

Recuperación de la información sobre la producción de la investigación científica de la Universidad Nacional Autónoma de México en Scopus y Web of Science

Information retrieval on the production of scientific research of the National Autonomous University of Mexico in Scopus and Web of Science

Carlos Benigno García López,* Jorge Mejía Ruíz, Jovv Valdespino Vázquez,*** Gabriel Cabrera Heredia,*** Crescencio Juárez Flores,*** Elsa Guzmán Rodríguez,** Omar Hernández Pérez.****

RESUMEN

Conocer la producción científica de los investigadores de la Universidad Nacional Autónoma de México (UNAM) es una prioridad en la Coordinación de la Investigación Científica. Con esta finalidad la Dirección General de Bibliotecas (DGB) de la UNAM realizó una investigación teniendo como base a 1 614 investigadores, para conocer su producción científica consultando las bases de datos Scopus y Web of Science. Se analiza la problemática en la búsqueda y recuperación de información bajo el nombre de los investigadores, situación que da origen al Catálogo de Artículos - Coordinación de la Investigación Científica, UNAM (CIC). El CIC contiene las referencias bibliográficas y textos completos de la producción científica de los investigadores pertenecientes a la Coordinación de la Investigación y la normalización (control de autoridad) de los nombres de cada uno de los investigadores e instituciones a las que están afiliados.

PALABRAS CLAVE: Producción científica, Coordinación de la Investigación Científica, Universidad Nacional Autónoma de México, Control de autoridad, Scopus, Web of Science.

*Subdirección Técnica. Dirección General de Bibliotecas (DGB). Universidad Nacional Autónoma de México (UNAM). Primer piso de la Biblioteca Central, Ciudad Universitaria, 14510, México D.F., México. Correo electrónico: carlos_garcia@dgb.unam.mx.

**Departamento de Procesos Técnicos. Dirección General de Bibliotecas (DGB). Universidad Nacional Autónoma de México (UNAM). Basamento de la Biblioteca Central, Ciudad Universitaria, 14510, México D.F., México. Correos electrónicos: tecnico@dgb.unam.mx; elsagr@dgb.unam.mx; omhdbc@dgb.unam.mx.

***Departamento de Análisis, Desarrollo y Mantenimiento de Sistemas. Dirección General de Bibliotecas (DGB). Universidad Nacional Autónoma de México (UNAM). Primer piso de la Biblioteca Central, Ciudad Universitaria, 14510, México D.F., México. Correos electrónicos: jowval@unam.mx; gabriel@dgb.unam.mx.

****Departamento de Catálogo Colectivo de Publicaciones Periódicas. Dirección General de Bibliotecas (DGB). Universidad Nacional Autónoma de México (UNAM). Décimo primer piso de la Biblioteca Central, Ciudad Universitaria, 14510, México D.F., México. Correo electrónico: crescencio@dgb.unam.mx.

Abstract

Knowing the scientific production of researchers from UNAM is of high priority in the Coordination of Scientific Research. To this end the General Directory for Libraries at UNAM conducted an investigation on 1 614 scientific researchers to know their production, by searching Scopus and Web of Science data bases. The problem is analyzed in the retrieval of information when searching by the name of researchers, a situation that gave rise to the creation of the Paper Index - Coordination of Scientific Research UNAM (CIC). CIC contains bibliographic references and full texts of the scientific production of researchers from the Coordination of Scientific Research and the standardization (authority control) of the names of each of the researchers and institutions they are affiliated to.

KEYWORDS: Scientific production, Coordination of Scientific Research, National Autonomous University of Mexico, Authority control, Scopus, Web of Science

Introducción

A petición de la Coordinación de la Investigación Científica de la Universidad Nacional Autónoma de México (UNAM), la Dirección General de Bibliotecas realizó una investigación de 1 614 autores para conocer su producción científica y saber con cuáles instituciones nacionales o internacionales colaboran. Ésta se llevó a cabo en las bases de datos internacionales Scopus y Web of Science y permitió, además, la creación de una nueva base de datos que contiene la producción científica de los investigadores pertenecientes a dicha Coordinación.

Las bases de datos Scopus y Web of Science contienen resúmenes y referencias bibliográficas de literatura científica revisada por pares y juntas reúnen más de 27 mil títulos de revistas. Permiten conocer lo que se publica en el área de la investigación científica, proporcionan información de los autores, su afiliación y la citación de sus trabajos. Las opciones de búsqueda son por autor, título del artículo, título de la revista, afiliación y temas. Son instrumentos de apoyo a la investigación que aportan, en algunos casos, información en texto completo siempre y cuando la biblioteca tenga la suscripción a las revistas.

A fines de 2012 se inició el desarrollo de la base de datos que lleva el nombre de "Catálogo de Artículos - Coordinación de la Investigación Científica, UNAM" (CIC), la cual contiene las referencias obtenidas de Scopus y Web of Science.

Esta base de datos contenía 21 882 registros proporcionados por la Coordinación de la Investigación Científica y venían con elementos mínimos como el autor, título del artículo, título de la revista, volumen, número de la revista, páginas, en mayúsculas compactas las cuales fueron convertidas en altas y bajas. Estos registros fueron investigados en fuentes internacionales y complementados con los demás autores participantes, su función, temas proporcionados por los mismos autores, y se normalizó la o las afiliaciones de los autores.

Más adelante se tomó la decisión de agregar a la nueva base otros documentos tales como los libros publicados y las tesis donde los investigadores hayan tenido alguna participación en la creación o realización de las obras.

Durante el proceso se observó la falta de uniformidad de los nombres de los autores en las bases de datos consultadas. Las inconsistencias generan dispersión de la información, aunada a la falta de un catálogo de autoridad que normalice las diferentes formas del nombre de una persona.

La recuperación de la información se realizó por la afiliación del autor, en este caso por la Universidad Nacional Autónoma de México y por todas las variables del nombre que se detectaron durante la búsqueda, todo lo cual, de inicio, parecía una labor prácticamente imposible.

Inicialmente, se desarrollaron los registros de autoridad de los nombres de 1 614 investigadores de la UNAM y, posteriormente, de otros autores que fueron recuperados en la búsqueda del periodo de 1960 a 2013. Adicionalmente se relacionó el catálogo de autoridad de LIBRUNAM con todos los registros de la base de datos CIC.

Obtención de los registros bibliográficos

Para obtener la mayor parte de la producción científica de los investigadores de la UNAM se determinaron las siguientes acciones:

1. La selección de las bases de datos Scopus y Web of Science, las cuales cuentan con las principales publicaciones periódicas científicas revisadas por pares, además de la citación de los trabajos indizados.
2. La recuperación de los registros del periodo 1960 a 2013
3. La definición de los identificadores: autores, título del artículo, fuente, año de publicación y volumen, entre otros, los cuales deberían estar declarados y ser compatibles con los elementos de la plantilla de la base de datos CIC.
4. La determinación de la afiliación, como la opción de búsqueda más factible para obtener los datos, realizándola por las diferentes variantes de la Universidad Nacional Autónoma de México.
5. La selección del formato Excel, para la salida de los registros y poder transferirlos a la base de datos CIC.
6. Agregar los registros bibliográficos de libros y tesis que se encuentran en los catálogos LIBRUNAM y TESIUNAM, respectivamente.

Obtención de los registros en SCOPUS

En la base de datos Scopus se utilizó la opción “Affiliation search”, la cual permite buscar en el identificador (afiliación) el nombre de una institución relacionada con los autores de los artículos; se ingresa el nombre de la institución y automáticamente se resolverán las variantes de nombre. (Imagen 1)

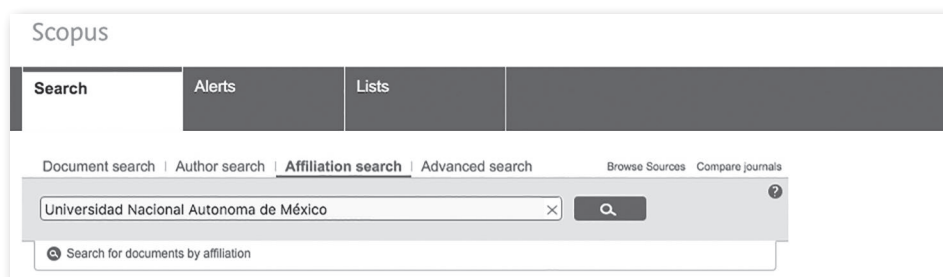


Imagen 1. Opción de búsqueda affiliation search.

Se muestra una lista de todas las instituciones afiliadas en la que puede seleccionarse la institución de interés. Los resultados incluyen el número de documentos que ha publicado una institución y el vínculo. Se puede elegir entre visualizar los resultados por número de documentos, ciudad o país. (Imagen 2)

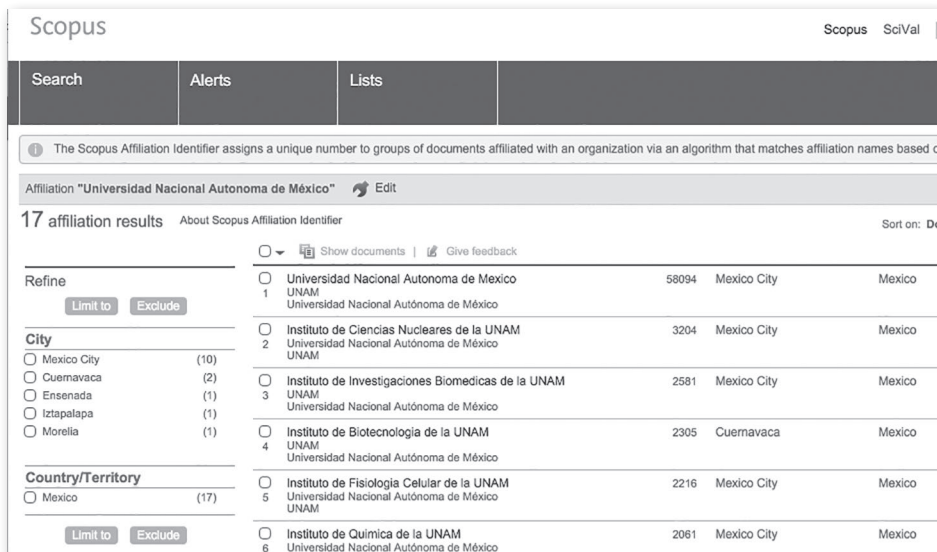


Imagen 2. Presentación de los resultados por afiliación.

Se eligió visualizar el número de documentos por institución para transferirlos al formato de salida Excel. Una limitante de Scopus es que no permite bajar los registros en forma global, por lo que se hizo en bloques de 2 mil registros delimitados con los siguientes identificadores: año de publicación y dentro del año, por área temática, seleccionando solamente aquellos campos que correspondan a la investigación científica. (Imágenes 3 y 4)

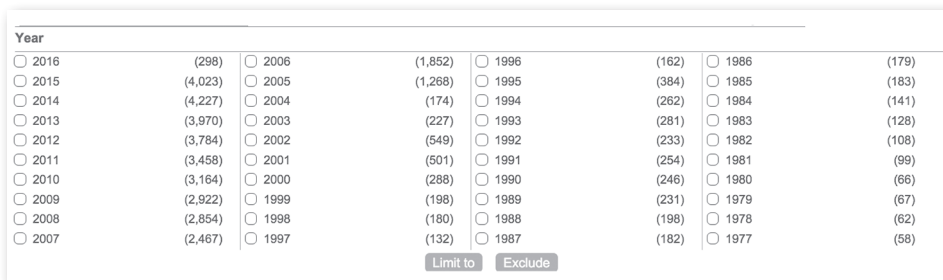


Imagen 3. Opción para limitar los resultados por año de publicación.

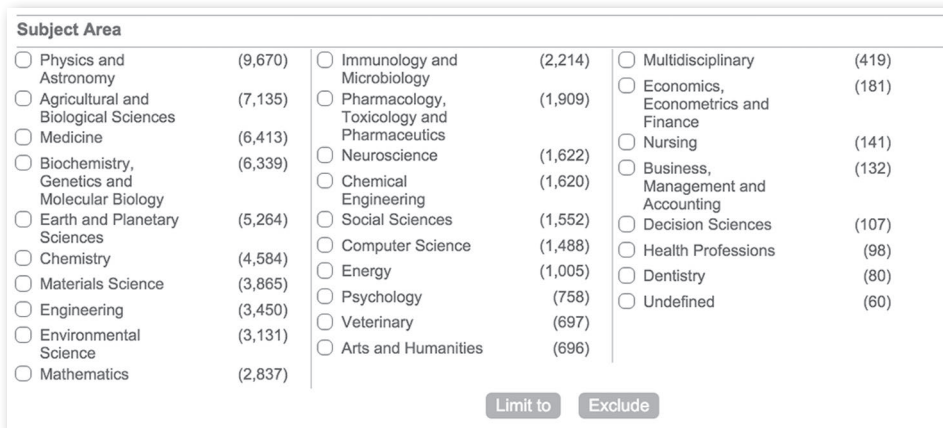


Imagen 4. Áreas temáticas para delimitar los resultados de la búsqueda.

Con el propósito de que la transferencia fuera lo más transparente, se seleccionaron los datos comunes entre las bases Scopus y CIC. (Imagen 5)

Imagen 5. Selección de los campos comunes entre Scopus y CIC.

Por último, se determinó el CSV (Excel) como el formato de salida de los registros para después analizarlos, prepararlos y transferirlos a la base de datos CIC. (Imágenes 6 y 7)

Imagen 6. Opción para seleccionar el formato de salida de los registros.

	A	B	C	D	E	F	G	H	I
1	Authors	Title	Year	Source title	Volume	Issue	Art. No.	Page start	Page end
2	Alvarez-Venicio V., Gutiérrez	Incorporation in Langmuir	2012	Thin Solid Films		526		246	251
3	Grether M., de Llano M.,	Generalized Bose-Einstein	2012	Journal of Superconductivity and Novel Magnetism				1	5
4	Campos I., Jimenez J.L.,	Electromagnetic momentum	2012	Revista Brasileira de Ens		34	2	2303	
5	García-Calderon G., Herr	Delay time as a quantum t	2012	Physical Review A - Ator		86	6	62118	
6	Cordero S., García-Calde	Analytical study of quadrat	2012	Physical Review A - Ator		86	6	62116	
7	Vera-Cruz P., Toscano R.	Synthesis and crystal struc	2012	CrystEngComm		14	24	8606	8614
8	Jaime L.G., Pichardo B.,	Regions of dynamical stabi	2012	Monthly Notices of the I		427	4	2723	2733
9	Holwerda B.W., Pirzkal N	Quantified HI morphology-	2012	Monthly Notices of the I		427	4	3159	3175
10	Abelev B., Adam J., Adar	Measurement of electrons	2012	Physical Review D - Part		86	11	112007	
11	Boyer D., Dean D.S., Mej	Optimal fits of diffusion co	2012	Physical Review E - Stati		86	6	60101	
12	Abelev B., Adam J., Adar	Measurement of the cross	2012	Physical Review Letters		109	25	252302	
13	Flores-Marquez E.L., Gal	Fractal dimension analysis	2012	Nonlinear Processes in C		19	6	693	701
14	Abelev B., Adam J., Adar	Pion, kaon, and proton pro	2012	Physical Review Letters		109	25	252301	
15	Díaz-De-Anda A., Flores	Experimental study of the	2012	Journal of Sound and Vil		331	26	5732	5744
16	Carballo J.L., Cruz-Barral	A revision of the genus My	2012	AIP Conference Proceed		1451	1	165	194
17	Laralde H.	Maximum-entropy distribu	2012	Physical Review E - Stati		86	6	61117	
18	García J.A., Sil J.L.R., Van	Study of colonial manuscri	2012	Materials Research Soci		1374		17	25
19	Ocana S.Z., Pesqueira M	Xavier Guerrero, De MÁ©	2012	Materials Research Soci		1374		73	85
20	Arroyo E., Lara A.C., Espi	The influence of glass in th	2012	Materials Research Soci		1374		61	72
21	Ruvalcaba Sil J.L., Espino	Materials Research Society	2012	Materials Research Soci		1374		ix	xi
22	Melgar E., Solís R., Ruval	Technological and materia	2012	Materials Research Soci		1374		103	114

Imagen 7. Presentación de los archivos en formato Excel para analizarlos, prepararlos y transferirlos a la base de datos CIC.

Obtención de los registros en Web of Science

Para obtener los registros en Web of Science se siguió el mismo procedimiento que en la base de datos anterior. En Web of Science se utilizó la opción “Organizaciones-Nombre preferido”, que permite buscar el nombre preferido o variantes de una institución. Posteriormente se delimitó la búsqueda a Science Citation Index Expanded (SCI-EXPANDED) 1900-presente, para obtener registros que correspondan a la investigación científica. (Imagen 8)

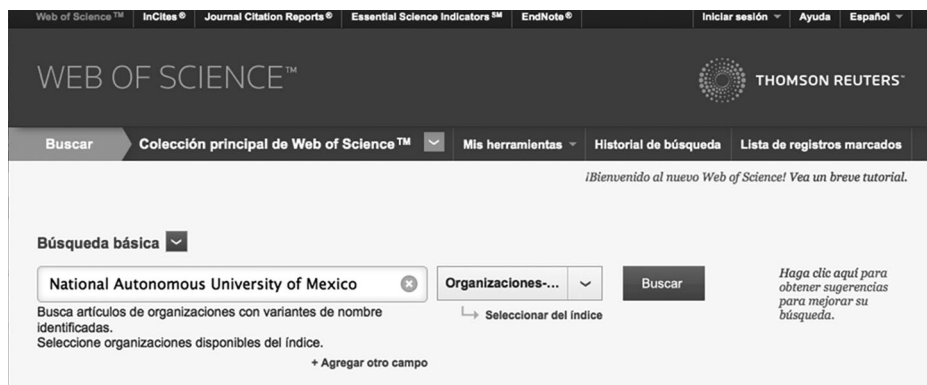


Imagen 8. Opción de búsqueda Organizaciones-Nombre preferido.

En la siguiente lista se muestra el número de resultados obtenidos y la forma de cómo se realizó la búsqueda, refinando los resultados por año de publicación. (Imagen 9)

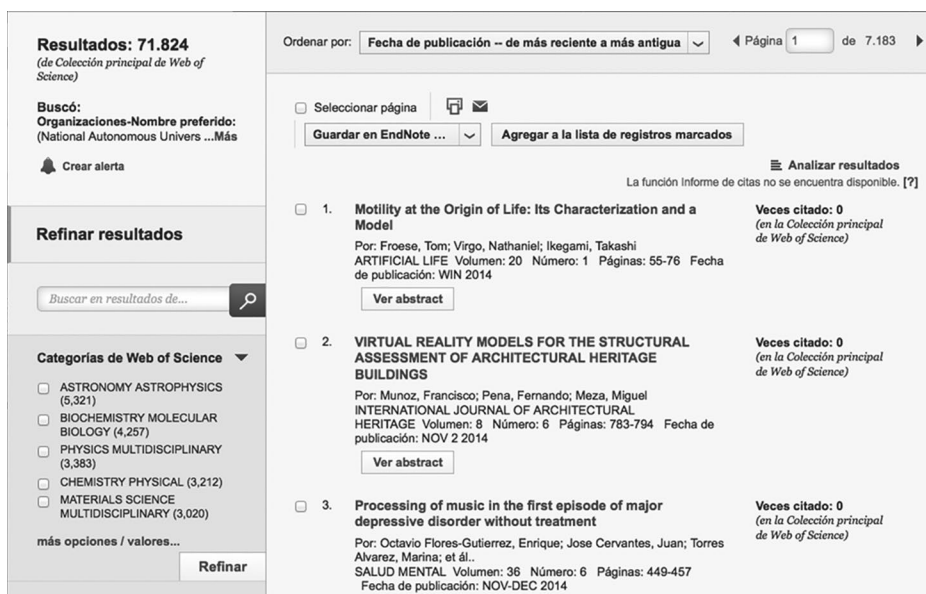


Imagen 9. Presentación de los resultados por Organizaciones-Nombre preferido.

Todos los registros bibliográficos que aparecen en la página de resultados son registros fuente. Éstos proceden de revistas, libros, conferencias y patentes que se indexaron. Cada registro fuente tiene un registro completo al que se puede acceder.

Se seleccionó el formato de salida (Formato delimitado por tabulador (WIN, UTF-8)) para exportar los registros. Una limitante de Web of Science es que no permite exportar los registros en forma global, por lo cual se tuvo que limitar en bloques de 500 registros. (Imágenes 10 y 11)

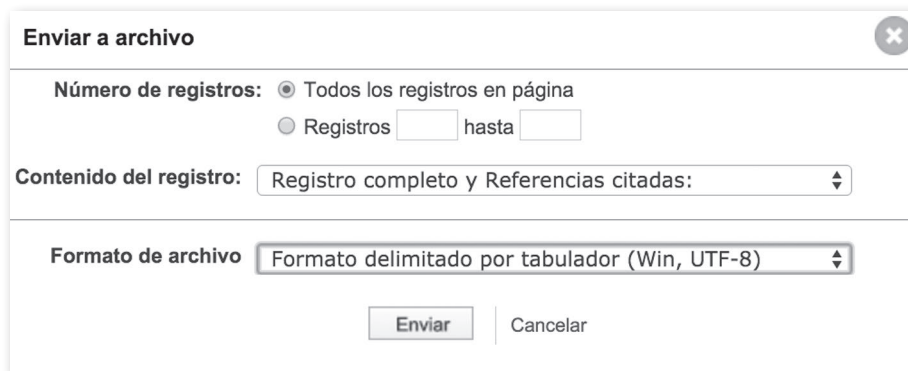


Imagen 10. Opción para seleccionar el formato de salida de los registros.



Imagen 11. Aviso de advertencia que indica la cantidad de registros que pueden exportarse.

Se obtiene un archivo.txt que se convirtió en Excel para ser modificado con los datos comunes entre las bases de datos. (Imagen 12)

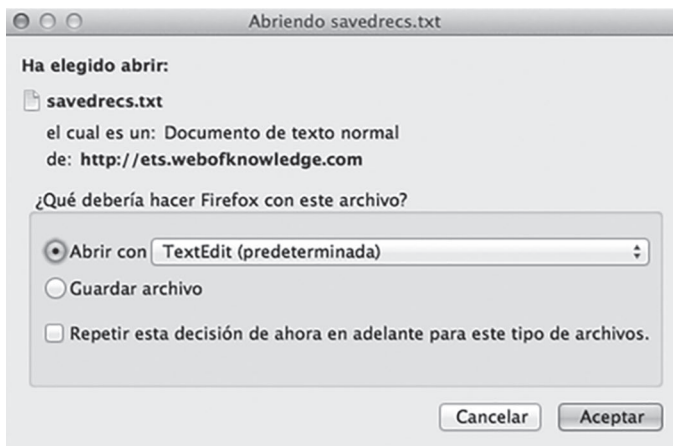


Imagen 12. Archivo.txt con los registros exportados de Web of Science.

Los archivos de salida en formato Excel, con los elementos comunes, son enviados para ser analizados y preparados para transferirlos a la base de datos CIC. (Imagen 13)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	PT	AU	BA	BE	GP	AF	BF	CA	TI	SO	SE	BS	LA
2	J	Arana, MD; Morrone, JJ; Ganem, MA; Luna, ML; Giac Daniel Arana, Marcelo; Morrone, Juan J Diversity and IHERINGIA SERIE BOTANICA											Portuguese
3	J	Cruz-Ramirez, H; Ramirez-Alarcon, R; Morelos, FJ; Qu Cruz-Ramirez, Hector; Ramirez-Alarcon, Observation OPTICS EXPRESS											English
4	J	Alvarez-Venicio, V; Gutierrez-Nava, M; Amelines-Sari Alvarez-Venicio, V.; Gutierrez-Nava, M.; Incorporatio THIN SOLID FILMS											English
5	J	Rovito, SM; Parra-Olea, G; Vasquez-Almazan, CR; Lun Rovito, Sean M.; Parra-Olea, Gabriela; V Deep diverge BMC EVOLUTIONARY BIOLOGY											English
6	J	Unger-Saldana, K; Pelaez-Ballestas, I; Infante-Castane Unger-Saldana, Karla; Pelaez-Ballestas, Developmen BMC CANCER											English
7	J	Trejo-Becerril, C. Perez-Cardenas, E; Taja-Chayeb, L; Trejo-Becerril, Catalina; Perez-Cardenas Cancer Progr PLOS ONE											English
8	J	De La Vega-Beltran, JL; Sanchez-Cardenas, C; Krapf, Luis De La Vega-Beltran, Jose; Sanchez-Mouse Sperr JOURNAL OF BIOLOGICAL CHEMISTRY											English
9	J	Luksic, M; Hribar-Lee, B; Vlachy, V; Pizio, O				Luksic, Miha; Hribar-Lee, Barbara; Vlach Structural an JOURNAL OF CHEMICAL PHYSICS							English
10	J	Villa-Galaviz, E; Boege, K; del-Val, E				Villa-Galaviz, Edith; Boege, Karina; del-V Resilience in PLOS ONE							English
11	J	Ormeno-Orrillo, E; Menna, P; Almeida, LGP; Ollero, F Ormeno-Orrillo, Ernesto; Menna, Pame Genomic bas BMC GENOMICS											English
12	J	Acosta, B; Smolentseva, E; Beloshapkin, S; Rangel, R; Acosta, Brenda; Smolentseva, Elena; Be Gold support APPLIED CATALYSIS A-GENERAL											English
13	J	Ayala, A; Dominguez, CA; Loewe, M; Zhang, Y				Ayala, Alejandro; Dominguez, C. A.; Loew Rho-meson r PHYSICAL REVIEW D							English
14	J	Cordero, S; Garcia-Calderon, G				Cordero, Sergio; Garcia-Calderon, Gastc Analytical sti PHYSICAL REVIEW A							English
15	J	Garcia-Calderon, G; Hernandez-Maldonado, A				Garcia-Calderon, Gaston; Hernandez-M Delay time a: PHYSICAL REVIEW A							English
16	J	Pizio, O; Sokolowski, S; Sokolowska, Z				Pizio, O.; Sokolowski, S.; Sokolowska, Z. Electric doub JOURNAL OF CHEMICAL PHYSICS							English
17	J	Cruz-Cordova, A; Rocha-Ramirez, LM; Ochoa, SA; Gor Cruz-Cordova, Ariadnna; Rocha-Ramire: Flagella from PLOS ONE											English
18	J	Sanchez-Chavez, G; Pena-Rangel, MT; Riesgo-Escovar Sanchez-Chavez, Gustavo; Teresa Pena- Insulin Stimu PLOS ONE											English
19	J	Abelev, B; Adam, J; Adamova, D; Adare, AM; Aggarw: Abelev, B.; Adam, J.; Adam: ALICE Collabi Measuremer PHYSICAL REVIEW D											English
20	J	Boyer, D; Dean, DS; Mejia-Monasterio, C; Osharin, G Boyer, Denis; Dean, David S.; Mejia-Mo Optimal fits r PHYSICAL REVIEW E											English
21	J	De Benedetti, F; Brunner, HI; Ruperto, N; Kenwright, De Benedetti, Fabrizio; Bri PRINTO; PRC Randomized NEW ENGLAND JOURNAL OF MEDICINE											English
22	J	Garcia-Diaz, MT; Lopez, JA; Steffen, W; Richer, MG Garcia-Diaz, Ma. T.; Lopez, J. A.; Steffen A DETAILED I ASTROPHYSICAL JOURNAL											English
23	J	Fowler, MA; Duhamel, J; Bahun, GJ; Adronov, A; Zara Fowler, Michael A.; Duhamel, Jean; Bah Studying Pyri JOURNAL OF PHYSICAL CHEMISTRY B											English
24	J	Abelev, B; Adam, J; Adamova, D; Adare, AM; Aggarw: Abelev, B.; Adam, J.; Adam: ALICE Collabi Measuremer PHYSICAL REVIEW LETTERS											English
25	J	Gil, ALAP; Navarro, LB; Vera, MP; Petricevich, VL				Perez Gil, A. L. Alvarez; Barbosa Navarr: Anti-inflam: JOURNAL OF ETHNOPHARMACOLOGY							English
26	J	Camarillo, L; Luna, R; Nacher, V; Romo, R				Camarillo, Liliana; Luna, Rogelio; Nache Coding perce PROCEEDINGS OF THE NATIONAL ACAD							English
27	J	Jolleys, MD; Coe, H; McFiggans, G; Capes, G; Allan, JC Jolleys, Matthew D.; Coe, Hugh; McFigg Characterizir ENVIRONMENTAL SCIENCE & TECHNOL											English
28	J	Abelev, B; Adam, J; Adamova, D; Adare, AM; Aggarw: Abelev, B.; Adam, J.; Adam: ALICE Collabi Pion, Kaon, a PHYSICAL REVIEW LETTERS											English
29	J	Rosales-Mendoza, S; Govea-Alonso, DO; Monreal-Esc Rosales-Mendoza, Sergio; Govea-Alons: Developing p VACCINE											English
30	J	Ambriz-Tututi, M; Monjaraz-Fuentes, F; Drucker-Colli Ambriz-Tututi, Monica; Monjaraz-Fuent Chromaffin c LIFE SCIENCES											English
31	J	Diaz-de-Anda, A; Flores, J; Gutierrez, L; Mendez-Sanc Diaz-de-Anda, A.; Flores, J.; Gutierrez, L Experimenta JOURNAL OF SOUND AND VIBRATION											English
32	J	Marquez-Ramirez, SG; Delgado-Buenrostro, NL; Chiri Gissela Marquez-Ramirez, Sandra; Lauri Titanium dio: TOXICOLOGY											English
33	J	Lopez-Alarcon, M; Rodriguez-Cruz, M; Vital-Reyes, VS Lopez-Alarcon, Mardias; Rodriguez-Cruz, PPAR gamma GENE											English
34	J	Perez, JD; Carrillo-Salgado, C; Garcia, E; Ruiz-Mar, G; de Jesus Perez-H, Jose; Carrillo-Salgado, Neuroprotec JOURNAL OF NEUROIMMUNOLOGY											English
35	J	Mateos-Jasso, A; Zavala-Hidalgo, J; Romero-Centeno, Mateos-Jasso, Adriana; Zavala-Hidalgo, Variability of CONTINENTAL SHELF RESEARCH											English
36	J	Vazquez, F; del Rio, JA				Vazquez, F.; del Rio, J. A. Thermodyna JOURNAL OF APPLIED PHYSICS							English

Imagen13. Presentación en formato Excel de los archivos para ser analizados y preparados para transferirlos a CIC.

Análisis y preparación de los archivos para su importación a CIC

Antes de almacenar los registros bibliográficos en la base de datos CIC se realizó la normalización de los datos en los archivos Excel de las bases de datos Scopus y Web of Science. El proceso de normalización consistió en determinar las etiquetas MARC21 para cada uno de los elementos que se encontraban en los archivos de salida (archivo Excel). Para realizar la normalización se desarrollaron las siguientes actividades:

1. Diseño de una base de datos en el gestor "MYSQL", con el propósito de almacenar temporalmente la información de los registros bibliográficos para normalizarlos antes de agregarlos al gestor.
2. Desarrollo de una aplicación para la lectura de los archivos con formato Excel y comparación de los registros almacenados en la base de datos a través de los datos de título, tema y DOI.
3. Desarrollo de aplicaciones para transformar la información de la base de datos temporal considerando diversos procesos:
 - a. Agregar etiquetas de acuerdo al formato MARC, así como códigos de sub-campo.
 - b. Generar elementos con información fija requerida por el gestor.
 - c. Establecer claves de acuerdo a catálogos.
 - d. Separar autores, domicilios, instituciones y temas.

Previo a la importación de los registros a CIC se desarrollaron también aplicaciones para normalizar la información contenida en los archivos Excel mencionados anteriormente, ya que su formato no tenía la estructura adecuada para su ingreso.

En el proceso de normalización se diseñó una base de datos temporal con el propósito de realizar las revisiones pertinentes antes de almacenar la información en forma definitiva en la base de datos CIC. De las observaciones detectadas la más importante es la referente a la duplicidad de los registros de los artículos, alcanzando un porcentaje del 12 % aproximadamente, aunque por la forma del contenido había pocas probabilidades de reducirlos ya que no eran iguales. Algunos títulos diferían hasta en 8 caracteres, con patrones diferentes entre los que destacan los siguientes: signo α y la leyenda "alfa" o "α"; signo β y la leyenda "beta" o "β"; números romanos y texto del número; números arábigos y texto del número; errores de tecleo y signos especiales.

La solución para reducir los registros de artículos duplicados de acuerdo a los puntos antes señalados se realizó a través del desarrollo de una aplicación con diversos módulos que efectuaron los siguientes procesos:

- Eliminación de los espacios en el campo de título del artículo.
- Eliminación de los siguientes caracteres: ",,;:[]{}+?%&/-i*+'='_!|#\$()<>.
- Conversión a letras mayúsculas.
- Clasificación alfabética de la cadena resultante de las etapas anteriores.
- Aplicación del algoritmo de distancia de Levenshtein (1965), que consiste en que dadas dos secuencias de caracteres debe medir cuáles son los cambios mínimos para que a partir de una cadena de caracteres obtengamos la otra. En el caso de la duplicidad de los títulos de los artículos se tomó la cadena del nombre del artículo y se comparó con los siguientes cinco registros de cadenas de artículos, respetando el orden en el que fueron establecidos por el proceso de clasificación y, además, definiendo parámetros acerca de cuántos fallos se permitirían en las comparaciones.

Archivo de salida	Proceso de comparación
10Boron distribution measurement in laser ablated B4C thin films using (n,?) reaction and LR-115 passive detector	(10)boron distribution measurement in laser ablated b4c thin films using (n,alpha) reaction and lr-115 passive detector
A variation of menger theorem for long paths	A variation of Menger's theorem for long paths
Leaf survival in relation to herbivory in 2 tropical pioneer species	Leaf Survival in Relation to Herbivory in Two Tropical Pioneer Species
1-t2 (1-?), a basic function to quantify magnetic measurements in high-Tc superconductors at low external magnetic fields.	1-t2/(1-alpha), a basic function to quantify magnetic measurements in high-Tc superconductors at low external magnetic fields

Tabla 1. Algunos ejemplos de las cadenas de caracteres que fueron detectadas en el proceso de comparación.

Normalmente los procesos de comparación de cadenas de caracteres se realizan contra un directorio o índice, pero el resultado concluye determinando únicamente si son o no iguales. En el caso de los títulos de los artículos era importante verificar que hubiera el mínimo de duplicidad, con el propósito de proporcionar servicios de búsqueda adecuados.

A continuación se presentan algunos ejemplos de las cadenas de caracteres que fueron detectadas en el proceso de comparación.

Problemática

Una vez que los registros de Scopus y Web of Science fueron importados a la base de datos CIC se detectaron una serie de problemas en los índices de autor y afiliación de la CIC. Esta situación provocó que los resultados en la búsqueda no fueran confiables. Entre éstos destacan:

- Las diferentes formas de registrar a las instituciones afiliadas de los autores.
- Las diferentes formas de registrar a los autores de los artículos.
- El cambio de orden de los autores de acuerdo al documento original.
- Autores diferentes bajo una misma forma del nombre.

Como ejemplo, el caso del nombre de la Universidad Nacional Autónoma de México. La forma variante UNAM se encontró con puntos y espacios, con puntos sin espacios, diferentes formas abreviadas del nombre completo, con errores de digitación, con forma variante del nombre en otro idioma (inglés):

U. N. A. M.

U.N.A.M.

Univ. Nac. Auton.

Univ. Nac. Auton. de México

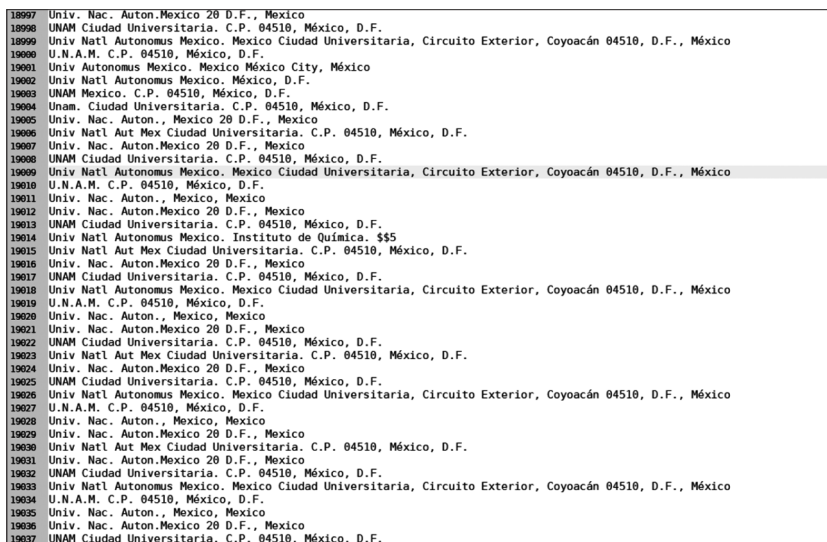
Univ. Nac. Auton. Mexico

Univ. Nac. Auton., Mexico

National Autonomous University, Mexico

National Autonomous University, Mexico City, Mexico

Con respecto a las entidades subordinadas en el índice de CIC, por afiliación se detectaron una gran variedad de formas variantes. Otro problema fue el de los códigos postales y direcciones de las diferentes afiliaciones, las cuales aparecían con inconsistencias de digitación, en abreviaturas y en puntuación. Este problema se corrigió directamente en el archivo de las afiliaciones, se eliminaron las direcciones y se unificaron los organismos; además, se detectaron autores sin o con diferentes afiliaciones. (Imagen 14)



18997	Univ. Nac. Auton.Mexico 28 D.F., Mexico
18998	UNAM Ciudad Universitaria. C.P. 04510, México, D.F.
18999	Univ Natl Autonomus Mexico. Mexico Ciudad Universitaria, Circuito Exterior, Coyoacán 04510, D.F., México
19000	U.N.A.M. C.P. 04510, México, D.F.
19001	Univ Autonomus Mexico. Mexico México City, México
19002	Univ Natl Autonomus Mexico. México, D.F.
19003	UNAM Mexico. C.P. 04510, México, D.F.
19004	Unam. Ciudad Universitaria. C.P. 04510, México, D.F.
19005	Univ. Nac. Auton., Mexico 20 D.F., Mexico
19006	Univ Natl Aut Mex Ciudad Universitaria. C.P. 04510, México, D.F.
19007	Univ. Nac. Auton.Mexico 28 D.F., Mexico
19008	UNAM Ciudad Universitaria. C.P. 04510, México, D.F.
19009	Univ Natl Autonomus Mexico. Mexico Ciudad Universitaria, Circuito Exterior, Coyoacán 04510, D.F., México
19010	U.N.A.M. C.P. 04510, México, D.F.
19011	Univ. Nac. Auton., Mexico, Mexico
19012	Univ. Nac. Auton.Mexico 28 D.F., Mexico
19013	UNAM Ciudad Universitaria. C.P. 04510, México, D.F.
19014	Univ Natl Autonomus Mexico. Instituto de Química. 555
19015	Univ Natl Aut Mex Ciudad Universitaria. C.P. 04510, México, D.F.
19016	Univ. Nac. Auton.Mexico 28 D.F., Mexico
19017	UNAM Ciudad Universitaria. C.P. 04510, México, D.F.
19018	Univ Natl Autonomus Mexico. Mexico Ciudad Universitaria, Circuito Exterior, Coyoacán 04510, D.F., México
19019	U.N.A.M. C.P. 04510, México, D.F.
19020	Univ. Nac. Auton., Mexico, Mexico
19021	Univ. Nac. Auton.Mexico 28 D.F., Mexico
19022	UNAM Ciudad Universitaria. C.P. 04510, México, D.F.
19023	Univ Natl Aut Mex Ciudad Universitaria. C.P. 04510, México, D.F.
19024	Univ. Nac. Auton.Mexico 28 D.F., Mexico
19025	UNAM Ciudad Universitaria. C.P. 04510, México, D.F.
19026	Univ Natl Autonomus Mexico. Mexico Ciudad Universitaria, Circuito Exterior, Coyoacán 04510, D.F., México
19027	U.N.A.M. C.P. 04510, México, D.F.
19028	Univ. Nac. Auton., Mexico, Mexico
19029	Univ. Nac. Auton.Mexico 28 D.F., Mexico
19030	Univ Natl Aut Mex Ciudad Universitaria. C.P. 04510, México, D.F.
19031	Univ. Nac. Auton.Mexico 28 D.F., Mexico
19032	UNAM Ciudad Universitaria. C.P. 04510, México, D.F.
19033	Univ Natl Autonomus Mexico. Mexico Ciudad Universitaria, Circuito Exterior, Coyoacán 04510, D.F., México
19034	U.N.A.M. C.P. 04510, México, D.F.
19035	Univ. Nac. Auton., Mexico, Mexico
19036	Univ. Nac. Auton.Mexico 28 D.F., Mexico
19037	UNAM Ciudad Universitaria. C.P. 04510, México, D.F.

Imagen 14. Archivo en el que se unificaron las formas variantes del nombre de la UNAM, entidades subordinadas y eliminación de dirección.

Una vez realizados los cambios de los casos mencionados se ingresaron a la base de datos CIC en la forma correcta, mostrándose ahora una sola forma autorizada del nombre de la UNAM y de sus entidades subordinadas. (Imagen 15)

No. Registros	Índice: INS - Afiliación
3988	Universidad Nacional Autónoma de México
1	Universidad Nacional Autónoma de México. Acad Fis, Colegio Ciencia & Tecnol
1	Universidad Nacional Autónoma de México. Acad Geociencias
1	Universidad Nacional Autónoma de México. Acad Inst Quim
1	Universidad Nacional Autónoma de México. Anim Care Unit, Fac Med
1	Universidad Nacional Autónoma de México. Anim Unit, Sch Psychol
1	Universidad Nacional Autónoma de México. Appl Math & Syst Inst
1	Universidad Nacional Autónoma de México. Appl Math & Syst Res Inst
1	Universidad Nacional Autónoma de México. Appl Sci Ctr, Precis Engrn Lab
1	Universidad Nacional Autónoma de México. Area Radiol, Secc Imagenol, Fac Med Vet & Zootecnia, Hosp Vet Especialidades, Depto Med Cirugia & Zooto




Imagen 15. Índice de afiliación unificada del nombre de la UNAM.

El índice de afiliación incluye las instituciones de educación superior extranjeras; en el caso de los artículos en colaboración la afiliación de los otros autores. (Imagen 16)

No. Registros	Índice: INS - Afiliación
53	University of California
1	University of California. Astronomy Department
22	University of California at Berkeley
11	University of California at Berkeley. Astronomy Department
1	University of California at Berkeley. Department of Civil and Environmental Engineering
3	University of California at Berkeley. Department of Integrative Biology
2	University of California at Berkeley. Geography Department
57	University of California at Berkeley. Lawrence Berkeley Laboratory
1	University of California at Berkeley. Mechanical Engineering Department
12	University of California at Berkeley. Museum of Vertebrate Zoology






Imagen 16. Índice de afiliación de las instituciones de educación superior extranjeras (artículos en colaboración).

Otro problema que afectaba la recuperación de información fue la sinonimia en la forma de los nombres de los autores, por ejemplo: Cuevas, S (Cuevas, Salvador, y Cuevas, Sergio). En ambos casos el sistema juntó los artículos de autores diferentes bajo una misma forma; una vez hechos los respectivos registros de autoridad se determinó su atribución uno a uno. (Imágenes 17 y 18)

008	121129 a az n aab a n d
035	a MX011000065269
040	a MXMxUNAM b spa c MXMxUNAM e rda
1001	a Cuevas García, Sergio
372	a Dinámica de fluidos, magnetohidrodinámica, termodinámica de procesos irreversibles
373	a Centro de Investigación en Energía, Universidad Nacional Autónoma de México
374	a Profesor a Investigador
375	a hombre
377	a spa a eng
4001	a García, Sergio Cuevas
4001	a Cuevas, Sergio
4001	 a Cuevas, S. q (Sergio)
670	a Dinámica y transferencia de calor en flujos de metales líquidos en campos magnéticos intensos, [tesis para obtener el grado de Doctor en Ciencias (Física), UNAM], 1994: b portada (presenta Sergio Cuevas García)
SYS	000065269

Imagen 17. Registro de autoridad para evitar conflicto de sinonimia en la forma de un nombre (Cuevas, Sergio).

010	a nr2002032121
035	a MX011000062561
040	a CaQMCCA b spa c CaQMCCA d MXMxUNAM e rda
1001	a Cuevas, Salvador
371	a Instituto de Astronomía, Circuito Exterior s/n, Ciudad Universitaria, Coyoacán b Ciudad de México d México e 04510 m chavoc@astro.unam.mx u http://www.astrocu.unam.mx/IA/index.php?option=com_content&view=article&id=440&Itemid=46&lang=es
372	a Desarrollo de instrumentación de alta resolución espacial utilizando óptica adaptativa y alta resolución espectral
373	a Departamento de Instrumentación, Instituto de Astronomía, Universidad Nacional Autónoma de México
375	a hombre
377	a spa a eng
4001	a Cuevas Cardona, Salvador
4001	a Cardona, Salvador Cuevas
4001	a Cuevas Cardona, Salvador Carlos
4001	a Cardona, Salvador Carlos Cuevas
4001	a Cuevas, S. q (Salvador)
670	a Light and time, a masterpiece, c2001: b portada portada (Salvador Cuevas)
670	a Caracterización de sistemas hexápodos para instrumentación en grandes telescopios, [asesor de tesis para obtener el título de Físico, UNAM], 2011: b portada (asesor Salvador Cuevas Cardona)
670	a Scopus, 24 agosto, 2012: b (Cuevas, Salvador; variante: Cuevas, S.; National Autonomous University of Mexico, Institute of Astronomy) u http://www.scopus.com/authid/detail.url?authorId=7004196258&eid=2-s2.0-84856530572
SYS	000062561

Imagen 18. Registro de autoridad para evitar conflicto de sinonimia en la forma de un nombre (Cuevas, Salvador).

Los registros de autoridad que permiten la identificación de una persona incluyen los siguientes elementos: nombre preferido, nombre variante, lugar de nacimiento, fechas asociadas, campo de actividad, afiliación, género, idioma y la fuente que se consultó para obtener los datos. (Imagen 19)

<i>Nombre aceptado</i>	100	1	a	Ostrosky, Patricia
<i>Dirección</i>	371	—	a	Instituto de Investigaciones Biomédicas, Sede del Circuito Escolar 1er. piso, Lab. A122, Ciudad Universitaria, Delegación Coyoacán Ciudad de México México 04510 ostrosky@unam.mx Medicina genómica Toxicología ambiental Departamento de Medicina Genómica y Toxicología Ambiental, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México Profesora Investigadora mujer spa eng
<i>Campo activada</i>	372	—	a	
<i>Alimación</i>	373	—	a	
<i>Ocupación</i>	374	—	a	
<i>Género</i>	375	—	a	
<i>Lengua</i>	377	—	a	
<i>Ac. no acept.</i>	400	1	a	Ostrosky Shejet, Martha Patricia
<i>Ac. no acept.</i>	400	1	a	Ostrosky Shejet, Martha P. (Patricia)
<i>Ac. no acept.</i>	400	1	a	Shejet, Martha Patricia Ostrosky
<i>Ac. no acept.</i>	400	1	a	Shejet, Martha P. Ostrosky (Patricia)
<i>Ac. no acept.</i>	400	1	a	Ostrosky, P. (Patricia)
<i>Ac. no acept.</i>	400	1	a	Ostrosky Wegman, Patricia
<i>Ac. no acept.</i>	400	1	a	Wegman, Patricia Ostrosky
<i>Ac. no acept.</i>	400	1	a	Ostrosky Wegman, P. (Patricia)
<i>Ac. no acept.</i>	400	1	a	Ostrosky, Wegman P. (Patricia)
<i>f. fuente</i>	670	—	a	La radiación al servicio de la vida, 1991: portada (Patricia Ostrosky) p. 4 de la portada (Patricia Ostrosky es bióloga egresada de la UNAM, maestría en ciencias biomédicas-genética en la Universidad de Tel-Aviv y el doctorado en ciencias biomédicas-farmacología en la UNAM)

Imagen 19. Registro de autoridad con los elementos núcleo para la identificación de una persona.

Carlos Benigno García López, Jorge Mejía Ruíz, Jovv Valdespino Vázquez, Gabriel Cabrera Heredia, Crescencio Juárez Flores, Elsa Guzmán Rodríguez, Omar Hernández Pérez

También se incluyó el ISNI (International Standard Name Identifier), con el fin de que se le identifique unívocamente en procesos como catalogación y localización de documentos. (Imagen 20)

ISNI:	0000 0000 8118 6468
Name:	Morán-Zenteno, Dante Morán Zenteno, Dante J. Zenteno, Dante J. Morán
Creation class:	Language material Text
Creation role:	author
Related names:	American association of petroleum geologists Sanchez-Barreda, Luis Wilson, James Lee (1920-...)
Titles:	Geología de la República Mexicana geology of the Mexican Republic, The
Notes:	
Sources:	VIAF BNF LC BOWKER

Imagen 20. Número ISNI asignado y agregado al registro de autoridad del investigador.

Otra finalidad de crear registros de autoridad para los nombres de los investigadores es permitir la recuperación de un artículo sin que importe la forma que haya elegido el autor como nombre preferido, por ejemplo, el investigador Pérez Ramírez José G., utiliza el seudónimo Xim Bokhimi para firmar sus trabajos. (Imagen 21)

1001	a Bokhimi, Xim
370	a El Oro de Hidalgo, Estado de México
371	a Instituto de Física, Departamento de Física Química, Cubículo 46, Ed. Principal, Ciudad Universitaria b Ciudad de México d México e 04510 m bokhimi@fisica.unam.mx u http://www.fisica.unam.mx/departamentos/fisquim/index.php?option=com_investigadores&idInvestigador=67&Itemid=2
372	a Ciencia y Tecnología del México Milenario a Orden Local en Sólidos
373	a Departamento de Física Química, Instituto de Física, Universidad Nacional Autónoma de México
374	a Investigador a Profesor
375	a hombre
377	a spa a eng
4001	a Bokhimi, Xi
4001	a Bokhimi, X.
4000	a Xim Bokhimi
4001	a Pérez Ramírez, José Guadalupe
4001	a Ramírez, José Guadalupe Pérez
4001	a Pérez Ramírez, José G.
4001	a Ramírez, José G. Pérez
4001	a Pérez Ramírez, J. G. q (José Guadalupe)
4001	a Ramírez, J. G. Pérez q (José Guadalupe Pérez)
670	a Síntesis y caracterización de nanoestructuras de sulfuro de bismuto, [asesor de tesis para obtener el grado de Doctor en Ciencias Químicas, UNAM], 2008: b portada (tutor: Dr. José Guadalupe Pérez Ramírez)
670	a Scopus, 17 agosto, 21012: b (Bokhimi, Xim ; variantes: Bokhimi, Xi.; Bokhimi, X; Xim Bokhimi; Universidad Nacional Autónoma de México, Instituto de Física) u http://www.scopus.com/authid/detail.uri?authorId=35586881400&eid=2-s2.0-79953728222
670	a Scopus 22 agosto 2012: b (Pérez-Ramírez, José Guadalupe; variantes: Pérez Ramírez, J. G.; Pérez Ramírez, José G.; Instituto de Física) u http://www.scopus.com/authid/detail.uri?authorId=7007162412

Imagen 21. Registro de autoridad que ayuda al usuario a buscar a una persona sin que importe la forma que elija para su nombre preferido.

La importancia de los catálogos de autoridad radica en permitir la recuperación bajo cualquier forma del nombre del autor (preferida o variante) y guiar al usuario a sus obras.

Además, en esta base de datos se desarrollaron registros de autoridad para las afiliaciones de los autores de todas las entidades académicas existentes, incluyendo las entidades subordinadas.

Obtención de los registros en LIBRUNAM y TESIUNAM

Con el propósito de tener completa la producción científica de los 1 614 investigadores de la Coordinación de la Investigación Científica se consideró apropiado agregar otros recursos a la base de datos CIC.

Los registros de los nuevos recursos se tomaron de los catálogos desarrollados por la DGB:

- De los libros en los que el investigador tuviera una participación como creador o como colaborador y se encuentren registrados en el catálogo LIBRUNAM. (Imagen 22)
- De las tesis en las que el investigador tuviera participación como asesor, director o sinodal, entre otras modalidades, además de su propia tesis, y se encuentren registradas en el catálogo TESIUNAM. (Imagen 23)

Artículo	El agua en México vista desde la Academia / Blanca Jiménez, Luis Marín, editores ; Dante Morán, Óscar Escolero, Javier Alcocer, coordinadores ; Víctor H. Martínez, edición digital
Descr. Física	403 páginas : ilustraciones, mapas
Restricciones	Acceso sólo para usuarios de REDUNAM
Restricciones	*
Formato adicional	Disponible también en formato impreso
Formato adicional	Disponible también en formato impreso
Tema	<ul style="list-style-type: none"> • Abastecimiento de agua -- México. • Agua -- Leves y legislación -- México. • Agua -- Contaminación -- México.
Tema	<ul style="list-style-type: none"> • Libros electrónicos.
Autor	<ul style="list-style-type: none"> • Jiménez Cisneros, Blanca E. (Blanca Elena), editor • Marín, Luis (Luis Ernesto), editor • Morán Zenteno, Dante J., editor • Escolero Fuentes, Oscar A., editor • Alcocer Durand, Javier editor • Martínez, Víctor H., editor
Coautor Corporativo	<ul style="list-style-type: none"> • Academia Mexicana de Ciencias.
Rec. electrónico	 Texto completo :

Imagen 22. Registro LIBRUNAM agregado a la base de datos CIC.


No. de sistema	000048806
Autor	<ul style="list-style-type: none"> • Ramírez Perez Amaya, Erick Yaotzin sustentante
Artículo	Estratigrafía Cretácico-Terciaria de la región de Jolalpan-Huehuetlan, estado de Puebla / tesis que para obtener el título de Ingeniero Geologo, presenta Erick Yaotzin Ramirez Perez Amaya ; asesor Enrique Alejandro Gonzalez Torres , Dante Jaime Moran Zenteno.
Descr. Física	ii, 63 páginas : ilustraciones.
Tesis	Ingeniero Geologo UNAM, Facultad de Ingeniería, 2005
Restricciones	Acceso en línea sin restricciones star
Restricciones	*
Autor	<ul style="list-style-type: none"> • Gonzalez Torres, Enrique Alejandro, asesor • Morán Zenteno, Dante J., asesor
Coautor Corporativo	<ul style="list-style-type: none"> • Universidad Nacional Autonoma de Mexico, Facultad de Ingeniería, institución que otorga el título
Rec. electrónico	 Texto completo :

Imagen 23. Registro TESIUNAM agregado a la base de datos CIC.

La base de datos cic

Algunos datos importantes de la base de datos CIC:

Registros	86 972 registros
Artículos cosechados en Scopus y Web of Science.	74 538
Artículos en texto completo.	43 213 (57.8 %)
Registros duplicados entre ambas bases datos.	10 744
Obras monográficas.	2 653 registros
Tesis donde los investigadores están como asesores o sustentantes.	9 781 registros

La base de datos CIC tiene varias opciones de búsqueda que permiten al usuario satisfacer sus necesidades de información: ya sea por palabras o por índice específico. (Imagen 24)

Imagen 24. Opciones de búsqueda de la base de datos cic, por palabras o por índice.

En la opción de búsqueda Otros índices el usuario puede recuperar información de elementos más específicos, tales como título de la revista y afiliación del investigador. (Imagen 25)

Imagen 25. Opciones de búsqueda específicas disponibles en cic.

Una de las ventajas de la base de datos CIC es la relación que tiene con los catálogos de autoridad (temas, personas y entidades corporativas), que permite la unificación de las formas autorizadas o variantes. (Imágenes 26 y 27)

No. Registros	Índice: AUT - Autor
74	Morán Zenteno, D. (Dante) - [Registro de autoridad]
74	Morán Zenteno, D. J. (Dante Jaime) - [Registro de autoridad]
74	Morán Zenteno, Dante - [Registro de autoridad]
74	Morán Zenteno, Dante J. - [Registro de autoridad]
3	Morán Zenteno, Dante J. (Dante Jaime)
74	Morán Zenteno, Dante Jaime - [Registro de autoridad]
2	Morancho, A.H.
1	Morancho, Ah
1	Morandi, M.T.
26	Morando, M




Imagen 26. Índice de autor donde se muestra la relación con el catálogo de autoridad [Registro de autoridad] para la normalización del nombre de una persona.

No. Registros	Índice: SUB - Tema
27	Petróleo - [Registro de autoridad]
1	Petróleo -- Aspectos ambientales -- Sonda de Campeche
4	Petróleo (Combustible) - [Registro de autoridad]
4	Petróleo como combustible - [Registro de autoridad]
1	Petróleo, Industria del - [Registro de autoridad]
1	Petróleo -- Lodos de perforación - [Registro de autoridad]
1	Petróleo -- Perforación de pozos - [Registro de autoridad]
12	Petróleo pesado - [Registro de autoridad]
13	Petróleo -- Pozos - [Registro de autoridad]
1	Petróleo -- Pozos -- Perforación - [Registro de autoridad]




Imagen 27. Índice de temas donde se muestra la relación con el catálogo de autoridad [Registro de autoridad] para la normalización de los temas que tratan los documentos.

Una vez realizada la búsqueda, en los registros se pueden visualizar elementos básicos para su identificación. (Imagen 28)

No. de sistema	000083532
Artículo	Geochemistry of thermal springs and geodynamics of the convergent mexican pacific margin
Descr. Física	páginas 251-262
Tema	<ul style="list-style-type: none"> • Continental margin • Subduction zone • Hot springs • He-isotopes • Rivera and cocos plates • Mexico • Slab-derived fluids • Fore-arc region • Distribucion geografica. • He-3/he-4 ratios • Southern mexico • Jalisco block • Isotopic composition • Southwestern mexico • Corteza continental. • Subducted rivera
Autor	<ul style="list-style-type: none"> • Taran Sobol, Iouri autor. Universidad Nacional Autónoma de México. Instituto de Geofísica. • Morán Zenteno, Dante J. autor. Universidad Nacional Autónoma de México. Instituto de Geología.

Imagen 28. Visualización de un registro en cic con los elementos básicos para su identificación.

Conclusiones

Las bases de datos de Scopus y Web of Science conjuntamente son una herramienta de información científica de gran valor para los investigadores de la Universidad Nacional Autónoma de México ya que permite conocer toda su producción.

El haber completado las formas abreviadas de los nombres de las entidades subordinadas le dio consistencia a la base de datos de artículos de investigación científica.

En la base CIC se puede recuperar información por diferente afiliación de los autores y por entidades subordinadas como institutos y centros de investigación y direcciones generales.

Los catálogos de autoridad representan una buena alternativa para proporcionar uniformidad y consistencia en la recuperación de los principales puntos de acceso de los artículos de investigación científica debido a que con una sola búsqueda recupera las formas variantes del nombre del autor a través del nombre preferido.

Sería conveniente que Scopus y Web of Science vincularan los registros de autoridad de autor con los que cuentan con los registros de los artículos, esto les permitiría una mejor recuperación de información.

Además, la base CIC se puede actualizar semestralmente o anualmente. 

Obras consultadas

Scopus [en línea]. <<http://www.scopus.com/>> [Consulta: 4 septiembre 2015].

Web of Science [en línea]. <<http://wokinfo.com/webtools/searchbox/>> [Consulta: 4 septiembre 2015].

