos criterios de selección: uno basado en el título del trabajo, usando palabras comodín como forma de descarte y selección; y el segundo basado en el título de la revista. Para el criterio dos se usaron de nuevo las palabras comodín en combinación con los temas que describen el contenido de las revistas en el *LATINDEX*, el uso de esta herramienta para delimitar disciplina ha formado parte de otras investigaciones, como en el caso de las ciencias agrícolas (Saavedra-Fernández *et al.*, 2002).

Los criterios que sirvieron de estrategias de búsqueda en *PERIÓDICA*, en la *Figura 4*, tienen una columna para el total de registros identificados, pero no representan el resultado final, pues todavía se eliminaron duplicados y se normalizó, reduciendo la cantidad de trabajos en la BDB.

Criterio 1	Palabras Comodín	Palabras comodín para LATINDEX	Total de Registros
Título de trabajo	<ul style="list-style-type: none"> • "quim*" • "chem*" • "farm*" • "pharm*" • "tox*" 		2477
Criterio 2	Palabras Comodín	Palabras comodín para LATINDEX	Total de Registros
Revista	<ul style="list-style-type: none"> • "quim*" • "chem*" • "farm*" • "pharm*" • "tox*" 	<ul style="list-style-type: none"> • "*quim*" • "*celu*" • "*farma*" • "*tox*" • "*petrol*" • "*genet*" • "*ferm*" • "*conta*" • "multi*" 	1498

Figura 4. Estrategias de selección y filtrado en la estrategia de búsqueda en *PERIÓDICA*.

- *CAB Abstracts*

La estrategia de búsqueda fue de lo general a lo particular. Se inició pidiendo la identificación en el campo “All Searchable Fields” (AF), es decir en todos los 52 campos del registro bibliográfico, aquellos trabajos

que pudieran contener “mexico”, para después empezar a reducir los registros que se recuperaron a través del periodo del estudio, el tipo de documentos de interés y por último los CABICODES, que es un lenguaje controlado interno, seleccionan después aquello que categorizan en química. La recuperación fue en archivos con no más de 200 registros, que es lo que permite el instrumento.

En la *Figura 5* se reportan 718 trabajos recuperados de la BDb CAB, a los cuales aún se les aplicó un proceso que redujo la cantidad de registros en la BDB final.

Estrategia general	Total de Registros
«mexico» (af)	2889
Filtrado:	
2000 - 2004	
Sólo: Journal, Journal article, Journal Issue	
Not «New Mexico»	1218
CABICODES	718

Figura 5. Estrategia de búsqueda aplicada en *CAB Abstracts*.

3. *Tratamiento de los registros bibliográficos*

Dado que la estructura de la referencia bibliográfica en cada BD así como la forma de recuperación es diferente en cada caso, se usó la paquetería de Office para procesar y reestructurar la información de forma que posteriormente se pudiera migrar a Bases de Datos Bibliométricas. La transformación que sufren los registros asegura que se puedan planear el diseño de las BDB, la cantidad de tablas y su contenido; reconocer los datos que deben ser normados y de los que se puede derivar más información, y que se pueda inferir el tipo de indicadores que es posible obtener. WoS es manejable en el sentido de poder recuperar un formato en columnas por campo del registro bibliográfico, así que el mayor trabajo se encuentra en separar los autores, las direcciones y las categorías. CAS y *CAB Abstracts* se recuperan en texto plano, por lo que el proceso es más largo y minucioso en el que es necesario primero integrar todos los registros en un solo archivo, para después organizarlo en columnas por campo de registro bibliográfico, desagregar los elementos que lo requieran, y por último migrar a la BDb. El caso de *PERIÓDICA*, aunque de igual forma tomó tiempo, al final se completó de forma más rápida, ya que desde el momento

que se tuvo el concentrado general, se empezaron a hacer ajustes de arreglo y desagregación.

4. *Diseño de la base de datos bibliométrica (BDB) y 5. Migración de información*

En estos pasos se estructura completamente el instrumento, las BDB tienen al final una estructura:

- a) WoS, 9 tablas con la información que se extrajo originalmente; y dos derivadas, una por país y otra por entidad geográfica y que contiene los estados de la República mexicana.
- b) CAS, 9 tablas con información original, y una derivada por entidad geográfica.
- c) *PERIÓDICA*, 9 tablas con información original, y dos derivadas (país, entidades geográficas).
- d) *CAB Abstracts*, 9 tablas con información original, y una derivada por entidad federativa.

Dado que nuestro objetivo no era construir un instrumento único que incluyera toda la información recuperada, sino observar qué visión de la investigación química ofrece cada BDB, se construyeron cuatro bases de datos bibliométricas con características comunes hasta donde fue posible, pero sobre todo aprovechando las diferencias y así obtener un panorama amplio con las correspondientes coincidencias, como se muestra en los resultados.

En todo momento de la metodología se usó la paquetería de Office.

6. *Normalización y obtención de resultados*

La etapa de normalización fue uno de los momentos más importantes, pues el campo que en todos los casos se estandarizó fue el de dirección, así que después de terminar la primera BDB y haber asignado siglas y formas de las direcciones de adscripción, se retomó el proceso para hacer comparables y compatibles las formas en que aparecen las instituciones que generan investigación química en México en las otras BDB. Dada la particularidad de falta de normalización del campo de dirección y la vasta información que incluye: nombre de la institución (junto con departamentos, escuelas y facultades, laboratorios, entre otros), nombre del autor, seguido de la dirección postal (número, estado y país), y para aprovechar todo este contenido, se separaron todos los elementos; así pudimos potencializar un campo plano, y proyectar conteos y el análisis por la institución, la institución/colaboración (ahora ya normalizada), los geográficos y el geográfico/colaborativo (también se normaliza) por estado o país. Esta estandarización de las direcciones de adscripción así como la identificación

de sus limitaciones y alcances en cada caso, es información que será usada en otros trabajos dentro del marco de esta investigación. Cada base de datos le asigna diferente relevancia al campo de dirección, especialmente al momento de recuperar el registro bibliográfico completo; únicamente en el caso de WoS es posible recuperar todas las direcciones de los autores que firman el trabajo, el resto de las BDb consultadas sólo proporcionan la dirección del primer autor, o la que se deriva de la estrategia de búsqueda.

Hubo otros casos, por ejemplo, en que la normalización se aplicó al agrupar en formas comunes los casos en que se desglosa mucho la información, caso específico son los tipos de documento; por ejemplo en CAS el tipo de documento “journal” podía ser: Journal article, Journal article/Conference/paper, Journal/issue/Conference proceedings, y al revisar una muestra nos dimos cuenta de que en su mayoría eran sólo del tipo “Artículos”, así que se renombra para efectos de su medición. Otro campo que fue necesario nombrar en los registros extraídos de CAS fue el del título de la revista, dado que hay una porción de la información que se recupera en forma de títulos abreviados.

La *Tabla II* muestra los campos finalmente seleccionados y normalizados que componen cada una de las BDB derivadas de cada una de las bases de datos bibliográficas ocupadas para el estudio de la química en México. El encabezado “Campos Construidos” enumera la información específica derivada de la información disponible de la adscripción del autor; la reestructuración de este campo consistió en separar la parte institucional (dirección), la parte geográfica (estado), y la parte de la colaboración científica internacional (país).

Tabla II. Estructura de las bases de datos bibliométricas (BDb)

Campos del registro	WoS	CAS	PERIÓDICA	CAB Abstracts
Clave	•	•	•	•
Autor	•	•	•	•
Título	•	•	•	•
Fuente	•	•	•	•
Año	•	•	•	•
Categoría, Disciplina o Tema	•	•	•	•
Dirección de Adscripción	•	•	•	•
Idioma del artículo	•	•	•	•
Campos Construidos				
Dirección / País / Estado	•		•	
Dirección/ Estado		•		•

7. Visualización, interpretación y análisis de los resultados

Este último paso tiene como objetivo ubicar y generar el tipo de conteos, así como los resultados y su visualización. Hay que tomar en cuenta que lo que hay en las BDb no es suficiente para entender la caracterización que se puede hacer de la investigación química de México, así que por un lado fue posible organizar en tablas, figuras y una matriz gran parte de la información resultante, y por el otro fue necesario integrar y recuperar datos que completaran el panorama, como el país de edición de las revistas y otros datos relevantes.

RESULTADOS

Evolución general de la producción en química de cada BDB (2000-2004)

La producción mexicana en química a partir de las cuatro BDb que se revisaron marca dos comportamientos que representan el tipo de información que cada una indiza, y reflejan en su cobertura la representatividad de las instituciones mexicanas que cultivan la química. Por un lado están la BDb especializada CAS y la multidisciplinaria WoS, siendo la primera la que domina la producción general; y por otro lado están la BD regional *PERIÓDICA* y la que recoge la visión de la investigación aplicada CAB, en ambos casos su proceso de indización está muy alejado de los patrones de las bases más visibles, como se observa en la *Figura 6*.

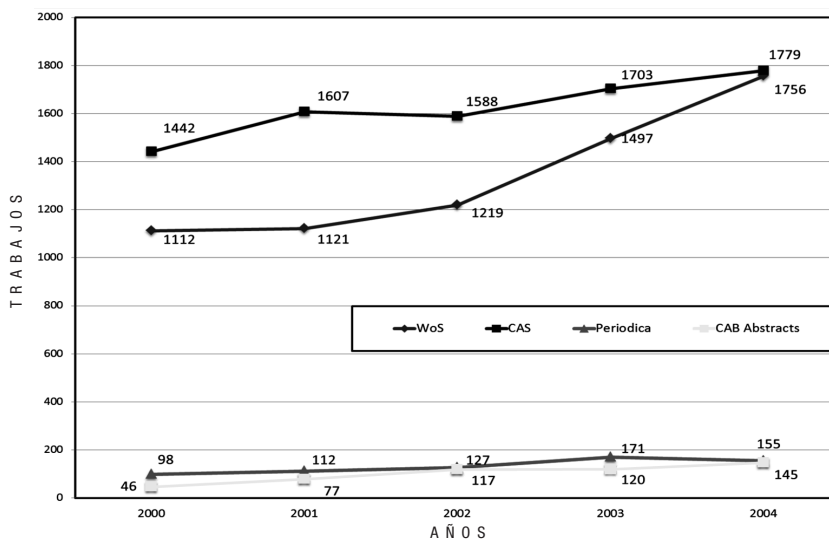


Figura 6. Producción anualizada de las cuatro BD para la química en México 2000-2004.

Teniendo nuestro estudio las estrategias de búsqueda basadas en la dirección de adscripción, hay que acotar que la dinámica de la producción reflejada en la *Figura 6* está sujeta a los procesos de normalización donde se identificaron 241 instituciones, localizadas en las 32 entidades federativas de México en el caso de WoS, y 124 en CAS ubicadas en sólo 27 estados. La relación del número de instituciones en *PERIÓDICA* es de 83 dependencias distribuidas en 27 estados, mientras que en CAB se normalizaron 54 instituciones en 30 estados. Toda esta información sirvió de base para determinar el curso de la investigación en futuros trabajos basados en esta dinámica.

A partir de estos primeros resultados procedemos a conocer la estructura de las BDB para darles significado a estas primeras cifras.

Nuestra metodología permitió arribar a diferentes tipos de resultados. Con respecto a los de caracterización general, la *Tabla III* muestra la conformación de las diferentes bases de datos con respecto al idioma. Como era de esperarse, el inglés y el español son los idiomas que más aparecen en los trabajos mexicanos en química y hay poca representación de otras lenguas incluyendo el portugués. Más del 90% de los documentos recuperados de las BD internacionales de WoS y CAS fueron publicados en inglés, *PERIÓDICA* en cambio recuperó el 81% de documentos en español, lo que indica la importancia de esta fuente para no dejar de lado los trabajos de la química en México publicados en revistas de alcance local o regional. En el caso de *CAB Abstracts* el porcentaje fue menor ya que el 28% correspondía a registros escritos en español o en español e inglés.

Tabla III. Idioma de los trabajos mexicanos de la química 2000-2004

Idioma	WoS	CAS	<i>PERIÓDICA</i>	<i>CAB Abstracts</i>
Alemán	1	0	1	0
Español	253	397	539	104
Francés	7	3	0	1
Hebreo	0	1	0	0
Inglés	6437	7688	123	362
Inglés/Español	2	18	0	35
Japonés	0	1	0	0
Portugués	3	0	0	3
Ruso	2	3	0	0
Total	6705	8111	663	505

Al nivel del medio utilizado para la comunicación de resultados, la *Tabla IV* muestra que a través del WoS se recuperaron primordialmente dos tipos de documentos, artículos de revista y de memorias (*journal article* y *conference*

paper) más un número importante de revisiones (*review* o revisión), CAS recoge además de artículos de revista otros tipos de fuentes como son: revisiones y cartas al editor (*letter*). En cuanto a *PERIÓDICA* fueron tres los tipos de documentos que se tomaron en cuenta (artículo, revisión bibliográfica y nota breve o noticia) y en *CAB Abstracts*, artículos de revista casi en su totalidad. A pesar de esta diversidad, dentro de las cuatro BDB el artículo es el medio principal que varía desde el 99% con respecto al CAB hasta un 85% en el caso de *PERIÓDICA*.

Como resultado de la normalización de algunos de los campos del registro bibliográfico, la *Tabla IV* enlista los tipos de documento en que se agrupan las diversas formas en que clasifican la información las BDB, anotándolos en el idioma en el que se recuperó la información y resaltando de esta forma las particularidades de cada instrumento consultado.

Tabla IV. Tipos de documento de los trabajos mexicanos de la química 2000-2004

Tipo de documento	WoS	CAS	PERIÓDICA	CAB Abstracts
<i>Journal Article</i> (Artículo)	5510	7612	570	503
<i>Conference Paper</i> (Memorias)	924	11	0	2
<i>Letter</i> (Cartas)	38	13	0	0
<i>Review</i> (Revisión)	190	475	72	0
(Nota breve o noticia)	0	0	21	0
Editorial Material	33	0	0	0
<i>Correction</i> (Correcciones)	10	0	0	0
Total	6705	8111	663	505

Otro indicador importante son las características de las revistas más representativas en cada BDB. En la *Tabla VI* se aprecia que el mayor número de trabajos por mucho (187) fueron publicados en la *Revista de la Sociedad Química de México* y recuperados de la BDB de *PERIÓDICA*, lo que indica la importancia de esta fuente local para difundir los resultados de la investigación en esta área. La misma revista ostentaba 125 trabajos en CAS ubicada en tercera posición, lo que podría sugerir una indización selectiva de este servicio a favor de los trabajos más enfocados hacia la química pura; en *PERIÓDICA* se incluyen trabajos sobre temáticas más variadas incluyendo títulos de revista especializados en educación. *CAB Abstracts* también evidenciaba dos revistas mexicanas entre las primeras tres, pero esta vez especializadas en ciencias agrícolas. Nos llama la atención que en el caso de *PERIÓDICA* se trate de revistas disciplinarias y en *CAB Abstracts* más bien de temáticas. Las BDB WoS y CAS incluyen dos títulos de revistas editadas en México: *Revista Mexicana de Física* y *Archives of Medical Research*, ambas de disciplinas relacionadas con la química y donde ésta

encabeza el ranking; en el caso de *PERIÓDICA*, también se recuperan estos títulos pero no con la cantidad de trabajos para figurar en la *Tabla V*. CAB incluye el *Archives of Medical Research*, y está en la misma situación que en *PERIÓDICA*.

En las BDB de corte internacional las revistas mexicanas aparecen en los primeros lugares. La mayoría de las revistas mexicanas en la lista son editadas por instituciones nacionales de educación superior, lo que demuestra la importancia de éstas para el desarrollo de la ciencia en México, ya que desde su formación los químicos y otros científicos entienden la trascendencia de estos canales de comunicación para dar a conocer sus hallazgos.

La mayor parte de las revistas que se muestran en la *Tabla V* se editan en México o Estados Unidos, lo que indica una forma de publicación para los químicos que están en instituciones mexicanas. Tanto en WoS como en CAS se puede observar una tendencia a publicar en revistas relacionadas con temas como física y medicina, a diferencia de *PERIÓDICA*, es decir que tal vez aquellos temas que tienen interés transdisciplinario o de la aplicación de la química son los que se ponen a consideración de la comunidad internacional, mientras que la publicación local tiene una visión más disciplinar.

Tabla V. Revistas más representadas en las cinco BD para la química en México 2000-2004

BDB	Química	País de edición	Totales
WoS	Abstracts of Papers of the American Chemical Society	EUA	136
	Journal of Chemical Physics	EUA	76
	Journal of Organometallic Chemistry	EUA	63
	Revista Mexicana de Física	MEX	59
	Journal of Molecular Structure-Theochem	Países Bajos	54
	Archives of Medical Research	MEX	53
CAS	Revista Mexicana de Física	MEX	159
	Archives of Medical Research	MEX	139
	Revista de la Sociedad Química de México	MEX	125
	American Institute of Physics Conference Proceedings	EUA	112
	Physical Review D: Particles, Fields, Gravitation and Cosmology	EUA	90
	Proceedings of the Western Pharmacology Society	EUA	86
<i>PERIÓDICA</i>	Revista de la Sociedad Química de México	MEX	187
	Revista del Instituto Mexicano de Ingenieros Químicos A.C.	MEX	55
	Educación Química	MEX	50
	Revista Mexicana de Ingeniería Química	MEX	46
	Revista de Educación Bioquímica	MEX	35
	Boletín de Educación Bioquímica	MEX	31

CAB Abstracts	Terra	MEX	50
	Journal of Agricultural and Food Chemistry	EUA	50
	Agrociencia	MEX	28
	Journal of Food Science	EUA	17
	Revista Internacional de Contaminación Ambiental	MEX	12
	Journal of Biological Chemistry	EUA	11

Para observar la forma en que las BDb se solapan se tomaron las diez revistas que mayor número de trabajos concentran y se ubicó la o las fuentes donde coincidían, incluyendo títulos duplicados en tanto que nuestro objetivo era identificar las BDb en los que coincide la selección en la indización, y no la frecuencia con la que esta situación se presenta. En el caso de la revista preferida para publicar, la *Revista de la Sociedad Química de México* es una de las que difiere del resto de las coincidencias, pues es indizada por la fuente especializada CAS y por la visión local que ofrece PERIÓDICA. En el resto de los títulos las BDb predominantes son WoS y CAS que en la dinámica de nuestro estudio son las que caracterizan las formas de publicación de los investigadores con adscripción a instituciones mexicanas. La otra excepción al comportamiento predominante es el *Journal of Agricultural and Food Chemistry* que se indiza tanto en CAB como en WoS y CAS.

De igual forma se refuerzan las características de transdisciplinariedad con la física y la medicina, y surge el predominio de la visión especializada y de corriente principal.

Tabla VI. Revistas de mayor producción mexicana en química 2000-2004

Título de Revista	BDB donde se encuentra	Trabajos por BDb	Total acumulado
Revista de la Sociedad Química de México	CAS	125	312
	PERIÓDICA	187	
Revista Mexicana de Física	WoS	59	218
	CAS	159	
Archives of Medical Research	WoS	53	192
	CAS	139	
Journal of Chemical Physics	WoS	76	137
	CAS	61	
Journal of Agricultural and Food Chemistry	WoS	51	101
	CAB	50	

	CAS	16	117
Journal of Organometallic Chemistry	WoS	63	
	CAS	52	115
Revista de Investigación Clínica	WoS	45	
	CAS	55	100
Journal of Molecular Structure: THEOCHEM	WoS	40	
	CAS	54	94
Physical Review D: Particles, Fields, Gravitation, and Cosmology	WoS	2	
	CAS	90	92
Thin Solid Films	WoS	49	
	CAS	41	90

La *Tabla VII* encuadra la asociación de los descriptores temáticos de cada BDb consultada con la frecuencia de aparición en las BDB que se construyeron, sólo se tomaron los primeros diez resultados que de forma descendente muestran la diversidad de temas que caracterizan los trabajos en química según lo reportado por instituciones mexicanas. Aunado a la descripción temática correspondiente a cada BDb consultada está la frecuencia con la que aparecen en las BDB construidas, esta frecuencia no representa números absolutos, puesto que en algunos casos como en WoS puede asignarse más de un tema asociado a la revista de publicación, mientras que CAS utiliza un sistema jerárquico donde a partir de cada sección CAS hay subsecciones que precisan la descripción del contenido del registro bibliográfico; la frecuencia entonces equilibra las diferencias de los instrumentos.

En el caso de las JCR (*Journal Citation Reports*) se sabe que a cada revista le puede corresponder más de una categoría, en nuestro caso se contabilizaron todas las categorías pertenecientes a cada revista, y se obtuvo un total de 112, y de éste las 10 que con más frecuencia identifican el periodo de 2000-2004 a los trabajos de corriente principal de instituciones mexicanas, lo que da cuenta de la variedad de disciplinas en las que se usan y publican las investigaciones químicas, no obstante el predominio de los títulos clasificados dentro de la categoría de química predominante.

En el caso de CAS se obtuvieron 79 secciones asociadas al contenido de los registros bibliográficos recuperados, de los cuales los primeros 10 que se muestran en la *Tabla VII* tienden a concentrarse de forma menos dispersa que las JCR, por ejemplo en el caso de las dos categorías JCR Química y Física, la

Tabla VII. Temas por Base de Datos para la química en México 2000-2004

	JCR	Frecuencia	CAS Sections	Frecuencia	PERIÓDICA	Frecuencia	CABICODES	Frecuencia
1	Química	2025	Farmacología	526	Bioquímica	164	Química de Suelos y Mineralogía	221
2	Física	850	Espectroscopia Óptica, Electrónica y de Masas, y otras propiedades relacionadas	435	Química orgánica	143	Química de Alimentos	109
3	Ciencias de Materiales	728	Hormonas en Mamíferos	408	Ingeniería química	129	Productos de Cultivo	104
4	Ingeniería	651	Fenómenos Nucleares	366	Química de alimentos	65	Composición y Calidad de los Alimentos	79
5	Farmacología y Farmacia	471	Bioquímica Patológica en Mamíferos	310	Química analítica	55	Contaminación y Degradación	66
6	Bioquímica y Biología Molecular	442	Inmunoquímica	275	Fitoquímica	35	Física de Suelos	66
7	Ciencias Ambientales y Ecología	337	Fenómenos Eléctricos	249	Genética	32	Cosechas de Campos	62
8	Ciencia y Tecnología de los Alimentos	300	Combustibles Fósiles, Derivados y Productos Relacionados	248	Química inorgánica	28	Biología de Suelos	56
9	Agricultura	216	Toxicología	227	Fisicoquímica y química teórica	27	Cultivos Hortícolas	51
10	Ciencia de las Plantas	206	Genética Bioquímica	219	Toxicología	25	Procesamiento de Alimentos (General)	49

distancia entre una y otra es de más de 1000 apariciones, en cambio en las dos secciones CAS que encabezan el listado (Farmacología y Espectroscopia óptica, electrónica y de masas y otras propiedades relacionadas) hay menos de 100 apariciones. Los temas en *PERIÓDICA* acumulan 123, los primeros 10 se relacionaron íntimamente con la química en sus diferentes manifestaciones. Los CABICODES claramente representan tanto los 10 de la tabla como los 150 de la base de datos bibliométrica, la parte aplicada de los trabajos que se indizan en *CAB Abstracts*.

Partiendo de que las estrategias de búsqueda están basadas en la dirección de adscripción, y aunado esto a las características propias de cada forma de identificar temáticamente los contenidos de los cuatro índices consultados, se puede observar en la *Tabla VII* que a excepción de los CABICODES, que son claramente más aplicados a las ciencias agrícolas, las JCR, las secciones CAS y los temas de *PERIÓDICA*, amén de la forma en que asigna la clasificación temática (al contenido o a las revistas), lo que muestran es la diversidad de temas que cultivan las instituciones mexicanas alrededor de la química, y a pesar de que se nombren en una forma general o más específica, hay constantes, como la toxicología y la relación con las ciencias físicas.

DISCUSIÓN Y CONCLUSIONES

Hemos confirmado que una adecuada selección de las BDb es un requisito indispensable al procesar investigación sobre los estudios cuantitativos de la ciencia. De suma importancia es también la vinculación de tales estudios con la problemática científica específica que se pretende esclarecer. Por ejemplo, CAS lleva muchos años siendo el principal servicio de información para la investigación en química (Levine-Clark y Kraus, 2007) pero más como fuente de referencias bibliográficas que para indagar sobre el comportamiento y los patrones de comunicación en las ciencias químicas. CAS presenta opciones de búsqueda de gran trascendencia para los químicos, como en los casos de la representación de estructuras químicas o por medio de fórmulas moleculares de interés limitado para las indagaciones bibliométricas. No obstante al limitar la búsqueda a una sola base, en este caso CAS, se deja fuera la opción de analizar las colaboraciones científicas o de recuperar toda la producción en revistas nacionales. En 2009 se publicó un trabajo en *Scientometrics*, que proponía usar las secciones CAS en lugar de los instrumentos tradicionales ofrecidos por Thomson Reuters, dadas las variaciones que ellos encontraron en el caso de la bioquímica; el trabajo discutía las limitantes de una sobre la otra en el caso de indicadores institucionales, mientras que argumentaba

estar a favor del trabajo que se hace cuando se asignan las secciones CAS a los trabajos indexados por las BDb: en nuestro caso las categorías JCR y las secciones CAS, más que alejarse de la visión que se da de la química que se cultiva en instituciones mexicanas, complementan este panorama. El uso exclusivo de las bases que cubren solamente las revistas de corriente principal, como hace WoS, da una visión parcial al centrarse en lo que la comunidad mexicana publica en inglés, lo cual resulta ser de interés y relevante para sus pares internacionales, pues por lo general se trata de investigaciones de frontera. No necesariamente así la amplia investigación que se publica en revistas nacionales, la cual suele ser de interés para una comunidad más pequeña y por ende, de relevancia más bien sólo a nivel local.

Otros trabajos donde se han usado bases de datos especializadas y se incluye la visión de corriente principal obtienen buenos resultados, como sucedió con en el análisis de la taxonomía biológica que hacen Michán y Llorente-Bousquets, donde el 8.88% del total de registros provinieron de *Science Citation Index*, mientras que el 52.04% y el 39.07%, correspondieron a las bases de datos especializadas BIOSIS y CAB, respectivamente (Michán y Llorente-Bousquets, 2010). En el caso del estudio de la actividad científica en ciencias de la salud en Latinoamérica, se identificaron 1,530 documentos en MEDLINE, frente a 2,392 en WoS y 1,381 en el Índice Médico Español (Costas *et al.*, 2008). Cuando Taylor *et al.* pusieron a prueba siete bases de datos con el objetivo de desarrollar estrategias de búsqueda para recuperar información en las ciencias sociales, los porcentajes se distribuyeron como sigue: MEDLINE 52%, *Social Sciences Citation Index* 46%, *Cumulative Index of Nursing and Allied Health Literature* 30%, *AgeInfo* 76%, *PsycInfo* 51% y *Social Services Abstracts* 41% (Taylor *et al.*, 2007). Escalona *et al.* comparan la cobertura en ingeniería química en dos bases de datos dedicadas a indizar trabajos de corriente principal, en WoS se encontraron 155,960 trabajos en un lapso de 10 años, mientras que en *Scopus* se recuperaron 165,285 en el mismo periodo (Escalona *et al.*, 2010).

El presente análisis ha mostrado también que el posicionamiento de las revistas en idioma español y el de las que están en inglés, en el caso de la producción de los científicos mexicanos en química, varía considerablemente de una BDb a otra. La estructura de cada BDb construida respecto al objetivo que se persigue influye sobre la visibilidad que se obtiene de la producción científica dentro de un campo, no solamente de manera cuantitativa sino también de manera cualitativa.

Así como en este estudio se trató de estandarizar en lo posible los registros para poder interpretar los resultados como un todo, también hay que valorar la realidad particular que refleja cada fuente. La actividad de quienes

trabajan la química en México está polarizada, sobre todo por la fuerte presión que existe para aparecer en los canales de corriente principal (WoS) y en los especializados en su área (CAS). Pero al mismo tiempo habrá temas, enfoques o tendencias que los conducirán a publicar en revistas que aparecen en canales más regionales (*PERIÓDICA*); por supuesto que los requerimientos de algunos temas permitirán su incursión en trabajos que tiendan a la aplicación de su trabajo y será CAB lo que refleje eso.

Entre 2000 y 2006, la producción científica en química representó el 12.17% de la producción científica total de México (CONACYT, 2007). La comunidad científica de químicos ostentaba para el año 2008, 1280 miembros del SNI (considerados los especialistas que producen la mayor parte de la investigación del país) repartidos en instituciones de educación superior y de investigación en un total de 38 entidades federativas del país. En ese mismo año la membresía total del SNI era de unos 14,681 investigadores, de los cuales los químicos representaban el 8.7% (CONACYT, 2008).¹³

México se posicionó en segundo lugar en producción científica total a nivel de América Latina (con 6,787 trabajos en 2005 y un promedio de 4,883 en el intervalo de 1996-2005), atrás de Brasil (con 15,777 en 2005 y un promedio de 10,343 durante 1996-2005) y antes que Argentina (5,182 trabajos en 2005 con un promedio de 4,147 en el periodo de 1996-2005) (CONACYT, 2006). En cuanto a las publicaciones en *Chemical Abstracts* para 2007, América Latina y el Caribe registraron un total de 21,949 de las cuales Brasil produjo 13,159 (el 60%), Argentina 2,973 (el 13.5%) y México 2,370 (el 10.8%) (RICYT, 2007). Argentina obtuvo un mayor número de publicaciones que México en esta fuente, una constante que se mantuvo entre 1990 y 2007, según las cifras de la RICYT.

Para concluir en términos generales, nuestro estudio apunta a las siguientes características de la investigación mexicana en química obtenidas a través de las bases de datos bibliográficas seleccionadas:

1. Los artículos en español y en inglés son el principal medio de comunicación de la comunidad de científicos mexicanos en el área de la química.
2. Las colaboraciones científicas que son un indicador importante en la estructuración de las disciplinas científicas por jugar un papel central en la caracterización de las comunidades científicas, son visibles en la BDb multidisciplinaria internacional WoS, aspecto que se pretende explorar en futuras publicaciones.

13 Cuento a partir del listado proporcionado por la Subdirección de Proceso de Evaluación- SNI. CONACYT, oct. 2008.

3. A pesar de que *PERIÓDICA* recupera títulos de revistas importantes para los químicos en México, esta fuente no es representativa de todo lo que éstos publican (ver conclusión 4). No obstante su carácter regional la convierte en fuente única e indispensable para cualquier indagación sobre la química mexicana. Dada la escasez de títulos especializados en español y el interés en alcanzar lectores y visibilidad a nivel internacional, los químicos mexicanos a menudo prefieren publicar en inglés en revistas editadas fuera de la región y que son recogidas exclusivamente por los grandes servicios internacionales multi y unidisciplinarios. Además, publicar en revistas de corriente principal es bien visto por los comités nacionales de evaluación.
4. Cuando lo que se pretende analizar es la estructuración de una disciplina en su contexto local, *PERIÓDICA* proporciona un conjunto de trabajos adecuados para este propósito. Además en *PERIÓDICA* la producción se concentra en 37 revistas (44% en las tres primeras), mientras que los trabajos en revistas internacionales se encuentran repartidos en un total de 1,559 publicaciones (WoS), 188 (CAB) y 1,640 (CAS). Así se puede hacer un estudio de la disciplina a partir de un pequeño número de revistas representativas.
5. Las revistas preferidas para publicar así como los temas asociados (ya sea a las mismas revistas o a los contenidos de los trabajos indizados) muestran una cualidad transdisciplinaria en la investigación química que deriva de instituciones mexicanas.
También con respecto a la metodología empleada en nuestro estudio podemos llegar a conclusiones generales. Tradicionalmente se usa la estrategia de extracción por bloque para a partir de esa gran cantidad de información hacer descartes y filtros, usualmente asociados a los descriptores temáticos; en nuestro caso esto no ofrecía resultados favorables, particularmente porque durante la etapa uno de nuestra metodología (ver *Figura 1*) pudimos reconocer trabajos donde los descriptores asociados con la investigación química recuperaban registros donde se tenía a la química como objeto de estudio (como en el caso del presente estudio), siendo ésta una intención que no formaba parte de nuestros objetivos.
Si bien el método que empleamos se aleja de las formas tradicionales de recuperación por bloque y el tratamiento “ex post”, nos representaron ventajas y nos dieron conclusiones como:
6. La elección de la fuente bibliográfica y el cuidado en el desarrollo de estrategias de búsqueda afectan el proceso general y los resultados cuando se estudia la actividad científica.

7. El método desarrollado tiene una vocación cíclica, lo que permite navegar en el proceso tratando de dar equilibrio a la importancia en cada etapa a pesar de que en nuestro caso las estrategias de búsqueda fueron especialmente importantes. Fácil tránsito y comunicación entre el contenido de las BDb exploradas, dado que nuestro punto de conexión no fue a partir de los descriptores temáticos propios y particulares de cada instrumento, sino la dirección de adscripción de los autores, común a todas las fuentes secundarias usadas. El método es reproducible, siempre y cuando se tenga un buen conocimiento y manejo de las BDb y sus herramientas; en nuestro caso hacer cumplir nuestras diferentes etapas de la metodología. Si bien estadísticamente es mejor trabajar con grandes cantidades de información en ocasiones más es solo más, en nuestro caso pretendíamos establecer una comunicación más directa con el autor del trabajo, es decir que el campo asociado de la dirección de adscripción fuera el que reprodujera a qué y a que no se puede reconocer cómo química, cumpliendo de esta forma con los objetivos del proyecto en el que este trabajo se inserta.

Así podemos vislumbrar en primer lugar que la elección de BDb tendrá siempre un impacto sobre las evaluaciones que realizan los tomadores de decisiones cuando éstos se apoyen en datos cuantitativos para organizar, estructurar, financiar y planificar la investigación científica. Segundo, esta elección tendrá también impacto para aquellos que realizan estudios sociales de la ciencia al utilizar estos datos para hacer un mapeo de un campo o para analizar una actividad científica. Nos queda por determinar cuáles serían estos impactos, lo que podría ser el tema de otros trabajos que estamos preparando dentro del marco de la presente investigación.

El presente estudio se realizó dentro del marco del convenio específico de cooperación "Ciencias, Tecnología y Sociedad" firmado por la Universidad Nacional Autónoma de México y el Institut de Recherche pour le Développement de Francia.

BIBLIOGRAFÍA

- Alonso Gamboa, J. O.; Reyna Espinosa, F. R., "Compilación de datos bibliométricos regionales usando las bases de datos CLASE y PERIÓDICA", en *Revista Interamericana de Bibliotecología*, 2005, vol. 28 (1), 63-78.

- Bolaños-Pizarro, M.; Vidal-Infer, A.; Navarro-Molina, C.; Valderrama-Zurián, J. C.; González-Alcaide, G.; Aleixandre-Benavent, R., “Análisis de la productividad científica y visibilidad de las publicaciones sobre usabilidad (1971-2005)”, en *Revista Española de Documentación Científica*, 2009, vol. 32(1), 9-21.
- Cañedo Andalia, R., “Estudios de solapamiento en la selección de las publicaciones seriadas y las bases de datos”, en *ACIMED*, 1999, vol. 7(3), 164-170.
- Consejo Nacional de Ciencia y Tecnología. CONACyT, *Indicadores de actividades científicas y tecnológicas*. Edición de bolsillo, México 2006, México, CONACyT, 2006.
- _____, *Anexo Estadístico del Informe General del Estado de la Ciencia y la Tecnología*, 2007, Capítulo III, Producción Científica y Tecnológica y su Impacto Económico, en <http://www.siiicyt.gob.mx/siiicyt/cms/paginas/IndCientifTec.jsp> [6 de febrero, 2012].
- Costas, R.; Moreno, L.; Bordons, M., “Solapamiento y singularidad de MEDLINE, WoS e IME para el análisis de la actividad científica de una región en Ciencias de la Salud”, en *Revista Española de Documentación Científica*, 2008, vol. 31(3), 327-343.
- Escalona, M.; Lagar, P.; Pulgarín, A., “Web of Science vs. SCOPUS: un estudio cuantitativo en Ingeniería Química”, en *Anales de Documentación*, 2010, vol.13, 159-175.
- Hood, W. W.; Wilson, C. S., “The scatter of documents over databases in different subject domains: how many databases are needed?”, en *Journal of the American Society for Information Science and Technology*, 2001, vol. 52(14), 1242-1254.
- _____, “Overlap in bibliographic databases”, en *Journal of the American Society for Information Science and Technology*, 2003, vol. 54(12), 1091-1103.
- Kleiche-Dray, M., “Institutionalization of Chemistry in Mexico during the twentieth century (1934-1970)”, en *Proceedings 6th International Conference on the History of Chemistry*, “Neighbours and territories. The evolving identity of Chemistry”, (1) Territories — Building a disciplinary identity—, Making a base for chemistry, pp. 739-750, 2008, Louvain-la-neuve: Mémosciences ASBL.
- Kleiche-Dray, M.; Casas-Guerrero, R., “La institucionalización de un campo científico: El caso de la Química en México en el siglo XX”, en *Redes (Revista de Estudios Sociales de la Ciencia)*, Buenos Aires, 2008, vol. 14, 47-73.
- Levine-Clark, M.; Kraus, J., “Finding chemistry information using Google Scholar: a comparison with *Chemical Abstracts Service*”, en *Science & Technology Libraries*, 2007, vol. 27(4), 3-17.
- Michán, L.; Llorente-Bousquets, J., “Bibliometría de la sistemática biológica sobre América Latina durante el siglo XX en tres bases de datos mundiales”, en *Revista de Biología Tropical*, 2010, vol. 58(2), 531-545.

- Neuhaus, C.; Daniel, H. D., "A new reference standard for citation analysis in Chemistry and related fields on the sections of Chemical Abstracts", en *Scientometrics*, 2009, vol. 78(2), 219-229.
- Pulgarín Guerrero, A.; Escalona Fernández, Ma. I., "Medida de solapamiento en tres bases de datos con información sobre ingeniería", en *Anales de Documentación*, 2007, vol.10, 335-344.
- Red de Indicadores de Ciencia y Tecnología (RICYT), Iberoamericana e Interamericana. Indicadores Comparativos, 2007, <http://www.ricyt.org/interior/interior.asp?Nivel1=1&Nivel2=2&Idioma> [5 de mayo, 2010].
- Russell, J. M.; Rosas, A. Ma.; Arvanitis, R., "Institutional production cutting across disciplinary boundaries: an assessment of chemical research in Mexico", en *Proceedings 5th International Conference of the International Society for Scientometrics and Informetrics*, 1995, p. 485-493. Medford, N. J. :LearnedInformation.
- Saavedra-Fernández, O.; Sotolongo-Aguilar, G.; Guzmán-Sánchez, M. V., "Medición de la producción científica en América Latina y el Caribe en el campo agrícola y afines: un estudio bibliométrico", en *Revista Española de Documentación Científica*, 2002, vol. 25(2), 151-161.
- Salvador Oliván, J. A.; Angós Ullate, J. M.; Fernández Ruíz, J., "Comparación y evaluación de las bases de datos ERIC e ISA sobre el tema 'Recuperación de la Información'", en *Revista Española de Documentación Científica*, 1999, vol. 22(1), 50-63.
- Sampson, M.; McGowan, J., "Errors in search strategies were identified by type and frequency", en *Journal of Clinical Epidemiology*, 2006, vol. 59, 1057-1063.
- Taylor, B.; Wylie, E.; Dempster, M.; Donnelly, M., "Systematically retrieving research: A case study evaluating seven databases", en *Research on Social Work*, 2007, vol. 17, 697-706.
- Thomson Reuters, *Web of Science. Guía de uso*, 2006, http://www.accesowok.fecyt.es/wp-content/themes/Wok/archivos/noticias2008/WebofScienceSpain_Sep08.pdf [17 de mayo, 2012]
- Yerkey, N.; Glogowski, M., "Scatter of Library and Information Science topics among bibliographic databases", en *Journal of the American Society for Information Science and Technology*, 1990, vol. 41(4), 245-253.

