

Enrique Orduña-Malea · Alberto Martín-Martín
Juan M. Ayllón · Emilio Delgado López-Cózar

Prologado por
Peter Jacsó
Anne-Wil Harzing

La revolución Google Scholar

Destapando la caja de Pandora académica

Granada · 2016

© LOS AUTORES

© UNIVERSIDAD DE GRANADA

La revolución Google Scholar.
Destapando la caja de Pandora académica

ISBN: 978-84-338-5941-9

Depósito legal: GR./ 837-2016

Edita: Editorial Universidad de Granada

UNE. Unión de Editoriales Universitarias Españolas

Diseño de la edición: Francisco Vega Álvarez

Printed in Spain

Impreso en España

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra sólo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley.

Contenidos

PREFACIO I

La revolución de Google Scholar: destapando la caja de pandora académica.....	9
---	---

PREFACIO II

Sacrificar un poco de precisión por una cobertura (mucho) más amplia	15
--	----

INTRODUCCIÓN

21

BLOQUE A

Google Scholar: el buscador académico universal	25
---	----

Capítulo 1 Los orígenes	27
----------------------------------	----

Capítulo 2 Las dos caras de <i>Google Scholar</i> (I): El buscador académico	41
---	----

Capítulo 3 Las dos caras de <i>Google Scholar</i> (II): La herramienta de evaluación	75
---	----

Capítulo 4 Capturando la Web académica: Funcionamiento general	99
---	----

Capítulo 5 El retrato de <i>Google Scholar</i> : Cobertura, tamaño y evolución	119
---	-----

Capítulo 6	
El proceso de búsqueda en <i>Google Scholar</i> :	
Prestaciones y servicios	133
Capítulo 7	
<i>Google Scholar</i> en la balanza: Fortalezas, debilidades y peligros.....	155

BLOQUE B

La familia Google Scholar: Productos derivados oficiales e independientes	171
---	-----

Capítulo 8	
Google Scholar Metrics: El impacto de las revistas científicas	173

Capítulo 9	
Google Scholar Citations: Los perfiles académicos en la Web.....	185

Capítulo 10	
Productos derivados independientes	217

CONCLUSIONES

<i>Google Scholar</i> : la caja de pandora académica.....	249
---	-----

BIBLIOGRAFÍA	253
--------------	-----

La revolución de Google Scholar: destapando la caja de pandora académica

La caja de Pandora fue un regalo de Zeus que no debía ser abierto. Los desarrolladores de *Google Scholar* (GS) pusieron muchas cosas en su caja de almacenamiento, algunas de ellas buenas, algunas otras malas. La mayoría de los usuarios no conocen lo que alberga la caja de GS y no se preocupan por aprenderlo, siempre y cuando encuentren algunos buenos artículos de revistas, libros, capítulos de libro, actas de congresos, tesis doctorales y patentes, más cuando muchos de éstos se encuentran fácilmente disponibles tanto en fuentes de acceso abierto como de pago, en las estanterías digitales de sus bibliotecas.

A los bibliotecarios y otros profesionales de la información les gustaría conocer el tamaño real aproximado de GS, el número de ítems (registros), revistas científicas, magazines, artículos, libros... y en qué medida están cubiertos, por cuántos años, en qué idiomas.

Los cuatro autores, profesores en diversas universidades en España, escribieron un libro excelente sobre la base de sus investigaciones, realizaron una extensiva e intensiva serie de pruebas y análisis de resultados. El hecho de estar familiarizado con sus artículos de revistas y conferencias, tanto en inglés como en español, hizo fácil que aceptara la invitación para escribir un prólogo a este libro, que explica las ventajas y los inconvenientes de GS y demuestra la competencia de los autores y su experiencia (incluso cuando no estamos de acuerdo en algunos temas). Se aprecia que para algunos de los asuntos controvertidos los autores citen artículos que critican y/o reconocen las “soluciones” de GS.

El único punto en el que discrepo profundamente con los autores es cuando éstos empuerñecen las consecuencias “letales” de un ejemplo específico: cuando GS atribuyó a I INTRODUCTION la autoría de 40.100 documentos y, dos años después, a 700.000 ítems aproximadamente, privando a los autores reales tanto de sus autorías como de sus citas.

En efecto, GS redujo esta autoría a cinco registros a fecha de septiembre de 2015, ante la cual indican los autores que “la mejora del sistema fue simplemente espectacular”. Esto es como alabar al abusador de una mujer de que éste ha reducido considerablemente los golpes a su esposa e hijos.

Jacso emphasizes shortcomings of Google Scholar in several articles; he shows inconsistencies in Google Scholar. For example [JACSO, 2006] one of the most prolific authors according to Google Scholar is “I Introduction” with 40,100 reported documents authored by him/her. We rerun the test on November 4, 2006 – the number of items authored by “I Introduction” (the number of results for the query *author: "I Introduction"*) increased to 689,000 (!). Another example, more related to informetrics

Figura A1. Extracto del famoso ejemplo de autor I INTRODUCTION
Source: Bar-Ilan, J. (2008)

Incluso para un autor muy solitario, la ventaja de agruparse en un cuarteto para escribir este libro es obvia, sobre todo para las secciones acerca de la génesis de GS, enriquecida por cientos de referencias a materiales primarios de investigación publicados acerca de GS durante los últimos doce años. Los trabajos citados, entre ellos muchos de los artículos de los autores así como artículos de conferencias, informes técnicos y comentarios acerca del bueno, el feo y el malo de GS, son muy valiosos para comprender sus ventajas e inconvenientes.

Lo “bueno” es la amplitud de los documentos primarios en términos de tipologías documentales, idiomas y la habilidad para buscar el texto completo de documentos primarios. El malo es que los *parsers* (analizadores) de GS, todavía de bajo nivel, siguen produciendo miles de autores fantasma creados a partir de las palabras de los títulos de los documentos, cabeceras de secciones y etiquetas de las opciones de los menús de búsqueda (FigA2). El autor, supuestamente llamado IIB STATISTICS según GS, es simplemente el título de la sección II. Background Statistics.

Resulta extraño por qué el título de la segunda sección fue el elegido por GS para generar el nombre falso de autor y no los títulos de la primera o la tercera sección. Los tres autores auténticos se ven privados de su estatus de autor y de las cinco citas recibidas, otorgadas por GS al autor fantasma. Esto no es un síntoma raro o casual, sino muy común.

Un síndrome relativamente nuevo de GS es que éste sustituye el resumen/extracto original de los artículos de revista o actas de congresos, tal y como se muestra en la siguiente captura de pantalla (Figura A3). En el caso de que no exista resumen, GS añade contenido absurdo creado a través de un *scraping* de la plantilla de la pantalla de búsqueda, añadiendo esa cadena sin sentido de caracteres.

[\[PDF\] Vascular imaging](#)

IIB STATISTICS - IEEE Transactions on Medical Imaging, 2005 - edge.cs.drexel.edu
 ANGIOGRAPHIC imaging is commonly used world-wide for the diagnosis of cardiovascular, cerebrovascular, and peripheral vascular disease. As modern image equipment has developed, angiographic imaging has also come to play a major role in the guidance of ...
 Cited by 5 Related articles All 6 versions Cite Save More

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 24, NO. 4, APRIL 2005

433

Vascular Imaging

I. MOTIVATION FOR THE SPECIAL ISSUE

ANGIOGRAPHIC imaging is commonly used world-wide for the diagnosis of cardiovascular, cerebrovascular, and peripheral vascular disease. As modern image equipment has developed, angiographic imaging has also come to play a major role in the guidance of interventional procedures. For instance, since 1979 and in the US only, the number of patient catheterizations has increased by 479% to a total of about 1.46 million interventions in 2002 [1].

This special issue is devoted to this exciting field of vascular imaging at the three levels of image acquisition, computational imaging, and computational simulation.

II. BACKGROUND STATISTICS

The first call-for-papers of the special issue on vascular imaging was announced in the IEEE TRANSACTIONS ON MEDICAL IMAGING issue of October 2003. It was also ad-

Figura A2. Ejemplo de autor fantasma (IIB STATISTICS) extraído de un título de sección

Skip-to Main Content Wiley Online Library Log in Register Address

About 552,000 results (0.05 sec)

[BOOK] Atoms in molecules
 RFW Bader - 1990 - Wiley Online Library
 Skip to Main Content Wiley Online Library. Log in / Register. Log In E-Mail
 Address Password Forgotten Password? Remember Me. ...
 Cited by 13000 Related articles All 4 versions Cite Save More

[BOOK] Protection for the carbonyl group
 TW Greene, PGM Wuts - 1991 - Wiley Online Library
 Skip to Main Content Wiley Online Library. Log in / Register. Log In E-Mail
 Address Password Forgotten Password? Remember Me. ...
 Cited by 8810 Related articles All 10 versions Cite Save More

[BOOK] Evidence-based medicine
 DL Sackett - 2000 - Wiley Online Library
 Skip to Main Content Wiley Online Library. Log in / Register. Log In E-Mail
 Address Password Forgotten Password? Remember Me. ...
 Cited by 8776 Related articles All 2 versions Cite Save More

Figura A3. Google Scholar añade un sumario a partir de las opciones en la plantilla de búsqueda en el sitio web de Wiley Online Library

En el caso de la colección digital del sitio web de *John Wiley*, GS se las arregla para producir 552.000 hits con el contenido que se ilustra a continuación, mostrando el número de hits en el más brillante, casi invisible, color gris, tal y como hacen los humanos cuando dicen algo en voz muy baja al ser conscientes de que lo que dicen no tiene sentido.

La prestación "fea" más característica de GS es el secretismo relativo a la información esencial, como por ejemplo el tamaño de la base de datos, el número de registros maestros, el número de citas totales, la lista de revistas cubiertas y la cobertura temporal cubierta. Es decir, datos que son fácilmente disponibles en bases de datos hospedadas por *EBS-CO*, *Proquest*, *Web of Science*, *Scopus*, *Journal Citation Reports* y *Scimago*. Como las prácticas de los médicos de vudú desde América del Norte a la India. Esto refuerza la sensación de que, a pesar del uso del adjetivo *Scholar* (o Académico en muchos países de habla hispana), son muchas las fuentes no académicas.

La cobertura del libro es excelente y completa, exceptuando la decepcionante omisión de la fértil obra de Eugene Garfield desde mediados de los 1950 en adelante acerca de la indización de citas y la búsqueda basada en referencias citadas. Su clásico trabajo publicado en la revista *Science* hace 60 años proporcionó la idea nuclear para la implementación de los índices de citas de ISI, Microsoft Academic Search (que se ha ido desvaneciendo), del servicio de *Scopus* y de *Google Scholar*. La columna vertebral de estos servicios de bases de datos son las referencias citadas y citantes (creadas manualmente por Garfield en tarjetas Hollerith, predecesoras de los enlaces digitales).

Las cerca de 270 páginas del libro se componen de dos partes principales (bloques). La primera parte trata de la génesis de GS y sus características como motor de búsqueda (una herramienta de descubrimiento de recursos) así como de su uso como una herramienta de evaluación que pretende cuantificar la productividad científica y su impacto en diferentes niveles, que van desde los autores individuales e investigadores hasta los artículos, actas de congresos, capítulos de libro y libros.

Las siete secciones dentro de este primer bloque dirigen a los lectores a través del proceso de cómo GS captura los datos, cuál es el tamaño y cobertura estimados, así como de los elementos del proceso de búsqueda a través del prisma crítico de los especialistas en búsquedas en bases de datos.

El segundo bloque discute las ventajas e inconvenientes de *Google Scholar Metrics for publications* y de *Google Scholar Citations*, así como de

algunas utilidades que ayudan en los procesos de input y output, incluyendo el excelente y gratuito software *Publish or Perish*.

El competente, objetivo y muy actualizado contenido del libro es muy necesario en tanto que las medidas basadas en métricas del rendimiento e impacto de la investigación son más y más ampliamente utilizadas por las relativamente objetivas evaluaciones del rendimiento a nivel individual de los investigadores y revistas. La evaluación de la investigación a nivel institucional requerirá a *Google Scholar* la disposición de información relativa a la afiliación de los autores en un formato y contenido consistentes, lo cual no es precisamente el punto fuerte de GS. La navegación por todos los elementos de los metadatos será esencial, pero esto podría revelar demasiado acerca de las limitaciones de GS.

El libro contiene ilustraciones muy informativas, aunque en la versión impresa utilizada para este prólogo no eran todas bien legibles. La edición digital puede ayudar con alta resolución e imágenes en color, gráficos y tablas. La inclusión de búsquedas en las que se pueda clicar a la hora de generar los perfiles de autor de GS, junto a un listado sistemático de las ventajas e inconvenientes (a partir del listado de resultados de GS de los propios autores) podría contribuir al uso de esta obra como un libro de texto digital tanto para los estudiantes como para los profesores. La incorporación de enlaces directos a las referencias citadas y la creación de consultas en las que se pueda hacer clic para poder repetir las pruebas de muestra, podrían constituir una característica excepcional para este excelente libro.

Professor Peter Jacsó

Library and Information Science Program
Department of information and Computer Sciences
Honolulu, Hawaii 2016 April 5

Sacrificar un poco de precisión por una cobertura (mucho) más amplia

Si me pidieran que resumiera mi “veredicto” sobre Google Scholar en tan solo diez palabras, la cita anterior sería un probable candidato. He estado siguiendo a Google Scholar desde su lanzamiento en 2004, aunque mi interés se incrementó en 2006, cuando necesitaba justificar el impacto de investigación en mi solicitud de promoción a Profesor Titular.

Dedicándome al área del Comercio internacional, descubrí que en la Web of Science estaban indizadas muy pocas de las revistas en las que había publicado. Aquellas revistas que sí se encontraban incluidas tenían bajos factores de impacto, debido en gran medida a que muchas de las revistas que las estaban citando no estaban (todavía) incluidas en Web of Science. A consecuencia de esto, hace diez años era muy difícil para un académico del área del Comercio internacional convencer a un tribunal de promoción de que había alcanzado tanto impacto como, por ejemplo, investigadores trabajando en disciplinas vecinas como la Economía o Psicología. En esas disciplinas, una proporción mucho mayor de revistas académicas sí estaban incluidas en la Web of Science.

Acceda a Google Scholar... y a publish or perish!

Acceda a Google Scholar! Cuando realizaba búsquedas sobre mi obra en Google Scholar, mi expediente parecía inmediatamente mucho más brillante. Resultó que mi trabajo era realmente muy citado; incluso las publicaciones no convencionales (como libros, capítulos de libro o incluso un ranking de revistas) tenían un fuerte impacto en términos de citas. Desafortunadamente, la interfaz de Google Scholar no hacía muy sencilla la tarea de agregar métricas de citas de una forma accesible, de tal forma que éstas pudieran ser comparadas entre investigadores.

Como consecuencia de ello, Publish or Perish (PoP) nació en Octubre de 2006 (<http://www.harzing.com/pop.htm>). PoP es una aplicación informática que recupera y analiza citas académicas. Utiliza Google Scholar para capturar citas brutas, analizarlas y presentar posteriormente un amplio rango de métricas derivadas. Para mí, esto representó además el comienzo de un nuevo “pasatiempo académico”, haciendo investigación en el campo de la Bibliometría. Desde entonces he publicado más de una docena de artículos en revistas como *Scientometrics*, *Journal of Informetrics* o *Journal of the Association for Information Science & Technology*. La mayoría de estos artículos han usado Google Scholar como fuente de datos.

Avances de Google Scholar durante la última década

Google Scholar ha recorrido un largo camino desde sus primeros días. Su cobertura ha mejorado de forma espectacular y en la actualidad es mejor que Web of Science o Scopus para cualquier disciplina (Harzing y Alakangas, 2016). Como consecuencia, existen actualmente muchos estudios bibliométricos que confían en Google Scholar (generalmente a través de Publish or Perish) para realizar su investigación. Una búsqueda en Google Scholar (incluyendo todas las palabras) para la expresión <Harzing “Publish or Perish”> genera cerca de 2000 resultados.

Disciplina	Citas en Scopus como % de citas en Google Scholar	Citas en Web of Science como % de citas en Google Scholar
Humanidades	11.5%	7.0%
Ciencias sociales	30.0%	22.7%
Ingenierías	57.6%	45.7%
Ciencias	64.2%	65.6%
Ciencias de la vida	70.5%	66.8%

Tabla B1. Comparando citas en Google Scholar con citas en Web of Science y Scopus

La tabla B1 resume parte de los principales resultados de un reciente estudio sobre 146 investigadores en Ciencias, Ciencias de la vida, Ingeniería, Ciencias sociales y Humanidades (Harzing y Alakangas, 2016). Tal y como es evidente a primera vista, tanto la Web of Science como Scopus pierden un gran número de citas en las Humanidades y Ciencias sociales, debido en gran parte a que estas bases de datos no incluyen libros y, en algunas disciplinas, cubren solamente una fracción de las revistas.

Sin embargo, incluso en las ingenierías, Ciencias y Ciencias de la vida, Google Scholar proporciona entre 1.5 y 2 veces más citas que Web of Science y Scopus.

Google Scholar y Publish or Perish han democratizado el análisis de citas

Soy la primera en reconocer que Google Scholar tiene importantes limitaciones, de las cuales las más importantes se encuentran resumidas en mi tutorial de Publish or Perish (Harzing, 2016; <http://www.harzing.com/poptips/index.htm>). Así mismo existe un peligro nada despreciable a que, gracias a un acceso sencillo a las herramientas bibliométricas, se propicie un cierto nivel de uso inexperto o sencillamente ignorante.

Sin embargo, Google Scholar ha jugado desde mi punto de vista un papel fundamental en la democratización del análisis de citas (Harzing y Mijhardt, 2015). Mediante Publish or Perish cualquier persona con un ordenador y acceso a Internet puede ejecutar análisis bibliométricos. No es de extrañar que el software sea usado en todo el mundo, desde investigadores y bibliotecarios particulares en más de 80 países hasta departamentos gubernamentales (por ejemplo, el *US Department of Energy*, *US Environmental Protection Agency* o el *US Agency for International Development*); desde agencias de financiación (como el *Social Sciences and Humanities Research Council - SSHRC* en Canadá o el *Centre National de la Recherche Scientifique - CNRS* en Francia) hasta laboratorios de investigación (por ejemplo, *Microsoft*, *Hewlett Packard* o *IBM*).

Es particularmente gratificante observar que el software es ampliamente utilizado en universidades altamente posicionadas, tales como *Harvard*, *Stanford*, *MIT*, *Oxford* o *Cambridge*; universidades que poseen un amplio acceso a otras alternativas comerciales. Sin embargo, es incluso más satisfactorio ver su igualmente alta utilización en universidades con escasos recursos en países como Armenia, Botsuana, Mongolia, Paraguay, Tayikistán y Uruguay. A nivel general, existen miles de bibliotecas por todo el mundo que listan el software como una alternativa gratuita a Scopus y la Web of Science. Google Scholar y Publish or Perish cubren claramente una necesidad!

Más cerca de (mi) casa, tanto el software como Google Scholar son particularmente populares en Italia, Grecia y Polonia. Países donde en muchas universidades no tienen acceso ni a Scopus ni a la Web of Science. Me parece particularmente agradable que tanto en Italia como en Grecia la aplicación haya sido regularmente utilizada para promover la transparencia y la meritocracia en nombramientos universitarios. Ade-

más, tanto en éstos como en otros países, los investigadores que publican en sus idiomas locales en lugar de en Inglés necesitan confiar en Google Scholar si desean obtener una cobertura de su trabajo más que incidental. Del mismo modo que los asiáticos del Este, también éstos podrían descubrir que la Web of Knowledge de Thomson Reuters no está particularmente bien versada en distinguir con precisión a los investigadores con nombres no anglosajones (Harzing, 2015).

¿El veredicto?

Utilizar Google Scholar significa sacrificar un cierto grado de precisión. En la mayoría de los análisis bibliométricos, especialmente en aquellos con elevados niveles de agregación o centrados en métricas robustas como el h-index, los inevitables deslices de Google Scholar no influirán significativamente en los resultados. Sin embargo, esto es poco consuelo para las (pocas) personas a las que se les ha “robado” sus publicaciones más citadas porque Google Scholar no ha analizado gramaticalmente bien su nombre. Si se requiriera de una precisión del 99-100%, Google Scholar necesitaría siempre ser triangulada con otras fuentes de datos.

En cualquier caso, como investigadora en el campo de las Ciencias sociales, estoy más que feliz por aceptar el error ocasional de Google Scholar a cambio de una cobertura que es muy superior a la de Scopus y Web of Science, y que no discrimina las prácticas de publicación fuera del campo de las Ciencias (de la Vida) (Harzing, 2013). De ahí mi cita de apertura: “Sacrificar un poco de precisión por una cobertura (mucho) más amplia”. A lo que podría añadir... “y una respuesta a la velocidad de un rayo”. Para el momento en que he puesto en marcha el sitio web de la Web of Knowledge o de Scopus, recordado mi contraseña y accedido, he terminado prácticamente mis búsquedas en Publish or Perish!

¿Por qué necesitamos este libro?

Aunque muchos investigadores han utilizado Google Scholar como fuente de datos, ninguno de ellos ha sido tan diligente en sus esfuerzos para proporcionar una mejor imagen de Google Scholar que el Grupo de Investigación EC3. Desde 2008, Emilio y su equipo han trabajado sin descanso para explorar el funcionamiento interno de Google Scholar. Por tanto, estoy encantada de que su experiencia hasta la fecha se haya integrado ahora en una monografía. Lo único que lamento es que no esté (todavía) disponible en inglés!

Referencias

- Harzing, A.W. (2016). Publish or Perish tutorial, a collection of tips to introduce users to the program's main functions in 80 easy chunks. Disponible en: <http://www.harzing.com/poptips/index.htm>
- Harzing, A.W. (2015). "Health warning: Might contain multiple personalities. The problem of homonyms in Thomson Reuters Essential Science Indicators". *Scientometrics*, 105(3), 2259-2270. Disponible en: <http://www.harzing.com/papers.htm#esi>
[Cobertura de prensa en:
<http://www.thetimes.co.uk/tto/science/article4543774.ece>
<https://www.timeshighereducation.co.uk/news/week-higher-education-%E2%80%933-september-2015>]
- Harzing, A.W.; Mijnhardt, W. (2015). "Proof over promise: Towards a more inclusive ranking of Dutch academics in Economics & Business". *Scientometrics*, 102(1), 727-749. Disponible en: <http://www.harzing.com/papers.htm#top40>
- Harzing, A.W. (2013). "Document categories in the ISI Web of Knowledge: misunderstanding the Social Sciences?". *Scientometrics*, 93(1), 23-34. Disponible en: <http://www.harzing.com/papers.htm#isi>

Professor Anne-Wil Harzing

Middlesex University London
Hendon, London 2016, January 29
www.harzing.com

Introducción

Cuando yo cursaba el bachillerato, allá por los años setenta del siglo pasado, mis manuales de literatura dividían invariablemente el estudio de un autor en dos partes: la dedicada a desvelar la vida y la enfocada a conocer la obra. Mis profesores, y ahora vienen a mi memoria las figuras del padre Severino y de Don Francisco, nos recalaban que no es posible entender la obra de un autor sin indagar sobre su vida, por lo que era importante no sólo conocer las creaciones literarias sino también sus entresijos; sabía máxima que no es más que una constatación existencial: en lo que hacemos está lo que somos. Añado yo que, en algunos casos como el de nuestro egregio premio nobel Ramón y Cajal, no se entiende la vida sin la obra; porque créanme hay casos en que la obra es la vida.

Pues bien, pretende esta breve introducción explicar cuál es la trama vital que explica la concepción y ejecución de esta obra. Para empezar diré que es el resultado de siete años dedicados de forma intensa y extensa a descubrir las entrañas de Google Scholar. Todo empezó allá por el año 2008 cuando los organizadores del *V Foro sobre evaluación de la calidad de la educación superior y de la investigación*, celebrado en San Sebastián, me encargaron la organización de una sesión docente sobre *Google Scholar*. Llevó por título "Qué es y cómo utilizar Google Scholar" y, en esa sesión, aparte de dar a conocer a una audiencia expectante en saber lo que era y prometía el nuevo buscador académico, intenté por primera vez explicar qué cubría, cómo funcionaba, qué ofrecía, por qué podía ser interesante su uso, concluyendo con un relato de sus fortalezas y limitaciones.

Venía yo de publicar una serie de artículos sobre la *Web of Science* y su universo de productos bibliográficos y bibliométricos, y *Google Scholar* apareció ante mí, desde el primer momento, como una alternativa novedosa y extraordinariamente sugestiva para buscar y acceder a documentos científicos imposibles de recuperar con las bases de datos tradicionales. Pero inmediatamente advertí de sus potencialidades para la evaluación de las publicaciones científicas; es por lo que la ponencia que presenté en el siguiente foro llevó ya por título "Google Scholar: ¿herramienta de evaluación científica?". A partir de ahí empecé a bucear y pes-

car en las vastas y cambiantes aguas por las que navegaba el buscador, quedando atrapado en sus redes.

Pero el impulso decisivo para llevar a buen puerto toda nuestra labor fue el feliz encuentro con los otros firmantes de esta obra: Enrique Orduña, primero, y Alberto Martín y Juan Manuel Ayllón, después. Enrique, que ya hizo unos primeros tanteos sobre el estudio de la presencia de las universidades españolas en *Google Scholar*, desplegó en su excelente tesis un portentoso abanico de técnicas y herramientas webmétricas, fundamentales para conocer los intestinos de *Google* y *Google Scholar*, como buscadores de información en la Web. Si feliz fue el encuentro con Enrique, más que dichosa y venturosa fue la incorporación al grupo de trabajo de Juanma, becario FPI (2013), y Alberto, becario FPU (2014). La tenacidad, laboriosidad y las brillantes habilidades técnicas de que hacen gala han sido el complemento perfecto para los resultados alcanzados por nuestro grupo. Formamos un auténtico equipo, y todos los logros alcanzados se deben a la concordia, fraternidad y camaradería que reina entre nosotros. La suma de habilidades diversas (siempre el todo es superior a las partes) es posible porque se asienta en personalidades caracterizadas por la bondad, generosidad, comprensión, compromiso y honestidad. En definitiva, la obra es el resultado inexorable de la vida.

Sobre estas bases hemos construido un más que exitoso conjunto de resultados. Y a las pruebas me remito: un balance del trabajo desplegado durante estos años ha dado por fruto:

- La confección de 9 productos de evaluación bibliométrica (*H Index Scholar*; *Journal Scholar Metrics*; *Publishers Scholar Metrics*; *Proceedings Scholar Metrics*; *Scholar Mirrors*; *Classics Scholar Profiles*; *Indice H revistas científicas españolas según Google Scholar Metrics* (ediciones 2007-2011, 2008-2012, 2009-2013, 2010-2014, 2011-2015); *La Biblioteconomía y Documentación española según Google Scholar Citations*).
- La publicación de 17 artículos en revistas, 21 documentos de trabajo y 15 informes.
- La edición de un boletín (<http://googlescholar digest.blogspot.com.es>) para dar cuenta de los resultados de investigación que se vayan publicando sobre *Google Scholar* como fuente de información y evaluación científica.

Todo este trabajo entusiasta y prolongado en el tiempo permite que pueda afirmar con orgullo - y sin ánimo de ser pretencioso - que, después de los creadores de *Google Scholar*, somos seguramente uno de los grupos que más saben acerca de los secretos que atesora el buscador.

Y ahora que ya conocemos las mimbres y la urdimbre toca hablar del cesto, esto es, de este libro. Y hay que empezar necesariamente por sus prolegómenos, que constituyen todo un lujo.

El título, como no podía ser menos, resume sintéticamente todo lo que pensamos sobre GS: ha supuesto una auténtica revolución en la forma de buscar y de encontrar información científica y de acceder a ella de manera fácil y rápida. Ahora ya queda lejana a la par que anacrónica la forma de buscar información científica anterior a la aparición de *Google Scholar*: dispersa en múltiples fuentes (catálogos de bibliotecas, bases de datos bibliográficas de revistas en distinta disciplinas, de patentes, de tesis, etc.), costosa (la mayoría de las bases de datos exigen cuantiosas suscripciones) y con complejas y poco amigables interfaces de búsqueda.

Hoy la simple Caja de búsqueda de *Google*, que se ha implantado como un estándar de facto, se encarga por sí sola de traducir las palabras que se escriben en ella en numerosos y pertinentes resultados de búsqueda escritos en cualquier lengua, producidos en cualquier país, bajo cualquier formato y tipología documental. Y todo ello ofrecido de forma gratuita. Con ello se hace realidad el sueño de Anurag Acharya de poner a disposición de los que menos recursos tienen toda la información científica que cuelga de la Web; sin duda una manera de ayudar a transformar el mundo en la buena dirección...

Subtitulamos la “caja de pandora académica” porque GS proporciona información que permite evaluar el interés y el impacto de los documentos, y de los autores que los publican. Se ha convertido, sin quererlo, en una valiosísima fuente de datos con múltiples usos bibliométricos.

El libro viene precedido de dos prólogos firmados por dos de sus mejores conocedores: Peter Jacsó, el investigador que más ha escudriñado GS, el que lo ha hecho más inquisitivamente y con más acierto, sometiendo al tercer grado al buscador para desvelar sus virtudes y carencias; y Anne-Wil Harzing, creadora de *Publish or Perish*, un magnífico software que ha hecho factible el uso bibliométrico de *Google Scholar*, contribuyendo decisivamente al masivo uso del buscador y a la popularización de la propia bibliometría, como herramienta de evaluación científica.

Dado que este es un libro con múltiples caras, con estos prefacios se muestran esos dos rostros que conviven en *Google Scholar*, la del buscador y la de la herramienta bibliométrica, la faz crítica frente a la faz loable...

Respecto a la estructura del libro, sólo recalcar que lo hemos dividido en dos secciones. La sección A dedicada a describir minuciosamente

las características y prestaciones de *Google Scholar*. Se traza su origen y evolución, se desmenuza su funcionamiento general, se examina su tamaño, cobertura y crecimiento, se pormenorizan las prestaciones y servicios que proporciona como buscador y se apuntan sus fortalezas, debilidades y peligros. Dado que *Google Scholar* presenta una doble cara, la de buscador (razón por la que nació y que realmente lo identifica) y la de herramienta de evaluación (una aplicación sobrevenida y no buscada inicialmente por sus creadores), se realiza un repaso exhaustivo de toda la literatura científica que ha indagado empíricamente sobre el comportamiento de *Google Scholar* hasta hoy en esta doble faceta; de manera que viene a ser un estado de la cuestión.

La sección B la hemos dividido en dos apartados; el primero dedicado en exclusiva a la descripción de los principales productos derivados de *Google Scholar*, y que conforman lo que denominamos la familia *Google Scholar*: *Google Scholar Citations* (Mis citas, en español), nacido en julio de 2011 como un servicio a los autores para que generen un perfil donde se muestren los documentos publicados y recogidos en *Google Scholar*, así como el número de citas que cada uno de ellos ha recibido, generando una serie de indicadores bibliométricos. Y *Google Scholar Metrics* (Estadísticas en español), nacido en abril de 2012 como ranking de publicaciones científicas ordenado en función del *h-index* a partir de los datos de citas de *Google Scholar*, que permitiera a los usuarios identificar las más influyentes revistas, congresos y repositorios. De forma sintética hemos descrito sus características, funciones y prestaciones, señalando sus fortalezas y debilidades. En el segundo apartado hemos incluido aquellos productos independientes a la compañía y que se han gestado para ofrecer distintas herramientas bibliométricas: *Publish or Perish*, *Scholarometer*, *H Index Scholar*, *Journal Scholar Metrics*, *Publishers Scholar Metrics*, *Proceedings Scholar Metrics* y *Scholar Mirrors*.

En fin, creo que este libro ofrece una revisión omnicompreensiva a la par que minuciosa sobre lo que es a día de hoy *Google Scholar* y sus derivaciones.

Emilio Delgado López-Cózar

Profesor de la Facultad de Comunicación y Documentación
Universidad de Granada
Huétor Vega, 11 de marzo de 2016

Bloque A

Google Scholar:

el buscador académico universal

Capítulo 1

Los orígenes

No es sencillo encontrar una definición exhaustiva y globalmente aceptada de motor de búsqueda académico (*academic search engine*). Podría ser considerado como un buscador especializado en localizar información académica en la Web, tal y como proponía Codina (2007), que incorpora además ciertos elementos de valor añadido (indicadores de naturaleza bibliométrica como número de citas a los documentos académicos) que puedan ser utilizados para la evaluación científica y que estén finalmente abiertos a cualquier tipología documental (no sólo artículos sino patentes, presentaciones, material docente, etc.), siendo ésta la aproximación seguida por Ortega (2014) en su completo y exhaustivo estudio comparativo de motores de búsqueda académicos.

En la actualidad cohabitan diversos motores de búsqueda académicos, cada uno con sus propias características y funcionalidades. Por ejemplo, *BASE*¹ es quien aporta un mayor número de contenidos en acceso abierto, *CiteseerX*² destaca por sus prestaciones para la exportación de registros. *Scirus*³ (actualmente deshabilitado) destacaba por una interfaz de búsqueda superior mientras que *Microsoft Academic Search*⁴ (que tras desactualizarse ha renacido de sus cenizas recientemente) por sus servicios de valor añadido. Finalmente, *Google Scholar*, el motor de búsqueda académico desarrollado por *Google Inc.*, es el que dispone de una mayor cobertura de material académico.

Dada la revolución que ha supuesto este buscador, tanto en la búsqueda como en la evaluación de la actividad científica, este libro está dedicado exclusivamente a conocer su historia, describir sus características y funcionamiento y presentar algunas de sus aplicaciones métricas más importantes, con el fin de conocer cómo ha contribuido definitivamente a la popularización y democratización de la evaluación de la actividad científica.

1 <http://www.base-search.net>

2 <http://citeseerx.ist.psu.edu>

3 <http://www.sciencedirect.com/scirus>

4 <http://academic.research.microsoft.com>

Google Scholar es un buscador web académico (Ortega, 2014) gratuito que indiza la literatura científica en un amplio abanico de disciplinas, tipologías documentales e idiomas. Está por tanto especializado en buscar e identificar material bibliográfico de carácter académico proporcionando además una serie de servicios adicionales de gran valor, tales como la disponibilidad de las referencias bibliográficas de los registros en distintos estándares (permitiendo su posterior descarga a gestores de referencias), el acceso directo a los textos completos (cuando esto es posible por cuestiones legales) o la indicación del número de citas recibidas por cada registro.

Estos y otros muchos servicios serán desgranados a lo largo de la presente obra con el propósito de facilitar una visión holística del entorno académico que *Google Scholar* supone en la actualidad, un enfoque amplio que resulta necesario para comprender el verdadero impacto que este buscador especializado está generando en la sociedad académica de principios del siglo XXI.

Nacimiento

No sería posible relatar la historia y los orígenes de *Google Scholar* sin indagar en la figura de Anurag Acharya, actual *distinguished engineer* en *Google*, co-creador y alma de este buscador. Acharya nació en Bikaner (Rajasthan, India), donde se graduó en *Computer Science & Engineering* en el campus de Kharagpur del *Indian Institute of Technology*.

Tal y como el propio Acharya ha reconocido en diversas entrevistas (Van Noorden, 2014), ya en su época de estudiante (a mediados de los 1980s) tuvo conciencia de la importancia que suponía el hecho de conocer la existencia de información “ahí fuera”. Aunque no se pudiera acceder a ella por diversos motivos, advertir que “existían” contenidos relevantes en algún lugar suponía un auténtico cambio en las reglas del juego.

Creer y estudiar en una zona en desarrollo implicaba, entre otras muchas cuestiones sociales y políticas, terribles dificultades a los estudiantes para poder costearse y tener acceso a materiales académicos que fueran realmente relevantes en sus estudios. Para un estudiante medio indio esto era prácticamente imposible. Sin embargo, si conocías la existencia de cierto artículo, libro, informe o material académico, siempre cabía la posibilidad de escribir una carta al autor solicitándole una copia (Levy, 2015), procedimiento que el todavía estudiante Anurag utilizó en más de una ocasión.