

## La cienciaometría como herramienta para analizar el impacto de la investigación científica en una región

Esmeralda Cervantes Rendón<sup>1</sup> y Victoriano Garza Almanza<sup>2</sup>

En los años sesentas Eugene Garfield creó el sistema de identificación de citas para determinar el impacto de las publicaciones científicas, siendo el método que ha perdurado hasta el momento, con algunas variantes como el índice H, el índice g, y el CENT (*s*cientific *curr*Ency Tokens), los cuales pueden ser analizados en diferentes bases de datos como Thomson Reuter, el Directorio de *Journals* de Acceso Abierto, Scielo y *Google Scholar*, entre otros, estos dos últimos más utilizados para la información en español, también se cuenta con diferentes programas de acceso abierto para el análisis estadístico, así como la presentación de resultados por medio de mapas.

Palabras clave: Cienciaometría; análisis de citas; *Google Scholar*; *SCImago Journal Ranking*, Scielo

La International Encyclopedia of the Social & Behavioral Sciences define a la cienciaometría como el *estudio de los aspectos cuantitativos de la comunicación científica, las prácticas de investigación y desarrollo, así como las políticas de ciencia y tecnología*. Esta disciplina está orientada a medir el avance e impacto de la ciencia en diferentes medios, para lo cual cuenta con diversas clases de herramientas, como son indicadores, programas, bases de datos y análisis estadístico.

De acuerdo a Chaviano (2004), los indicadores cienciaométricos *permiten identificar el avance de la ciencia, la caducidad de algunos campos científicos, el cambio cronológico de la producción, la*

*productividad de los autores, la colaboración entre los investigadores, el impacto de las publicaciones y el análisis y la evolución de las fuentes difusoras de los trabajos*. Por lo que logran producir una aproximación al quehacer de los investigadores, así como tener una idea concreta sobre la repercusión de sus trabajos, tanto en la comunidad científica como en la sociedad en general.

### Análisis de Citas

Una manera de identificar el impacto de una publicación es por medio de las citas (Szymanski et al, 2011). Es decir, cuantos investigadores consideran o han leído un trabajo para incorporarlo en una obra propia. Desde que Eugene Garfield

<sup>1</sup> Estudiante del Doctorado en Investigación. El Colegio de Chihuahua.

<sup>2</sup> Prof. Inv. Universidad Autónoma de Ciudad Juárez.

desarrolló este instrumento e incorporo en su base de datos, y lo dio a conocer al medio científico, en el *Journal Science* en el año de 1955 (Bensman, 2007), se ha constituido en el elemento más utilizado para determinar el impacto de un artículo científico, o de un autor, o de una institución en el quehacer de la ciencia.

Pero, ¿cómo se identifican las citas en un artículo? Para empezar, se puede recuperar la cantidad de citas que se han tenido dentro de una base de datos, siendo la más utilizada hasta el momento la de Thomson Reuter, la cual, cabe mencionar, fue desarrollada por Eugene Garfield, llamándola en ese momento “*Web of Science*”. Esta base de datos y este proceso de identificación de citas es altamente utilizado hasta el día de hoy, aun en investigaciones de instituciones de países cuyo idioma no es el inglés (Annibaldi et al, 2010; Sooryamoorthy, 2011). Sin embargo, actualmente existen otras bases de datos que también pueden ser utilizadas y que identifican las citas que se tienen dentro de ellas, como es el *Google Scholar*, que también rastrea las citas que son realizadas sobre los documentos que identifica (Repanovici, 2011).

Pero no es tan sencillo otorgar un lugar o un distintivo a un trabajo, a un

investigador o a una institución por el mero hecho de contabilizar sus citas. Existen varios factores que pueden influir en la otorgación de citas, como son: las auto-citas, citas entre colegas, la aparición de varias formas de escribir el nombre de una sola persona y, en algunos casos, las citas fantasmas (García-Pérez, 2011).

En el caso de las auto-citas, se debate que tanto peso otorgarle, y esto depende del enfoque del análisis, ya que si lo que se quiere conocer es el impacto que tiene ese trabajo en el medio científico, se le otorga un valor de cero, debido a que no es otro investigador el que esta conociendo el trabajo (Szymanski, op.cit.).

En el caso de citas frecuentes entre colegas, estas se contabilizan en un análisis de citas sencillo, mientras que otros indicadores como el *sScientific currEncy Token* (CENT), lo considera como medio punto, debido a que no es una divulgación entre todo el medio científico, sino solamente entre colegas (Szymanski op.cit.).

Por otro lado, dependiendo de la revista, puede escribir un mismo nombre de diferentes maneras, como utilizando solo las iniciales de los nombres, o el primer nombre completo y los demás con iniciales, u omitiendo algún nombre, en el caso de

nombres latinos, es mas frecuente esta confusión, ya que en muchas revistas solo colocan un apellido, o los dos, o los dos unidos por un guion, lo que puede confundir en el momento de identificar a un autor (García-Pérez op.cit.).

Mientras que las citas fantasmas, suelen darse un poco mas frecuente en los documentos que no son en ingles, y es cuando aparece que un autor cita a otro, pero no es así, indica García-Pérez.

Con el fin de mejorar la identificación de la calidad y el impacto de la investigación científica se han desarrollado diferentes indicadores a partir del análisis de las citas.

### **Indicadores derivados del análisis de citas**

Entre los indicadores mas destacados se encuentran el Índice H, que fue desarrollado en el 2005 por Hirsch. Está hecho básicamente para identificar el impacto de las publicaciones de un autor, (Arancebia y Carbajal, 2008). Consiste en un análisis cualitativo del desarrollo de un científico a lo largo de su carrera como investigador (Glanze y Persson, 2005), así como el estatus alcanzado a nivel internacional en las publicaciones (Braun et al, 2005). La forma de calcular el Índice H es ordenar de

una manera descendente cada uno de los trabajos de un autor, de acuerdo a las citas recibidas, lo que otorga un numero por el orden recibido y a este número se le denomina el Índice H (Hirsch, 2005).

Algunas de las ventajas del Índice H, incluyen (Rousseau, 2008):

- Es matemáticamente sencillo
- Es un mejor índice que el número total de publicaciones
- Incentiva a desarrollar un trabajo de mayor calidad
- Puede ser aplicado a cualquier nivel de agregación
- Combina dos tipos de actividades, que son el impacto de las citas y las publicaciones

Mientras que en el índice g, se determina de un conjunto de artículos clasificados en un orden decreciente de acuerdo a las citas obtenidas (Harzing and van der Wal, 2008).

Por otro lado, el CENT (sCientific currEncy Tokens), es utilizado para otorgar valor al impacto de las citas, con el fin de obtener el mejor dato de divulgación que existe sobre un documento, y por ello a la auto-cita no se le otorga valor, al ser la misma persona, considera que no tiene una difusión mayor, lo mismo para las citas constantes entre dos colegas, ya que no sale

del mismo círculo el documento y le da un valor menor (Szymanski op.cit).

### **Bases de datos utilizadas para el análisis del avance de la ciencia**

La principal base de datos utilizada es la Thomson Reuter, en donde solo se presentan *journals* considerados de alto impacto y mayormente en inglés. Esta base de datos fue creada por Eugene Garfield (Bensman, 2008), sin embargo si se desea realizar un análisis de publicaciones en otro idioma, no es posible utilizar esta base de datos. También se precisa tener una suscripción institucional para que la base pueda ser utilizada (Harzing and van der Wal op.cit.).

Otra base de datos que ha sido utilizada es la Scopus, basada en las publicaciones de Elsevier, que también cuenta con un gran número de publicaciones, mayormente en inglés, y es de alto impacto, aunque cuenta con 21% de su documentación total en otro idioma diferente al inglés (Gagolewski, 2011).

También existen bases de datos para analizar información de acceso abierto, como es el Directorio de *Journals* de Acceso Abierto (DOAJ por sus siglas en inglés) (<http://www.doaj.org/>). En este sitio

se pueden encontrar también documentos en español.

Existen repositorios de documentos que pueden ser utilizados para realizar este tipo de análisis como es el del proyecto SHERPA/RoMEO

(<http://www.sherpa.ac.uk/romeo/>) el cual es un repositorio de documentos archivados por los autores (Miguel et al 2011).

Por otro lado, también se tiene la base de datos Scielo (*Scientific Electronic Library Online*) (<http://www.scielo.org/>), que contiene información de acceso abierto en diferentes idiomas (Packer, 2010). La metodología que utilizan en la recuperación de documentos de acceso abierto, les permite obtener datos estadísticos y bibliométricos del impacto de sus publicaciones.

Se debe tener presente que para hacer un análisis cuantitativo, no es necesario que sea sobre una de estas bases de datos, puede ser también sobre otras, o incluso sobre documentos locales, como bases de datos locales o documentos de alguna convención, como de la Academia Europea de Alergia e Inmunología Clínica, en donde se analizaron las contribuciones realizadas en sus tres reuniones que llevaron a cabo en el periodo del 2001 al 2007 (Scutaru et al 2010).

También puede ser utilizado *Google Scholar* para realizar análisis de la información que se encuentra presente en este buscador, en el que se incluyen artículos, libros, presentaciones, así como documentos en diferentes idiomas distintos al inglés, como el español (Prat, 2009), (Harzing and van der Wal op.cit.), (Repanovici, 2011).

### **Otros parámetros que pueden ser identificados en un análisis cuantitativo**

Para un análisis cuantitativo, no solamente se puede medir las citas, sino también otros elementos dependiendo de lo que se quiera obtener, se pueden hacer análisis de palabras claves, de autores, de regiones, palabras en el título o revistas (Leydesdorff, 2010).

Un punto también importante para evaluar, es la identificación de redes de trabajo, para ello existen varios métodos, como son el análisis de la red social de colaboración, la identificación de las coautorías, así como métodos cualitativos como la observación, la entrevista y la encuesta (Abbasi et al, en prensa).

Por otro lado, el investigador no se encuentra solo con sus colegas, realizando proyectos, es importante tener el enfoque

socioeconómico, y por ello nos menciona Heimeriks (2012) que para la identificación de nuevos regímenes de la investigación es necesario identificar la interacción que existe entre el ambiente particular del investigador, la aparición de nuevos campos de estudio, así como el ambiente socioeconómico (Heimeriks and Leydesdorff, 2011)

Otro análisis es a través del SCImago Journal & Country Rank (SJR) (<http://www.scimagojr.com/>) que utiliza el algoritmo de Google PageRank se pueden realizar análisis de citas, del índice H y de la cantidad de publicaciones por países y/o temáticas. Una de sus ventajas es que elimina la autocitación en los datos que proporciona (Siebelt et al, 2010) y (Falagas et al, 2008).

SCImago emite un reporte anual de Instituciones, en donde en el del 2010, se pueden encontrar indicadores de 1) producción de la publicación, 2) porcentaje de producción realizado con instituciones extranjeras, 3) el radio entre el promedio del impacto de una institución comparado con el promedio mundial del impacto en la misma área y 4) la publicación de artículos en revistas consideradas de alto impacto (Bornmann et al preprint).

## Software disponible para el análisis cuantitativo

Los programas que se han desarrollado son para el análisis estadístico de los resultados, algunos que existen de arquitectura abierta son el *Publish or Perish* (Harzing and van der Wal op.cit.), y CITAN por sus siglas en inglés (Análisis de Citas) (Gagolewski, 2011).

El software *Publish or Perish*, fue diseñado para poder realizar análisis de citas, así como de otro tipo, sobre la documentación de *Google Scholar*, también se puede obtener el índice H, el índice g, así como el número de citas por documento (Harzing, 2007).

Tanto Google Scholar como el software *Publish or Perish* pueden ser utilizados para estudios regionales, y con el idioma español, así como con información de acceso abierto (Chirroque-Solano y Padilla-Santoyo, 2009), (Efron y Brennan, 2011).

Mientras que CITAN realiza análisis estadístico R (<http://www.e-project.org/about.html>) utilizando la base de datos SciVerse Scopus, así como obtiene los índices h y g, entre las herramientas estadísticas que maneja se encuentran los modelos lineales y no lineales, series de tiempo, clasificación, *clusterin* y análisis de

microflechas (Gagolewski op.cit.) y se encuentra disponible a través del repositorio de CRAN (*The Comprehensive R Archive Network*).

## Generación de mapas del conocimiento

Independientemente del análisis que se realice, se obtienen bastantes resultados, los cuales pueden ser presentados de diferentes formas, la mayoría en tablas y gráficas, pero también existe una forma de presentar estos resultados, los cuales son por medio de mapas geográficos, así como mapas del conocimiento, para ello, existen software que han sido diseñados para cumplir con esta función.

Algunos de los programas encontrados son:

VOSviewer ([www.vosviewer.com](http://www.vosviewer.com)), es diseñado para crear mapas con información bibliométrica, en donde puede analizar términos, documentos, región, autores, coautores, entre otras cosas, presenta la información, otorgándole peso por medio del radio de cada término, y diferenciándolos por colores, además es de libre acceso (Waaiker et al 2011).

Otro software de libre acceso utilizado es el Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>), en donde se pueden representar relaciones entre autores,

temas o palabras clave (Leydesdorff, 2010b)

Para generar mapas sobre las redes de trabajo sociales se puede utilizar el Visone (<http://visone.info>) (Leydesdorff op.cit).

Mientras que para obtener mapas de excelencia, utilizando el Google Maps y otras herramientas, el equipo de Bornmann y Leydesdorff desarrollaron varios programas que pueden ser obtenidos desde el sitio oficial de Leydesdorff ([http://www.leydesdorff.net/mapping\\_excelence/index.htm](http://www.leydesdorff.net/mapping_excelence/index.htm)) (Bornmann et al, 2011)

### **Conclusiones**

Aun cuando la información que es considerada de alto impacto es publicada la mayoría en inglés, también se genera información en la lengua nativa del investigador, como es el español, por lo que un investigador, institución o región pueden generar publicaciones tanto en inglés como en español, y dependiendo de lo que se desee medir, será la metodología a utilizar. Por ejemplo si se requiere identificar el impacto internacional en revistas de alto impacto, el análisis se deberá de realizar sobre bases de datos que contengan estas revistas, como Thomson Reuters™, incluso

cuando el idioma natal del autor, institución o región no sea el inglés.

Sin embargo, existen otras bases de datos en donde se tienen artículos en idioma español, como el Directorio de *Journals de Acceso Abierto*, así como la base de datos de Scielo (*Scientific Electronic Library Online*), en donde se pueden realizar diferentes análisis cuantitativos en este idioma.

Además también puede ser utilizado *Google Scholar* con el software *Publish and Perish* para este tipo de análisis, que como se mencionó es de acceso libre. Por lo que actualmente es un poco mas viable hacer estudios del avance de la ciencia en idioma español incluso con programas de acceso libre, tanto para el análisis estadístico como para la presentación de resultados por medio de mapas.

Por lo que se puede concluir que dependiendo del enfoque de la investigación y del objetivo del análisis cuantitativo que se quiera realizar, dependerá las herramientas que se utilicen, así como las bases de datos y/o publicaciones consultadas.

## Referencias

- Abbasi A, Hossain L and Leydesdorff L, In Press. Betweenness Centrality as a Driver of Preferential Attachment in the Evolution of Research Collaboration Networks, *Journal of Informetrics*. Available at: <http://arxiv.org/abs/1111.6804>
- Annibaldi A, Truzzi C, Illunimati S and Scarponi G, 2011. Scientometric analysis of national university research performance in analytical chemistry on the basis of academic publications: Italy as case study. *Anal Bioanal Chem*, vol. 398, pp 17-26
- Arencibia JR, Carvajal ER 2008, Los índices H, G y R: su uso para identifica autores líderes en el área de la comunicación durante el periodo 2001-2006. *Acimed*, vol. 17, no. 4, pp 9
- Bensman, SJ 2007, Garfield and the impact factor, *Annual Review of Information Science and Technology*, vol. 41, pp 93-155
- Bensman, SJ 2008, Garfield and the Impact Factor: The creation, utilization, and validation of a citation measure. Part 2. The Probabilistic, Statistical, and Sociological Bases of the Measure, *Annual Review of Information Science and Technology*, pp 109
- Bornmann, L, Leydesdorff L, Walch-Solimena C and Ettl C, 2011. Mapping excellence in the geography of science: An approach based on Scopus data. *Journal of Informetrics*, vol.5, no. 4, p.537-546. Available at: <http://arxiv.org/abs/1105.4736>.
- Bornmann L, Moya-Anegón F and Leydesdorff L, Preprint. The new Excellence Indicator in the World Report of the SCImago Institutions Ranking 2011, *Journal of Informetrics*. Available at: <http://arxiv.org/abs/1110.2305>
- Braun, T, Glanzel W y Schubert A 2005, A Hirsch-type index for journals, *Scientist*, vol. 19, no. 22, pp 8
- Chaviano, G. 2004, Algunas consideraciones teórico-conceptuales sobre las disciplinas métricas, *Acimed*, vol. 12, no. 5, revisado 5 de septiembre 2011, <<http://eprints.rclis.org/handle/10760/5904#referencias>>
- Chiroque-Solano, R y Padilla-Santoyo, P 2009, Análisis de coautoría en la revista *Biblios*: Una aproximación desde Google Scholar, *Biblios*, vol. 33-34, revisado 15 de noviembre 2011, <<http://eprints.rclis.org/handle/10760/3859#.TxBY6Xj5ic>>
- Efron, N and Brenna, NA 2011, Citation analysis of Australia-trained optometrists', *Clinical and Experimental Optometry*, vol. 94, pp. 600-605
- Falagas, ME, Kouranos, VD, Arencibia, R and Karageorgopoulos, E 2008, 'Comparison of SCImago journal rank indicator with journal impact factor', *The FASEB Journal*, vol. 22, pp. 2623-2628
- Gagolewski, M., 2011. Bibliometric impact assessment with R and the CITAN package. *Journal of Informetrics*, vol.5, no.4, p.678-692. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1751157711000708>.
- García-Pérez, MA 2011. Strange attractors in the Web of Science database. *Journal of Informetrics*, vol 5, pp 214-218
- Glanzel, W y Persson, O 2005, 'H-index for Prize medalist', *ISSI Newsletter*, no. 1, vol. 4, pp. 15-23
- Harzing, AW 2007, *Publish or Perish*, revisado 27 de noviembre 2011, <<http://www.harzing.com/pop.htm>>
- Harzing, A. & Van Der Wal, R., 2008. Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*, vol.8, no.1, p.61-73. Available at: <http://www.int-res.com/abstracts/ese/v8/n1/p61-73/>.
- Heimeriks, G. & Leydesdorff, L., 2011. Emerging Search Regimes: Measuring Co-evolutions among Research, Science, and Society. *Innovation*, vol.24, no.2, p.1-31. Available at: <http://arxiv.org/abs/1101.2591>.
- Hirsch JE, 2005. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA*, vol. 102, pp. 16569-16572
- Leydesdorff, L., 2010. Eugene Garfield and Algorithmic Historiography: Co-Words, Co-Authors, and Journal Names. *Annals of Library and Information Studies*, p.1-28. Available at: <http://arxiv.org/abs/1005.5444>.
- Leydesdorff, L., 2010. What Can Heterogeneity Add to the Scientometric Map? Steps towards algorithmic historiography M. Akrich et al., eds. *Social Science Information*. Available at: <http://arxiv.org/abs/1002.0532>.
- Miguel, S., Chinchilla-Rodríguez, Z. & De Moya-Anegón, F., 2011. Open access and Scopus: A new approach to scientific visibility from the standpoint of access. *Journal of the American Society for Information Science and Technology*, vol. 62, p.1130-1145. Available at: <http://doi.wiley.com/10.1002/asi.21532>.
- Packer, A.L., 2010. The SciELO Open Access: A Gold Way from the South. *Canadian Journal of Higher Education*, vol.39, no.3, p.111-126. Available at: <http://ojs.library.ubc.ca/index.php/cjhe/article/view/479>.
- Prat, AM 2009, *Módulo de capacitación para la recolección y análisis de indicadores de*

*productos de las actividades de ciencia y tecnología*, Banco Interamericano de Desarrollo

Repanovici, A., 2011. Measuring the visibility of the university's scientific production through scientometric methods: An exploratory study at the Transilvania University of Brasov, Romania. *Performance Measurement and Metrics*, vol.12, no.2, p.106-117.

Rousseau, R 2008, 'Reflections on recent developments of the h-index and h-type indices', *Collnet Journal of Scientometrics and Information management*, vol. 2, no. 1, pp 8

Scutaru C, Quarcoo D, Sakr M, Shami A, Al-Mutawakel K, Vitzmhut K, Fischer TC, Zuberbier T and Groneberg-Kloft B 2010. *Journal of Occupational Medicine and Toxicology*, vol. 5, no.2, pp 1-8. Available at <http://www.occup-med.com/content/5/1/2>

Siebelt, M, Siebelt, T. Pilot, P, Bloem, RM, Bhandari, M and Poolman, RW 2010, 'Citation

analysis of orthopaedic literature; 18 major orthopaedic journals compared for Impact Factor and SCImago', *BMC Musculoskeletal Disorders*, vol. 11, no. 4, revisado 18 de noviembre 2011, <<http://www.biomedcentral.com/1471-2474/11/4>>

Sooryamoorthy, R 2011, 'Scientific publications of engineers in South Africa, 1975-2005', *Scientometrics*, vol. 86, pp 211-226

Szymanski BK, Lluís de la Rosa, J and Krishnamoorthy M 2012. An Internet Measurement of the Value of Citations. *Information Science*, vol. 185, no. 1, pp 18-31

Waijjer, C.J.F., Van Bochove, C.A. & Van Eck, N.J., 2011. On the map: Nature and Science editorials. *Scientometrics*, vol.86, no.1, p.99-112. Available at: <http://www.springerlink.com/index/10.1007/s11192-010-0205-9>.