

Comparación científica entre Web of Science (WoS) y Google Académico: estudio a partir de los autores más representativos de Brasil

Adilson Luiz Pinto

Universidad Federal de Mato Grosso (Brasil)

José Antonio Moreiro-González

Universidad Carlos III de Madrid (España)

Resumen

El objetivo de esta medición es analizar comparativamente la productividad y el índice de citación de los diez autores brasileños que muestran mayor frecuencia en ISI, y que pueden denominarse *Top 10 Brasil*. En la fase metodológica se ha realizado una comparación de la aparición de esos autores en Web of Science (WoS) y en un motor de búsqueda de cuarta generación como es Google Académico (GA). La recuperación de los datos en ambos recursos se hizo mediante búsqueda simple por autor. A la hora de refinar los datos en WoS utilizamos ProCite, mientras que para el GA empleamos FlashGet versión 1.9.2.1028. Para establecer la comparación entre los dos recursos nos hemos servido de índices científicos, así como de las características de cada fuente de información, analizando la parte cuantitativa y la cualitativa. Las representaciones de los datos se han visualizado mediante mapas de redes sociales, cuadros y tablas.

Palabras clave: Web of Science. Google Académico. Índice de frecuencia. Índice de citación. Índice *h*. Red social – colaboración científica. Análisis comparativo.

Abstract

The aim of this measurement is to analyze comparatively the productivity and the citation index of ten Brazilian authors which show major frequency in ISI, and that which can be named as *Top 10 Brazil*. In the methodological phase we have made a comparison between the appearance of these authors in Web of Science (WoS) and in a fourth generation search engine called Google Scholar (GS). The information recovery in both resources was done by author in a simple search. To refine the information in WoS we used ProCite, whereas for GS we utilized FlashGet version 1.9.2.1028. To establish the comparison between both resources we have employed scientific indexes, as well as the characteristics of every information

resource, analyzing the quantitative part and the qualitative one. The information representations have been visualized by social networks maps, pictures and tables.

Keywords: Web of Science. Google Scholar. Index of frequency. Citation index. *h* index. Social network – scientific collaboration. Comparative analysis.

1. Introducción

El potencial de las bases de datos y de los recursos abiertos, como sucede con los motores de búsqueda, ha atraído en gran medida la atención de los investigadores, en especial cuando se aplican a fuentes fiables como son los paquetes del Institute for Scientific Information (Science Citation Index, Social Science Citation Index y Arts & Humanities Citation Index), denominado en el medio académico Web of Science (WoS), y el motor de búsqueda académico desarrollado por Google (Google Académico – GA).

Respecto a los motores de búsqueda, hemos alcanzado ya la cuarta generación. Ahora son más ricos en contenidos y pueden aplicarse a entornos especializados, como sucede con Google Académico (Google, 2005), buscador especializado desde el año 2004 en el mundo académico y que sigue, por una parte, las publicaciones de documentos en la Web, así como las citas recibidas y las tipologías documentales (Jacso, 2005). Su fin consiste en jerarquizar los resultados valiéndose de un algoritmo similar al que utiliza Google para las búsquedas generales, aunque al mismo tiempo emplea herramientas específicas a la hora de mejorar los resultados, con lo que consigue encontrar las informaciones solicitadas con mayor precisión y actualización (Gutiérrez Gallardo y López Guisado, 2005). Asimismo, puede actuar como un excelente recurso estadístico para averiguar y comparar la producción científica de un país, una institución o una autoridad (Noruzi, 2005).

Sin embargo, algunos investigadores (Thelwall, 2002; Bar-Ilan, 2005) han observado que el balance general de los recursos de los motores de búsqueda en la pantalla puede resultar “engañoso”, debido a la falta de transparencia en los algoritmos manejados para llegar al cálculo exacto de las búsquedas (Smith, 2006). Otro asunto relevante es la comprobación general de recursos, que incluye bastante material proveniente de esferas no científicas (Thelwall, 2003), principalmente porque el contenido de Google Académico reporta informaciones personales y no solamente de artículos científicos y libros.

El motor de búsqueda Google Académico ofrece también al usuario la posibilidad de recuperar el número de citas que cada documento ha recibido durante su existencia en Internet, con la ventaja de indicar las tipologías bibliográficas en las que estas apariciones han ocurrido.

El análisis de citas es uno de los criterios más frecuentemente empleados para seleccionar la información científica (Price, 2006), en especial a la hora de formar

y desarrollar colecciones o de hacer estudios de los fondos, así como para conocer la colaboración y el consumo de información dentro de una comunidad científica.

A través del análisis de citas se puede conocer la intensidad del empleo de una revista o de la obra de un referido autor, lo que proporciona una medida razonable de su importancia científica.

La utilización de las citas bibliográficas como criterio de evaluación de los artículos y de las publicaciones científicas se fundamenta en su carácter condicionante de las principales etapas de la investigación científica.

Por otra parte, las citas en Google Académico pueden considerarse un factor importante a la hora de comparar países con menor índice científico y menor visibilidad en las grandes bases de datos tales como WoS (Bar-Ilan, Levene y Lin, 2007). Este proceso de contrastar el potencial de un país, o incluso de una autoridad, es intrigante, porque hasta hace no muchos años no podíamos imaginar que Internet, con sus duplicaciones de documentos y una indización supuestamente descontrolada, llegaría a ser un indicador científico, lo que sucedió con la llegada de los motores de cuarta generación y de contenidos académicos.

Contando con estos hechos nos propusimos desarrollar un estudio sobre la producción y las citas recibidas por los diez autores brasileños más productivos según WoS, así como comparar los índices recibidos por ellos en Google Académico, de forma que se pudiesen visualizar tanto su índice científico como las citaciones, pretendiendo de esta forma evaluar la confiabilidad de esta herramienta respecto a los productos del sistema ISI.

Este estudio está centrado en los autores más importantes de Brasil, pero podría ser desarrollado en otro país, o entre autores aleatorios dentro de Web of Science, o incluso en otra base de datos de gran relevancia mundial, como Scopus.

Además, puede decirse que la metodología empleada introduce cierta innovación a la hora de enfrentarse al análisis de los datos ofrecidos por la Web y de tratarlos de una forma simple y eficiente.

2. Metodología empleada

Se trataba de aplicar el índice científico de visibilidad de los autores a los repertorios WoS y Google Académico mediante una limitación temporal de cinco años, entre 2002 y 2006, al tiempo que se averiguaba el grado de citación y de cocitación para conocer la intensidad de las citas recibidas y el grado de las citaciones en colaboración con otros autores. Para ello se aplicó un índice *h*, que permite visualizar comparativamente la tasa de citación de los autores, y un índice de los documentos representados en ambas fuentes, y posteriormente se determinaron las relaciones mediante una tasa de frecuencia y centralidad.

A la hora de seleccionar ambos repertorios percibimos que los países en desarrollo científico todavía no estaban utilizando las plataformas web a la hora de identificar la visibilidad de sus autores y de establecer comparaciones con otras fuentes, como sería el caso entre WoS y Google Académico.

La importancia de la comparación se deriva de los propios recursos manejados, que en la actualidad están considerados entre los mejores repertorios bibliográficos y que incluso ofrecen índices de citaciones. En este sentido, Google Académico resulta muy atractivo por contener también un índice del número de citaciones recibidas y referirlo a los documentos donde aparecen las citas, imitando el proceso que caracterizó la importancia del WoS en el medio académico.

Esta iniciativa de Google Académico puede funcionar principalmente en las instituciones que no tengan acceso a paquetes científicos, que son muchas en el caso de universidades del tercer mundo, porque estas no invierten en acceso a las bases de datos científicas.

En Brasil existe el Portal Capes, que es una iniciativa brasileña para dar a las 163 instituciones del consorcio acceso público a más de 90 bases de datos científicas, entre ellas Web of Science, Pro-Quest, Scopus, Wilson Web y WebSpirs, pero incluso aunque el país tenga acceso a estos paquetes es interesante que hagamos una comparación a modo de demostración.

Para recuperar la información en WoS hicimos un corte por país (CU = Brasil), que luego matizamos según la categoría de los autores seleccionados, para identificar así los diez más representativos en esa base de datos de acuerdo con el modelo de élite de los autores más importantes $E = \sqrt{N}$ (Price y López Piñero, 1973).

Respecto a Google Académico empleamos una estrategia inicial basada en buscar a partir de los nombres de los diez autores más importantes obtenidos anteriormente en WoS, para realizar después una búsqueda por el nombre completo de cada autor. Sin embargo, al considerar que esta no era la forma más correcta de recuperar con fiabilidad los datos, consultamos el manual de Google y decidimos usar el formato de búsqueda en forma de citación (*longo e*), de modo que conseguimos una recuperación de datos más fiable.

En ambos recursos aplicamos después un filtro con objeto de poder identificar el solapamiento de los registros. No encontramos ninguno en el caso de WoS, pero en Google Académico aparecieron 356 documentos duplicados o con caracteres sin identificar, debido a que esta herramienta tiene un sistema de control para atajar este problema, aunque pueden darse duplicaciones causadas por la presencia de títulos en otros idiomas distintos al inglés.

Realizamos un filtrado de los datos en Google Académico excluyendo toda la información de filiación de departamento y dejando solo datos de publicaciones. De estos hicimos un nuevo filtrado para seleccionar las duplicaciones y el sola-

pamiento de citas, pues la preocupación no era únicamente filtrar la información duplicada en cuanto a visibilidad, sino también en cuanto a las citas recibidas por los autores estudiados.

Así, en WoS recuperamos un total de 1946 registros, de los cuales 1201 eran artículos, y en Google Académico obtuvimos 21 551 documentos, entre artículos científicos, trabajos presentados en congresos, libros y documentos mencionados como [CITAS], de los que 9030 eran artículos científicos, los cuales fueron incluidos como muestra.

El tratamiento de los registros recuperados en WoS se llevó a cabo empleando ProCite (Saxton, 2001), ya que resulta adecuado para el análisis de los autores y de los documentos según el formato estándar ISI-CE. Para el tratamiento de los datos provenientes de Google Académico se empleó FlashGet versión 1.9.2.1028 con soporte de Internet Explorer, y posteriormente se exportaron a un archivo en formato XLS, que puede ser leído con Excel para Windows.

En relación con la representación gráfica de las colaboraciones científicas utilizamos CreatePajek con objeto de transferir los archivos creados en formato XLS a un lenguaje en .NET; después se utilizó Pajek (Batagelj y Mrvar, 2003) y NetDraw (Qian y Gross, 1999) para su conversión en imagen. Para la obtención y representación de la red de relaciones y coocurrencias se acudió a CiteSpace (Chen, 2006).

3. Resultados y discusión

A partir de la comparación de los textos obtenidos en WoS y en Google Académico puede afirmarse que aparecen como datos similares 1) la información necesaria para identificar una publicación (descripción bibliográfica); 2) los artículos de publicaciones científicas del tipo *peer review*; 3) los documentos y referencias de publicaciones *open access*; 4) las referencias de otros tipos de publicaciones, como libros, cartas científicas y *abstracts*; 4) los índices de citación.

Sin embargo, la ausencia de paralelismo entre ambas fuentes se deriva de que en Google Académico es posible encontrar tesis y documentos depositados en repositorios científicos (*e-prints*), *pre-prints* y *post-prints*, mientras que no lo es en WoS. Por el contrario, Google Académico no contiene índices bibliográficos con datos de patentes, frente a lo que ocurre en WoS, que sigue la línea marcada por otros recursos web, como Scirus (www.scirus.com).

Para obtener los resultados cuantitativos del análisis se optó por dividirlos en dos grupos, uno relativo a la frecuencia científica y otro basado en el índice de citas de los autores estudiados, con objeto de poder comparar la eficiencia de Google Académico con relación a las bases de datos de Web of Science solamente en artículos científicos.

Los autores examinados representan a las universidades más importantes de Brasil, como la Universidad Estadual de Campinas (Unicamp), la Estadual Paulista (Unesp), la de São Paulo (USP), la Federal de São Paulo (Unifesp), la Federal de Río de Janeiro (UFRJ), la Federal de Ceará (UFC) y la Federal de São Carlos (UFSCar).

Autor	Universidad	Área	ISI		Google Académico	
			Rango	Frec.	Rango	Frec.
Longo, Elson	Unesp	Química	1	282	1	2700
Tufik, Sérgio	Unifesp	Medicina	2	269	5	737
Varela, José Arana	Unesp	Química	3	216	3	964
Wardell, James Lewis	UFRJ	Química	4	187	7	570
Costa, Fernando Ferreira	Unicamp	Medicina	5	174	4	857
Ramires, José A. Franchini	USP	Medicina	6	173	8	552
Cendes, Fernando	Unicamp	Medicina	7	166	2	1300
Leite, Edson Roberto	UFSCar	Química	8	163	6	574
Souza Filho, Antônio Gomes	UFC	Física	9	161	10	321
Eberlin, Marcos Nogueira	Unicamp	Química	10	155	9	455

Tabla I: Autores más productivos en ISI y Google Académico.

Los diez autores más destacados alcanzaron un índice superior al del resto de autores que aparecen en la plataforma ISI, debido sobre todo a que su producción entre 2002 y 2006 fue mayor que la suma total de la producción de artículos científicos (1819) de todos los investigadores brasileños en 1981.

Ante estos datos es importante hacer referencia a las características de estos investigadores, que muestran una estrecha relación científica con autores destacados de las áreas médicas, químicas y físicas, como C. Glidewell, J. M. S. Skakle, A. Jorio, C. A. Paskicimas, C. Riccardi, S. Roizenblatt, M. Sonata, E. Orhan, T. M. Boschi, L. H. W. Gowdak, A. C. H. F. Sawaya, W. A. Hueb, E. J. H. Lee, L. A. Papale, R. Sparrapan, J. M. Annichino-Bizzacchi, Z. Brankovic, E. M. Krieger, C. E. Rochitte, A. Gruneis, J. M. Tsutsui, V. D. Bolzani, M. Grinberg, M. Wajngarten, C. H. del Carlo, M. Young, S. T. Saas y A. Ballestrero, entre otros de renombre internacional.

La intensidad de la colaboración de los autores se resume en las relaciones mantenidas con países importantes en cuanto a innovación científica y tecnológica, con especial incidencia en los casos de Escocia, Estados Unidos, Japón, España, Canadá, Francia, Yugoslavia, Inglaterra, Chile, Polonia, Italia, Portugal, Alemania, Serbia y Montenegro, Argentina, Dinamarca, Suiza, Bulgaria, Hungría, Holanda,

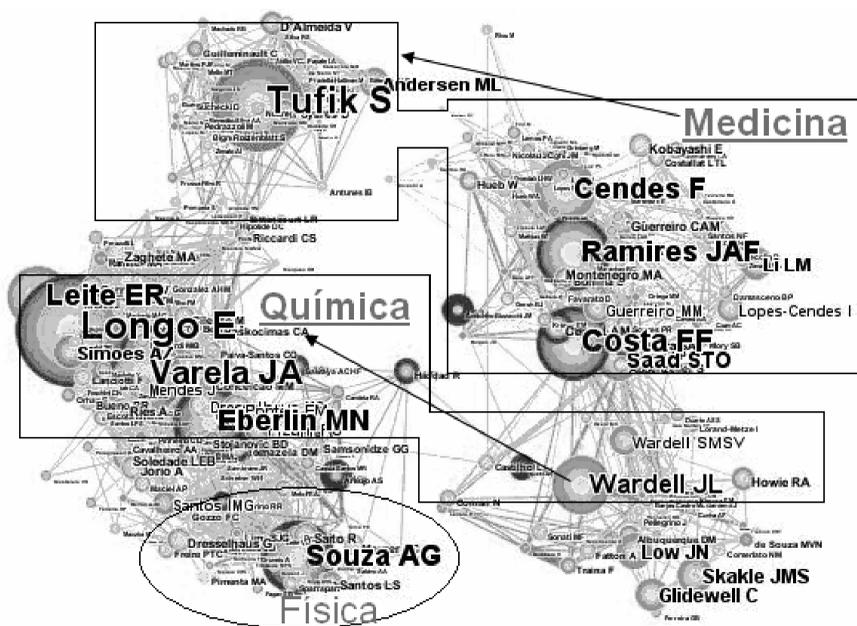


Figura 1. Red de colaboraciones científicas del grupo de autores estudiado.

Taiwán, Uruguay, Bélgica, Colombia, Finlandia, Austria, Chipre, Grecia, India, Irlanda del Norte, México, China, Rusia, Corea del Sur, República Checa, Israel, Noruega, Rumania, Eslovaquia y Suecia.

En lo tocante a la investigación en Google Académico no esperábamos visualizar un número muy elevado de productividad y de relaciones científicas, pero tratándose de investigadores de alto nivel todo resulta posible, incluso alcanzar un número de apariciones superior a los 9000 registros por parte de los diez autores.

El orden de los autores es lo menos importante, pero el análisis comparativo resulta fundamental y demuestra que recurrir a Google Académico como evaluador de la ciencia puede ser muy interesante para futuros trabajos.

Los resultados obtenidos en ambos repertorios respecto a la productividad son algo desiguales, pero hay que tener en cuenta que dentro de WoS solo representan datos de los artículos científicos, mientras que en Google Académico aparecen los artículos científicos junto a otros documentos monográficos, que pueden suponer el 40 o el 50% del total.

Otro dato interesante es que casi todas las publicaciones que están indizadas en WoS lo están también en Google Académico, pero en la mayoría de los casos

solo mediante la referencia bibliográfica, lo que tal vez se deba a que la gran mayoría de las revistas que aparecen en la plataforma ISI no son de libre acceso.

En ambos recursos la colaboración científica resultó ser muy elevada, principalmente por motivos institucionales, ya que los autores publican con otros de su misma universidad o de otros centros de investigación con los que mantienen convenios. La más representativa, según su productividad, sería la Universidad Estadual de Campinas, seguida por las universidades de São Paulo, Federal de São Paulo, Estadual Paulista, Federal de Río de Janeiro, Federal de São Carlos, Federal de Paraíba, Federal de Ceará, Federal de Río Grande del Sur, Federal de Minas Gerais, Federal Fluminense y Federal de Campinas Grande, con los centros externos MIT, Universidad de Aberdeen, Estatal de Andrews, Universidad de Belgrado, Instituto Fio Cruz, Universidad de Stanford, Universidad de Tokio, Instituto Heart, Universidad de Purdue, Universidad Jaume I, Universidad de Carolina del Sur, Universidad de California en Berkeley, Instituto Tecnológico de Georgia, CSIC español, Universidad de Harvard.

Como todos los autores pertenecen a áreas científicas de mucha relevancia social y científica en Brasil (medicina, química y física) es importante destacar que estas áreas incentivan la colaboración con las instituciones brasileñas debido a la gran creatividad que tiene este país a la hora de adecuar las investigaciones punteras a la realidad nacional y a la aplicabilidad en comparación con otros países, para una mejor identificación de los problemas tanto en el ámbito local como en el global.

La comparación entre los dos recursos demuestra que la visibilidad es mayor en Google Académico porque contiene un gran número de archivos en *open access*,

Autor	ISI			Google Académico		
	Cita	NDC	Media	Cita	NDC	Media
Souza Filho, Antônio Gomes	1267	107	11,84	2 728	79	34,53
Longo, Elson	1008	188	5,36	14 962	432	34,63
Eberlin, Marcos Nogueira	912	117	7,79	1 759	205	8,58
Leite, Edson Roberto	793	119	6,66	2 142	222	9,64
Tufik, Sérgio	614	96	6,39	2 537	339	7,48
Varela, José Arana	600	128	4,68	5 226	286	18,27
Ramires, José A. Franchini	378	51	7,41	5 474	291	18,81
Cendes, Fernando	368	72	5,11	14 078	559	25,18
Wardell, James Lewis	300	110	2,72	3 844	158	24,32
Costa, Fernando Ferreira	210	53	3,96	13 114	450	29,14

Tabla II. Número de citas de los autores estudiados (NDC = número de documentos citados; cita = total de citas recibidas; media = media aritmética del total de citas recibidas por número de documentos citados).

y debido también a la elevada cantidad de documentos escritos en el idioma materno de los autores estudiados, en este caso el portugués, mientras que en WoS la gran mayoría de los documentos se presentan en inglés. Ese factor es lógico, pues muchos autores tienen la necesidad de publicar en su lengua vernácula, dado que gran parte de su público es nacional.

También se analizaron las citas de los autores brasileños aportando la lista de los autores y los documentos publicados por ellos. El número de citas varió considerablemente entre ambos recursos.

El índice total de citas permite establecer un paralelismo entre las dos fuentes de información porque la suma final en Google Académico (GA) siempre mostrará un valor más elevado en relación con los paquetes ISI, debido al gran volumen de información recuperado.

Respecto al umbral medio de citas, y con el total de documentos indizados, puede establecerse un factor de media que se obtiene de dividir un dato por el otro y por los años estudiados. Estas serían, para cada autor, las medias de citas por documento publicado al año:

Elson Longo: 0,59 en WoS y 0,92 en Google Académico.

Sérgio Tufik: 0,38 en WoS y 0,57 en Google Académico.

José Arana Varela: 0,46 en WoS y 0,90 en Google Académico.

James Lewis Wardell: 0,26 en WoS y 1,12 en Google Académico.

Fernando Ferreira Costa: 0,20 en WoS y 2,55 en Google Académico.

José Antônio Franchini Ramires: 0,36 en WoS y 1,65 en Google Académico.

Fernando Cendes: 0,36 en WoS y 1,80 en Google Académico.

Edson Roberto Leite: 0,81 en WoS y 0,62 en Google Académico.

Antônio Gomes Souza Filho: 1,31 en WoS y 1,41 en Google Académico.

Marcos Nogueira Eberlin: 0,98 en WoS y 0,64 en Google Académico.

La media de cada uno de los autores fue muy importante para destacar su visibilidad dentro de Google Académico, donde 6 de los 10 autores resultaron tener, con diferencia, una mayor representación. Sin embargo, aparecen algunos con un índice más elevado en WoS (Leite) o con índices muy similares en ambos recursos (Souza Filho, Eberlin y Tufik).

Otra perspectiva interesante es el número de documentos citados en ambas fuentes. El total de citas recibidas por cada autor se dividió por el número de documentos citados. En este análisis lograron los puestos más destacados Eberlin, Leite y Tufik, que consiguieron mantener una media de citación equivalente en ambas fuentes, mientras que los demás autores obtuvieron índices muy distintos en ambos repertorios, por lo general más elevados en Google Académico que en WoS.

En relación con las colaboraciones que aparecen en las cocitaciones, han de destacarse las recibidas por el grupo compuesto por Longo, Leite, Varela, Souza Filho y sus colaboradores, que obtuvieron la mayor relación entre los autores es-

tudiados, reflejando de forma muy ajustada la red de productividad. Las demás relaciones se dieron en un ambiente de cocitación con autores no relacionados.

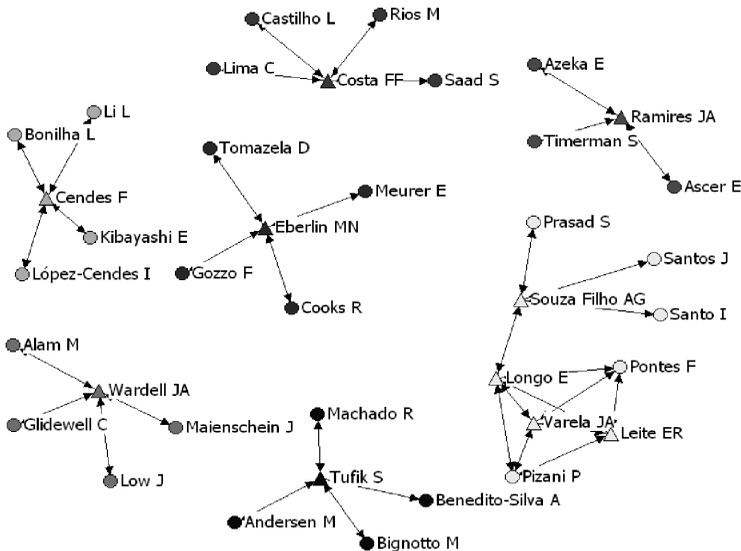


Figura 2. Red social de cocitación del Top 10 Brasil.

Las relaciones más importantes se representaron mediante triángulos para poder identificar la centralidad de cada subclúster (figura 2). Para esto no se contó con todos los autores que colaboraron con los estudiados, sino únicamente con los que presentaban mayor índice de cocitaciones.

Si se establece una comparación general entre ambos repertorios puede afirmarse que Google Académico está todavía en pruebas, y por lo tanto no puede decirse que sea del todo fiable, principalmente por no disponer de un índice de *citation report*, como sucede en los productos ISI. Este recurso resulta fundamental a la hora de poder visualizar el grado de citación.

Otro indicador interesante es el índice *h*, que permite establecer el número de documentos que han recibido al menos ese número de citas para un autor o un colectivo dado. Este factor se calcula basándose en la distribución de las citas que han recibido los trabajos científicos de un investigador (Hirsch, 2005): $N_c, \text{tot} = ah^2$.

Por ejemplo, buscamos en una base de datos o en un motor web todos los artículos de “Barquín BA” y los ordenamos por número de citas recibidas. A conti-

nuación listamos los primeros resultados (que son los que nos interesan), tal como se indica bajo estas líneas (documentos junto al índice de citas recibidas).

- 1. 27
- 2. 26
- 3. 25
- 4. 12
- 5. 8
- 6. 7
- 7. 5...

En nuestro ejemplo, el índice *h* de este autor sería de 6 apariciones, porque, una vez ordenados sus artículos por el número de citas recibidas, dicho número corresponde al número de rango del último artículo cuyo número de citas es igual o mayor que dicho número de rango. Para tener un índice *h* de 7 apariciones o más debería recibir, por lo menos, el mismo número de citas que el puesto ocupado en la lista de artículos.

Este indicador bibliométrico se aplicó a los dos repertorios estudiados para comprobar si existe una equivalencia entre ambas fuentes y poder averiguar si Google Académico es una herramienta útil también para las citaciones.

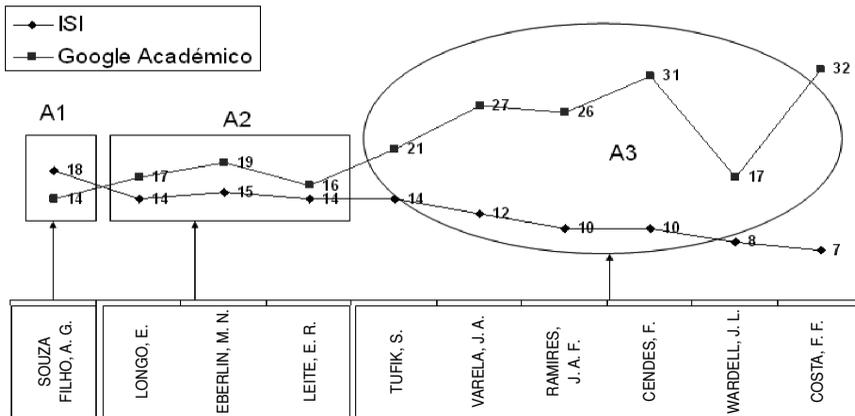


Figura 3. Índice *h* de los autores estudiados.

En la figura 3 se muestra el índice *h* con objeto de averiguar la incidencia de los autores en las dos fuentes investigadas. Se aprecian tres tramos o zonas diferentes: el primero con mayor índice en WoS (A1), el segundo con cierta igualdad entre ambos repertorios (A2) y el tercero con mayor índice en Google Académico (representado por un gran óvalo):

A1: el autor que consiguió obtener un índice mayor fue Souza Filho.

A2: aparecen como autores con igualdad de citaciones Longo, Eberlin y Leite.

La zona señalada con un óvalo corresponde a la mayor concentración de autores, siempre con un número más elevado por parte de Google Académico, mientras que para WoS se obtuvo un índice por debajo de 14 documentos.

Para la concepción del índice de documentos citados identificamos el grado de documentos presentes en ambas fuentes de información y resaltamos los que tenían más de 10 citaciones en común en ambos repertorios, de modo que se consiguieron identificar 140 documentos. Al mismo tiempo se buscaban esas citaciones (por encima de 10) en cada fuente por separado. En WoS se logró identificar 60 documentos, mientras que en Google Académico se llegó a 125, lo que supone la mayor frecuencia de ambos recursos.

Autor	WoS	GA	Ambos recursos
Longo, Elson	6	10	40
Tufik, Sérgio	7	44	13
Varela, José Arana	3	10	6
Wardell, James Lewis	6	8	4
Costa, Fernando Ferreira	9	11	11
Ramires, José A. Franchini	2	2	3
Cendes, Fernando	10	19	20
Leite, Edson Roberto	3	2	15
Souza Filho, Antônio Gomes	3	10	17
Eberlin, Marcos Nogueira	10	6	10

Tabla III. Representación de documentos citados en GA, en WoS y en ambas fuentes.

Finalmente, la comparación fue muy importante por dos motivos claves: primero porque los autores con mayor frecuencia en WoS alcanzan mayor índice también en Google Académico, fundamentalmente en el caso de los documentos citados; en segundo lugar porque un recurso puede añadir información a otro, lo que resulta especialmente interesante al ver que los datos en ISI son de pago, accesibles solo mediante suscripción, mientras que en Google Académico existe información para medir otros aspectos científicos sobre los autores investigados, como las tipologías documentales y el acceso a recursos públicos como SciELO, PubMed, MedLine y Latindex, entre otros.

4. Consideraciones finales

Las consideraciones finales tienen lógicamente que contestar a las dos preguntas que nos formulábamos para justificar nuestra propuesta de evaluar la visibilidad de algunos autores en ambos recursos: ¿Es posible establecer comparaciones entre ambos repertorios? ¿Es fiable el índice de citación en Google Académico?

Se muestra con claridad que ambos recursos son comparables por dos motivos: primero porque existen muchas instituciones de investigación que no tienen acceso a las bases de datos de ámbito mundial, en especial las de países emergentes; y en segundo lugar porque, en lo tocante a la recopilación de datos y de acuerdo con los resultados alcanzados en esta investigación, Google Académico demostró que puede servir como un recurso fiable, si bien presenta la limitación de ofrecer gran cantidad de documentos no disponibles en su integridad y obliga a hacer una depuración y refinamiento de los datos recuperados, aunque esto también ocurría con los productos de ISI en sus primeras versiones, cuando todavía estaba disponible en disquetes de 2 pulgadas.

El índice de citación de los documentos de Google Académico es fiable en parte, porque hemos realizado una depuración minuciosa y eliminado todo el solapamiento existente. Otro problema identificado fue que algunos documentos estaban desactivados de un sitio web. En este sentido también encontramos muchos enlaces de citas retiradas de Internet o con defectos de “error”.

Finalmente, consideramos que el motor de búsqueda académico puede perfeccionar su sistema, de la misma forma que se hizo en las bases de datos del ISI, para lograr una mejora considerable a través de la reducción de los solapamientos que resultan de la cuestión idiomática y de otros factores que afectan a la representación.

Referencias

- Bar-Ilan, Judit (2005). Expectations versus reality. Search engine features needed for Web research at mid 2005. // *Cybermetrics*. 9:1 (Oct. 2005). <http://www.cindoc.csic.es/cybermetrics/articles/v9i1p2.html> (2009-08-25).
- Bar-Ilan, Judit; Levene, Mark; Lin, Ayelet (2007). Some measures for comparing citation databases. // *Journal of Informetrics*. 1:1 (Jan. 2007) 26-34.
- Batagelj, Vladimir; Mrvar, Andrej (2003). Pajek analysis and visualization of large networks. // *Preprint Series*. 41:871 (2003) 2-26.
- Chen, Chaomei. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. // *Journal of the American Society for Information Science and Technology*. 57:3 (Feb. 2006) 359-377.
- Google (2005). Google Scholar support for scholarly publishers, 2005. <http://scholar.google.com/scholar/publishers.html> (2009-07-04).
- Gutiérrez Gallardo, Juan Diego; López Guisado, Ángel (2005). *Google*. A Coruña: Anaya, 2005.

- Hirsch, Jorge E. (2005). An index to quantify and individual's scientific research output. // Proceedings of the National Academy of Sciences of the United States of America 102:46 (Jan. 2005) 16569-16572.
- Jasco, Peter (2005). As we may search: comparison of major features of Web of Science, Scopus and Google Scholar citation-based and citation-enhanced databases. // Current Science. 89:9 (Nov. 2005) 1537-1547.
- Noruzi, Alireza (2005). Google Scholar: the new generation of citation indexes. // Libri. 55:4 (2005) 170-180. <http://www.librijournal.org/pdf/2005-4pp170-180.pdf> (2009-08-10).
- Price, Derek J. de Solla; López Piñero, José María (1973). *Hacia una ciencia de la ciencia*. Barcelona: Ariel, 1973.
- Price, Gary (2006). Google Scholar goes international. // SearchEngineWatch Blog, 2006. <http://blog.searchenginewatch.com/blog/060111-145555> (2009-08-04).
- Qian, Dongqiu; Gross, Mark D. (1999). Collaborative design with NetDraw. // Proceedings of Computer Aided Architectural Design (CAAD) Futures '99 (Atlanta, Georgia, 1999). Seattle: University of Washington, 1999 213-226.
- Saxton, Matthew L. (2001). ProCite 5.0. // The Library Quarterly. 71:2 (April 2001) 288-290.
- Smith, Alastair G. (2006). Google Scholar as a cybermetric tool: a comparison with the New Zealand PBRF research assessment. // 9th International Conference on Science and Technology Indicators (Leuven, Belgium, Sept 7-9, 2006). Leuven: Katholieke Universiteit, 2006. http://www.vuw.ac.nz/staff/alastair_smith/publns/GoogleScholarPBRF.pdf (2009-08-04).
- Thelwall, Mike (2002). Methodologies for crawler based web surveys. // Internet Research: Electronic Networking Applications and Policy. 12:2 (2002) 124-138.
- Thelwall, Mike (2003). What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. // Information Research. 8:3 (April 2003). <http://informationr.net/ir/8-3/paper151.html> (2009-08-09).

Recibido: 2009-09-11. Revisado: 2009-09-25. Aceptado: 2009-10-20