
Superposición entre las citas del Web of Science y Scopus: un estudio exploratorio

Overlapping between the Web of Science and Scopus: an explorative study

Suyin ORTEGA CUEVAS (1), Sergio MÁRQUEZ RANGEL (2), Leticia LÓPEZ HUERTA (3)

(1, 3) Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Biblioteca, Circuito escolar Ciudad Universitaria, México. D.F. C.P. 04510, suyin.ortega@iimas.unam.mx, leticia.lopez@iimas.unam.mx. (2) Dirección General de Bibliotecas, Circuito de la Investigación Científica, Ciudad Universitaria, México. D.F. C.P. 04510, smaquez@dgb.unam.mx

Resumen

Se presenta un estudio exploratorio de la superposición de citas entre el Web of Science y Scopus y sus implicaciones para los autores, los investigadores y las bibliotecas, realizado en la Biblioteca del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas de la Universidad Autónoma de México sobre las citas recibidas por las publicaciones de sus investigadores para los años 2011-12.

Palabras clave: Análisis de citas. Bases de datos de citas. Web of Science. Scopus. Superposición. Bibliotecas. Matemáticas. Méjico.

1. Introducción

El medio más común que utilizan los investigadores para dar a conocer los avances de sus trabajos son los artículos en revistas especializadas. Las bases de datos se encargan de difundir este conocimiento de forma simplificada, como citas bibliográficas. Por medio de estas referencias bibliográficas de las bases de datos, se dan a conocer los trabajos e investigaciones que se realizan en los institutos; y, por medio de las citas bibliográficas que reciben los artículos, se puede conocer de forma dinámica si los trabajos de investigación realizados en los institutos son consultados.

El número de citas es un indicador muy usado para la evaluación de los trabajos publicados en revistas académicas y sus autores. De ahí la importancia de localizar hasta la más mínima cita en los diferentes índices que se tienen para consultar. Para ello, los autores y las instituciones recurren a bases de datos de citas. Las citas han sido desde la década de 1950 un tema de interés gracias al surgimiento del Science Citation Index (SCI), la base de datos que contenía las referencias de los artículos publicados en un grupo seleccionado de revistas y que permitía localizar las citas de los artículos originales.

Abstract

An exploratory study of the coverage overlapping between the Web of Science and Scopus and its implications for authors, researchers and libraries is presented, which was carried out in the library of the Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas of the Universidad Autónoma de México for their researchers' citations for the period 2011-12.

Keywords: Citation analysis. Citation databases. Web of Science. Scopus. Overlapping. Libraries. Mexico. Mathematics.

Hoy en día, Web of Science (WoS) (Thomson Reuters, 2013ab) —que incorpora el SCI y otros servicios— y Scopus (Sciverse 2013abc) son los dos servicios más usados acceder a las citas que reciben las publicaciones científicas. Recientemente, ha entrado un nuevo agente de alcance internacional, Google Scholar, que aprovecha los índices de la empresa matriz, y que permite documentar citas no cubiertas por las otras dos bases de datos, que siguen un cuidadoso procedimiento de selección para determinar su cobertura.

Una función importante del Área de Servicios Especializados de la Biblioteca del Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (IIMAS) de la Universidad Autónoma de México es ayudar a los investigadores y a la institución a documentar las citas recibidas para facilitar sus procesos de evaluación interna y externa.

2. Objetivos

Revisando la bibliografía concerniente a las bases de datos del Web of Science y Scopus, se aprecia como tema recurrente la evaluación de los registros que proporcionan, surgiendo el interrogante de si existe duplicidad en esta información. Aprovechando la experiencia de diez años de elaboración de los análisis de citas por

nuestro Área de Servicios Especializados, se tomó la determinación de afrontar una investigación al respecto, apoyándose en otras investigaciones anteriores (Hood, 2005; Bar-Ilan, 2007; López-Illescas, 2008; Santa, 2010).

El estudio realizado tiene carácter exploratorio y se realizó sobre una muestra de dos años (2011 y 2012) de citaciones a los artículos que se producen en el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS).

3. Metodología

La metodología para elaborar este estudio, consistió en una revisión exhaustiva de las citas obtenidas en las bases de datos del WoS y de Scopus.

En WoS, se ejecutó la siguiente estrategia de búsqueda:

Address=(IIMAS) OR Address=(inst* inv* mat* apl*) OR Address=(inst* res* appl* math* sys*)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

Se obtuvieron 1.096 referencias con 9.563 citas; y una vez eliminadas de la hoja Excel las referencias que no tuvieran citas en el año de 2011-2012, quedaron reducidas a 445 referencias y 890 citas. En el Gráfico 1, se observa que la mayor parte de citas y referencias fue eliminada, quedando un pequeño conjunto uniendo las referencias y las citas exclusivamente del WoS.

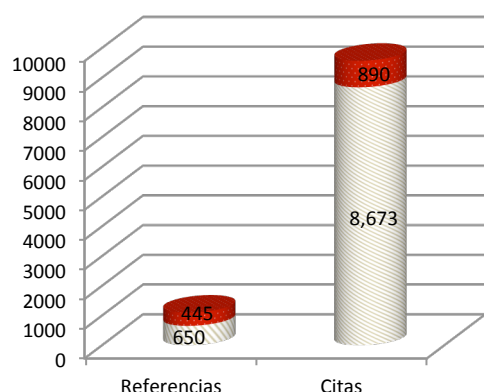


Gráfico 1. Referencias y citas recuperadas del WoS 2011-2012

En Scopus se realizó la siguiente estrategia:

Your query: (AFFIL(inst* res* appl* math* sys*) OR AFFIL("inst* inv* mat* apl*")) OR AFFIL("Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas")

En esta ocasión el resultado obtenido fue de 519 referencias con 4.163 citas. Exportando el archivo a una tabla Excel, se procedió también a eliminar las referencias que no tuvieran citas en el año de 2011-2012, resultando 201 referencias con 798 citas (Gráfico 2).

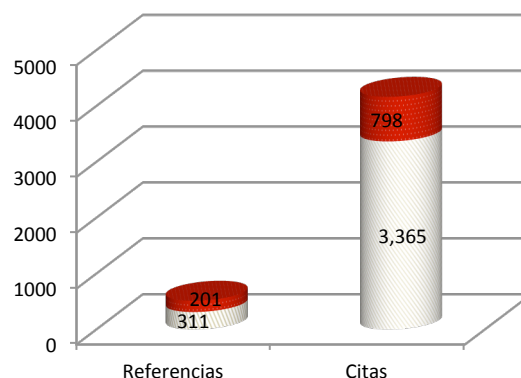


Gráfico 2. Citas y referencias obtenidas de Scopus 2011-2012

El siguiente procedimiento fue unir las referencias de las dos bases de datos en una hoja Excel, colocándolas en orden alfabético, obteniendo como resultado 646 referencias. Se eliminaron primeramente las que no se duplicaban en las bases de datos, quedando como resultado 276 referencias con 667 citas. (Aquí es importante destacar que las referencias que se eliminaron por no estar duplicadas fueron 370.)

Del resultado de 276 referencias, se procedió a identificar las que tuvieran número de citas iguales y desiguales, tal como se muestra en la siguiente tabla, en la que se destacan en rojo los números de citas iguales y en negro las desiguales (Figura 1).

| | | | | |
|-----|--|---|---|--------|
| 134 | Inverse spectral-scattering problem with two sets of discrete spectra for the radial Schrödinger | 2 | 0 | WOS |
| 135 | Inverse spectral-scattering problem with two sets of discrete spectra for the radial Schrödinger | 2 | 0 | Scopus |
| 136 | Jacobi matrices with rapidly growing weights having only discrete spectrum | 0 | 1 | WOS |
| 137 | Jacobi matrices with rapidly growing weights having only discrete spectrum | 0 | 1 | Scopus |
| 138 | KLM model for lossy piezoelectric transducers | 5 | 1 | WOS |
| 139 | KLM model for lossy piezoelectric transducers | 6 | 2 | Scopus |
| 140 | LMST-based searching and broadcasting algorithms over Internet graphs and peer-to-peer c | 1 | 0 | WOS |
| 141 | LMST-based searching and broadcasting algorithms over Internet graphs and peer-to-peer c | 0 | 1 | Scopus |
| 142 | Location-based localized alternate, disjoint and multi-path routing algorithms for wireless | 1 | 0 | WOS |
| 143 | Location-based localized alternate, disjoint and multi-path routing algorithms for wireless | 1 | 2 | Scopus |
| 144 | Lower energy icosahedral atomic clusters with incomplete core | 1 | 0 | WOS |
| 145 | Lower energy icosahedral atomic clusters with incomplete core | 1 | 0 | Scopus |

Figura 1. Identificación de referencias duplicadas

En los resultados obtenidos se localizaron 138 referencias que cubren la estrategia. De éstas, 84 tenían igual número de citas en cada año (el

60%) y 54 (el 40%) eran diferentes desde una hasta 5 citas de diferencia, lo que nos indica una disparidad en la información (Gráfico 3).

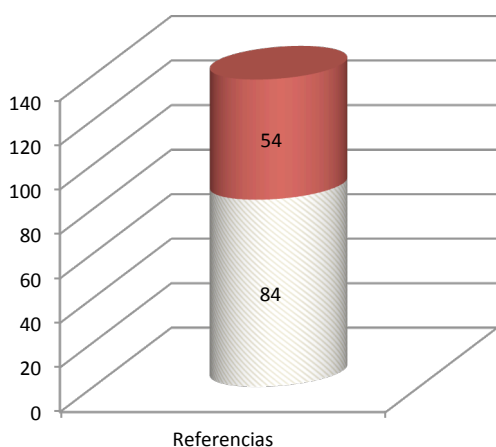


Gráfico 3. Referencias sin citas en WOS y Scopus

4. Resultados

Después del análisis anterior, se buscaron las 138 referencias en las dos bases de datos, pero, para facilitar el trabajo de cotejo de las citas, se hizo por separado: primero, las 84 referencias en las que coincidía el número de citas; y posteriormente, las 54 referencias que mostraban diferencias.

La Tabla I muestra que, para el primer grupo, el número de citas es muy semejante, mientras que, para el segundo, el número de citas crece exponencialmente y es en Scopus donde se localizan más.

| Referencias | Citas iguales | Citas desiguales |
|----------------|---------------|------------------|
| Scopus | 84 (184) | 54 (342) |
| WoS | 84 (172) | 54 (490) |
| Total de citas | 84 (356) | 54 (832) |

Tabla I. Número de referencias y (citas)

En la Tabla II se muestra el número de citas localizadas en los 84 artículos de WoS y Scopus, que suman 356 citas, de las cuales el 51% corresponden a Scopus y el 48% a WoS.

| Índice | Referencias | Citas |
|--------|-------------|-------|
| Scopus | 84 | 184 |
| WOS | 84 | 172 |

Tabla II. Referencias y citas recuperadas por base de datos

Una vez que se unieron las citas en un solo listado, se procedió a identificar cuántas de estas 356 eran realmente únicas. Resultó que 160 eran citas duplicadas en ambas bases; y, por tanto, quedaron solo 28 como únicas para Scopus y 12 para WoS, por lo que la diferencia entre las bases de datos resultó ser mínima (Tabla III).

| Índice | Citas | Únicas | Duplicadas |
|--------|-------|--------|------------|
| Scopus | 184 | 28 | 160 |
| WoS | 172 | 12 | |

Tabla III. Número de citas, citas únicas y duplicadas por base de datos

Otra variable interesante que se estudió fueron los tipos de documentos citados. En la Tabla IV se observa que ambas bases de datos también consideran otros materiales además de artículos de revistas científicas, en particular libros y conferencias, entre otros. En el caso de las citas recuperadas de WoS el 95% de estos materiales son artículos y el 4% corresponden a conferencias. El porcentaje en Scopus es muy similar, ya que el 91% son artículos de revistas, el 7.5% las conferencias y el 1.5% en libros.

| Tipo de documento | WoS | Scopus |
|-------------------|-----|--------|
| Artículos | 164 | 168 |
| Conferencias | 7 | 13 |
| Libros | 1 | 3 |
| TOTAL | 172 | 184 |

Tabla IV. Tipo de documentos en las citas duplicadas por base de datos

El segundo grupo analizado estuvo constituido por las 54 referencias que se buscaron en WoS y Scopus y que tenían ya de entrada un número de citas distinto. Los 54 artículos arrojaban un total de 832 citas, que se revisaron a detalle para validar la cantidad de únicas y duplicadas. La mayor parte de las citas se localizaron en Scopus (Tabla V).

| Índice | Referencias | Citas |
|--------|-------------|-------|
| Scopus | 54 | 490 |
| WoS | 54 | 342 |

Tabla V. Referencias y citas recuperadas

Del total de 832 citas obtenidas en el periodo consultado se identificaron 310 citas duplicadas, es decir, que el 37% de las citas se encuentra

en ambas bases de datos. Scopus cuenta con el mayor número de citas únicas 186; mientras que WoS solo tiene 32.

| Índice | Citas | Únicas | Duplicadas |
|--------|-------|--------|------------|
| Scopus | 490 | 186 | 310 |
| WoS | 342 | 32 | |

Tabla VI. Total de citas únicas y duplicadas entre WoS y Scopus

Como se podrá observar en las Tablas III y VI hay una pequeña variación, estas referencias que faltan en la suma de las citas en Scopus se debe a que dentro de la misma base de datos están duplicados los registros, caso que no se ve en WoS.

Por otro lado, también se analizó el tipo de material en el que aparecen estas citas (Tabla VII), observando que en WoS el 83.5% de la muestra correspondía a artículos, el 15.5% a conferencias y el 1% a libros; mientras que en Scopus el 66% estaba identificado como artículos, el 29 % como conferencias y 5% como libros.

| Citas | WoS | Scopus |
|--------------|-----|--------|
| Artículos | 284 | 326 |
| Conferencias | 54 | 141 |
| Libros | 4 | 23 |
| TOTAL | 342 | 490 |

Tabla VII. Tipo de documentos en las citas únicas por base de datos

Finalmente, haciendo un acopio de los resultados en ambas bases de datos, el total de citas recuperadas fue de 1558 y en la figura 2 se aprecia cómo se dividieron estas citas.

El total de citas analizadas en el periodo de 2011-2012 fue de 1558, entre las dos bases de datos. De las citas recuperadas, se eliminaron 370, pues al no tener una referencia con la cual comparar no podían ser parte de este estudio, disminuyendo un 24% la muestra obtenida. Del 76% restante que se analizó, se identificaron como citas duplicadas el 30% —es decir 470 citas— y como citas únicas el 46% —718 citas—.

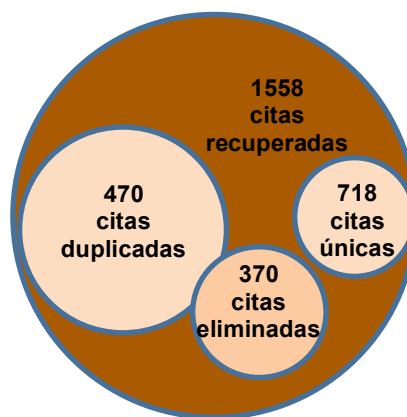


Figura 2. Total de citas recuperadas

5. Conclusiones

De acuerdo con lo anterior, las dos bases de datos analizadas sirven como complemento la una de la otra, ya que en ambas se detectaron una gran cantidad de citas únicas en un periodo muy corto y reciente.

En los resultados se detectó recíprocamente que el 82% de las citas se duplican entre ambas bases, lo que demuestra también que hay una gran coincidencia en los contenidos que analizan una y otra.

Las citas duplicadas obligan al especialista a realizar una doble búsqueda y un trabajo muy delicado de depuración, ya que los investigadores y las instituciones desean ofrecer datos reales, contabilizando citas únicas a sus trabajos de investigación.

Al realizar las búsquedas, se detectaron citas duplicadas dentro de la propia base de datos Scopus, lo que refuerza la necesidad de analizar con cuidado y detenimiento los resultados obtenidos.

Finalmente, el hecho de que se aprecie duplicidad en las citas trabajando tan sólo con las dos bases de datos más importantes, sugiere la relevancia de replicar estos estudios con todas las otras bases de datos que ofrecen este servicio, como Google Scholar.

Referencias

- Archambault, Eric; Campbell, David; Gringras, Yves; Larivière, Vincent (2009). Comparing biometric statistics obtained from the Web of Science and Scopus. // *Journal of the American Society for Information Science and Technology*. 60:7, 1320-1326.
- Ballew, Barbara S. (2009). Elsevier's Scopus database. // *Journal of electronic Resources in Medical Libraries*. 6: 245-252.

- Bar Ilan, Judith; Levene, Mark; Lin, Ayelet.(2007). Some measures for comparing citation databases. // *Journal of Informetrics*. 1: 26-34.
- Buchanan, Robert A. (2006). Accuracy of cited references: the role of citation databases. // *Collegue and research libraries*. 67:4, 292-303.
- De sutter, Bjorn; Den Oord, Aaron (2012). To be or not to be cited in computer science. // *Communications of the ACM*. 55:8 (Agosto 2012) 69-75.
- Escalona Fernández, M. I.; Lagar Barbosa, Pilar; Pulgarin Guerrero, Antonio. (2010). Web of Science vs. SCOPUS: Un estudio cuantitativo en ingeniería química. // *Anales de documentación*. 13: 159-175.
- Gavel, Ylva; Iselid, Lars (2007). Web of Science and Scopus: a journal title overlap study. // *Online Information Review*. 32:1, 8-21.
- Gorraiz, Juan; Schloegl, Christian (2008). A bibliometric analysis of pharmacology and pharmacy journals: Scopus versus Web of Science. // *Journal of Information Science*. 34: 715-725.
- Granda-Orive; José Ignacio de; Alonso-Arroyo, Adolfo; Roig-Vázquez, Francisco (2011). Which data base should we use for our literature analysis? Web of Science versus SCOPUS = ¿Qué base de datos debemos emplear para nuestros análisis bibliográficos? Web of Science versus SCOPUS. // *Archivos de Bronconeumonía*. 47:4, 213-217.
- Hood, W. W; Wilson, C. S. (2005) The relationship of records in multiple databases to their usage or citedness. // *Journal of the American Society for Information Science And Technology*. 56: 1004-1007.
- Huang, Hong; Andrews, James; Tang, Jiang (2012). Citation characterization and impact normalization in bioinformatics journals. // *Journal of the American Society for Information Science and Technology*. 63:3, 490-497.
- Jacso, Peter (2009). Database source coverage: hypes, vital signs and reality checks. // *Online Information Review*. 33:5, 997-1007.
- Kosmulski, Marek (2012). The role of references in scientific papers: cited papers as objects of research. // *Research Evaluation*. 21: 87-88.
- Kousha. Kayvan; Thelwall, Mike; Rezaie, Somayeh (2010). Using the web for research evaluation : the integrated online impact indicator. // *Journal of Informetrics*. 4, 124-135.
- Li, Jie; Burnham, Judy F; Lemley, Trey; Britton, Robert M. (2010). Citation analysis: comparison of Web of Science, Scopus, SciFinder and Google Scholar. // *Journal of Electronic Resources in Medical Libraries*. 7, 196-217.
- Lopez-Illescas, Carmen; de Moya, Félix; Moed, Henk F. (2008). Coverage and citation impact of oncological journals in the Web of Science and Scopus. // *Journal of Infometrics*. 2, 304-316.
- Meho, Lokman I.; Rogers Yvonne (2008). Citation counting, citation ranking and h-index of human-computer interaction researchers: a comparison of Scopus and Web of Science. // *Journal of the American Society for Information Science and Technology*. 59:11, 1711-1726.
- Santa, Samaly; Herrero-Solana, Víctor (2010). Cobertura de la ciencia de América Latina y el Caribe en Scopus vs Web of Science. // *Investigación Bibliotecológica*. 24:52 (Sep-oct.) 13-27.
- Sciverse (2013a). What does Scopus cover?. // Elsevier. <http://www.info.sciverse.com/scopus/scopus-in-detail/facts>. (2013-03-01).
- Sciverse (2013b). Scopus Content Selection and Advisory Board. // Elsevier. <http://www.info.sciverse.com/scopus/csab>. (2013-03-01).
- Sciverse (2013c). Scimago Research Group. // Elsevier. <http://www.scimago.es/>. (2013-03-01).
- Thomson Reuters (2013). History of citation indexing. // Science Home. http://thomsonreuters.com/products_services/science/free/essays/history_of_citation_indexing/ (2013-03-01).
- Thomson Reuters (2013). Web of Science. // Science : Products A-Z http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science/ (2013-03-01).

Enviado: 2012-04-19. Segunda versión: 2013-07-15.
Aceptado: 2013-08-26.
