

Viejas novedades en Google Scholar

Emilio Delgado López-Cózar

Facultad de Comunicación y Documentación

Universidad de Granada

edelgado@ugr.es

RESUMEN

Tras describir someramente el proceso de indexación de documentos en Google Scholar, se explica cómo se posicionan los documentos en el buscador académico y se repasan los criterios de inclusión de revistas en Google Scholar Metrics. Se concluye opinando sobre el uso o no de Google Scholar Citations para elaborar perfiles bibliográficos de revistas.

PALABRAS CLAVE

Google Scholar, Google Scholar Metrics, Google Scholar Citations, Revistas Científicas, Recuentos de citas, posicionamiento, ASEO

1. INTRODUCCIÓN

Perdonad que participe tan tardíamente en este hilo tan asombrosamente productivo que se ha gestado en la lista ([lwetel](#)) sobre las buenas prácticas en la indexación de revistas en *Google Scholar* (GS) y sus efectos en sus productos derivados *Google Scholar Citations* (GSC), *Google Scholar Metrics* (GSM), así como en la controversia sobre cuál debe ser el adecuado uso de los mismos. Buena señal, pues eso indica que lo que haga GS ya no es indiferente para la comunidad científica y profesional. Los que venimos predicando desde años, y empezamos como apóstoles en el desierto, acerca de la conveniencia de figurar en GS vemos con íntima satisfacción como las tornas han cambiado definitivamente.

Con toda la intención titulo este mensaje con un oxímoron porque el debate es tan interesante como poco novedoso. Desde que naciera GS allá por 2004 y sus productos derivados en 2011-2012 se ha escrito y opinado mucho –tal vez demasiado- e [investigado bastante menos de lo conveniente](#); el tema periódicamente viene reapareciendo como guadiana tal cual ha ocurrido ahora en *lwetel*. Aunque reconozco que esta personal opinión puede ser el efecto y sesgo lógico de llevar los últimos diez años dedicados monotématicamente a desvelar las entrañas de GS, GSC, GSM y a someterlos a una intensiva y extensiva “explotación” bibliométrica.

Con la intención de esclarecer y explicar lo más didácticamente posible estos asuntos recurriré para empezar a una metáfora que ya venimos empleando desde hace varios años (incluso titulamos así un producto [Scholar Mirrors](#)). GS es un espejo y en él se refleja lo que se manifiesta; y digo manifiesta porque si un documento, una revista o un autor no se manifiestan o lo hacen incorrectamente (bibliográficamente hablando) el resultado será ser invisible o figurar defectuosamente. No vale, pues, esperar que el espejo diga cosas de nosotros que no somos, tal cual ocurría con la malvada madrastra

de *Blancanieves*. La realidad al final, como la belleza de *Blancanieves* y la fealdad de la bruja, se impondrán inexorablemente.

Recientemente en la lista INCYT escuchaba lamentos, lloros e incluso gritos sobre el maltrato que *Web of Science* y *Scopus* infringían a revistas latinoamericanas por ignorarlas y no incluirlas en sus fondos. Alguna de las personas que más chillaron, editora de una de esas revistas, incluía un enlace a la misma en el pie de su firma. Pues resulta que el enlace estaba roto y, algo más, la revista ni siquiera era ya editada. ¿Cómo es posible reclamar estar sin ser?

Las bases de datos inclusivas como GS, a diferencia de las selectivas, no ponen ninguna traba a que una revista sea indizada. Lo único que hay que hacer es ser pulcros en su edición digital y en su manifestación bibliográfica (metadatos en la web, incrustados en el propio documento o ¡en donde sea!...). Normalización bibliográfica, uniformidad en la descripción... ¿os suenan estos términos de pretéritas épocas cuando en nuestro mundo reinaba la catalogación? Nada nuevo bajo el sol...

En definitiva, para que alguien pueda hacer algo por ti, lo primero es que lo hagas tú mismo. Si ni siquiera tú eres capaz de hacerlo... no lo va a hacer el sistema por ti. Además, aplicar los estándares para la correcta interoperabilidad de la información digital es algo que depende en exclusiva de nuestra voluntad pero, sobre todo, de nuestra pericia técnica.

2. INDEXACIÓN EN GOOGLE SCHOLAR

La correcta indización de una revista en GS o cualquier otro motor de búsqueda depende directamente de que los robots, spiders, crawlers de que se sirven puedan rastrear con facilidad y rapidez los sitios web. Permittedme de nuevo una metáfora. Los robots son eso, robots; son programados para que actúen de forma simple y lógica. Esta sería su elemental forma de actuación:

1. Lo primero es encontrar el sitio web (la puerta de entrada). Para ello, urls sencillas, cortas, con pocos niveles y significativas. Si no la encuentran pasarán de largo.
2. Lo segundo es que la puerta esté abierta. Muchas webs sin saberlo no permiten el paso a los robots. Hay que poner en robots.txt: *User-agent: Googlebot Allow: /*
3. Una vez que el robot franquea la puerta conviene que los contenidos a indizar (esto es los artículos y sus pdfs) figuren inmediatamente y bien organizados (una lista es de artículos en una página html con enlace al texto completo es más que suficiente). Si el robot ha de pasar sucesivamente por varias estancias abriendo puertas y, además, los objetos figuran amontonados y no organizados en “estanterías virtuales” se irá, se perderá o dará palos de ciego. Os propongo un simple test de accesibilidad Por tanto, no más de tres niveles de profundidad aparte de la home.
4. A los robots le gustan los metadatos de la web... Todo lo que está bien etiquetado será bien empaquetado. A GS le privan especialmente las metaetiquetas de *Highwire*,

Eprints, *BE Press*, *Prism*. Si no hay más remedio utiliza *Dublin Core*. Aunque os parezca increíble *Dublin Core* no tiene campos inequívocos para el título de la revista, el volumen, el fascículo y los números de página. No es *Dublin Core* el estándar preferido como alguien dedujo erróneamente.

De manera práctica, buena parte de estos problemas serían resueltos si la revista adopta el gestor de revistas OJS (*Open Journal Systems*), recomendado abiertamente por GS. En el caso de los repositorios Google se inclina por *Eprints*, *Digital Commons*, *DSpace*.

¿Qué hacen los robots de GS cuando los metadatos no figuran o son inexactos? Dirigirse directamente a los documentos. Si están en pdf, han desarrollado unas heurísticas para identificar los principales elementos de la referencia bibliográfica (título artículo, autores, referencia bibliográfica del artículo en encabezamiento o pié de página, lista de referencias bibliográficas). Si no se hace en este orden y marcando claramente en con los recursos tipográficos apropiados (tamaño y estilo de la letra), el parser fallará. Y eso explica muchos de los monstruosos errores que podemos cazar en GS.

Evidentemente si en el pdf somos capaces de incrustar el nombre del autor, el título del trabajo, resumen y palabras clave, miel sobre hojuelas... Nunca estorba bien sea para un motor de búsqueda o para otro tipo de programas como los gestores bibliográficos tipo Mendeley...

Viendo el funcionamiento de GS lo que resulta más que sorprendente es los buenos resultados que ofrece. Peinar la variopinta web con documentos producidos en las condiciones más diversas, en los formatos y lenguas más singulares y en las ubicaciones más heterogéneas es una hercúlea tarea. Y ojo, GS ha sido capaz de generar controles internos que pueden detectar errores sistemáticos.

Justamente eso es lo que ocurrió el año pasado cuando *Dialnet* quedó fulminado de GS. El [apagón digital de la producción científica española](#) que descubrimos se debió a una partida defectuosa de registros bibliográficos antiguos en *Dialnet* (un número elevado de registros tenían alteradas la cadena de firmas). GS después de comparar esos registros con otras fuentes consideró que *Dialnet* no era una fuente fiable, con las terribles consecuencias que todos ya conocéis. Que sirva este hecho de aviso a navegantes.

Y no hay diferencias entre lo que pueda exigir GS u otros sistemas. En *Crossref* ya sabemos que si usted quiere que sus documentos estén correctamente referenciados y contengan sus listas de referencias citadas debe suministrar la lista de referencias y autorizar al sistema su libre uso. Si no lo hace, no se queje, por favor de que su revista no es "revista fuente" para el sistema y es maltratada...

En conclusión, reitero lo dicho anteriormente: la visibilidad de una revista en los motores de búsqueda depende de la propia revista. Si usted lo hace bien, normalmente, su espejo le recompensará.

Como el texto ya excede más de lo aconsejable en este tipo de mensajes, me remito al capítulo 4 del libro [La revolución Google Scholar](#) donde explicamos esto con el detalle

debido... y a unos próximos mensajes en la lista donde trataré sobre lo que ha quedado pendiente: posicionamiento en GS, Buen uso de los perfiles en GSC, Indexación en GSM, las alternativas a GS...

3. POSICIONAMIENTO EN GS... *EL SER O NO SER*

Si en el epígrafe anterior me ocupaba del estar, o sea, de los requisitos necesarios a cumplir para estar indexado correctamente en GS, y por extensión en cualquier otro motor de búsqueda, en este me centraré en el ser, esto es, en el lugar ocupado en la lista de resultados que nos devuelve un buscador ante cualquier consulta. Es evidente que sin estar es imposible ser. Pero una vez que estamos, todos queremos ser. Ser o no ser, el drama hamletiano trasladado a la página de resultados del buscador. Porque ser o no ser vistos, -y por ende visitados, leídos y citados-, depende de la posición que ocupen los documentos en dicha página de resultados. Si no estamos en la primera o primeras páginas de resultados mal asunto: seremos fantasmas deambulantes... Apelo a vuestra experiencia personal ¿Cuántos de vosotros pasáis de la primera página de resultados? La respuesta precisa la tienen los ingenieros de Google... disponen del comportamiento parametrizado estadísticamente de miles de millones de consultas. Y es por ello, entre otras razones, por lo que el buscador no ofrece más de 1000 resultados por consulta y más de 100 resultados para ser visualizados en una página. En el caso de Google Scholar hace unos años se bajó de 100 resultados por página a 20, mostrando por omisión 10. Es evidente que GS ya sabía, cuando tomó esta decisión, que los usuarios sobrepasan a duras penas las primeras páginas de resultados.

En las BD (bases de datos) tradicionales (*WoS, Scopus, Pubmed...*), en cambio, lo importante es estar. Una revista indexada “cover to cover” tenía ya garantizado que sus contenidos podían ser buscables por cualquier científico que consultara la BD. En dichas BD se muestran todos los resultados de la búsqueda y éstos pueden ser visualizados (10,25,50,100,200 por página) y ordenados según el gusto del usuario (más recientes, más antiguos, más o menos citados, alfabéticamente por nombre autor, revista, congreso...). Por defecto, los resultados se ordenan por fecha, apareciendo en primer lugar los más recientes. Las BD entienden que es la novedad el criterio que deben primar en la visualización. En resumen, el usuario posee libertad de acción; la BD pro-pone y el usuario dis-pone.

Esta filosofía y modo de actuación cambió radicalmente en los motores de búsqueda, en general, y en Google/Google Scholar, en particular. Ahora es el buscador el que pro-pone y dis-pone. El margen del usuario es mínimo: se nos ofrece un máximo de resultados a los que acceder (1000) y visualizar por página (10/20), y se dis-ponen los resultados ordenados por “relevancia”, dando solo la opción al usuario de reorganizarlos por fecha de publicación. ¿Y qué es la relevancia para GS? Para responder a esta pregunta hay que hablar del algoritmo de posicionamiento empleado por GS.

Siendo verdad que el algoritmo de posicionamiento de Google (el PageRank, la joya de la corona sobre la que se asienta el éxito de esta empresa) y que el de GS, siguiendo la

opaca política de la casa, es el secreto de la *Coca-Cola* desde hace años sabemos, mejor dicho intuimos, que son cientos de variables los que los conforman; pero de entre todas ellas son las citas recibidas por los documentos el elemento clave.

Que *Larry Page* y *Sergei Brin* en su famoso artículo “PageRank: Bringing order to the web” donde proponen el algoritmo que sirvió de base a Google se inspiraron en las ideas de *Garfield* sobre la citación para determinar la importancia y calidad de una página web es un hecho bien conocido. Lo que sabemos, en el caso de GS, por centrarnos en el motor que nos ocupa, gracias a las indagaciones de [Beel, Gipp y Wilde](#) en 2010 es que las citas juegan un papel fundamental. Así lo certificamos nosotros en 2014, donde sobre una muestra de [64000 documentos altamente citados entre 1950 y 2013](#) se vio a las claras que los documentos se ordenaban de acuerdo a las citas que recibían. Recientemente en un artículo publicado en [EPI](#) se confirman, de nuevo, los resultados anteriores añadiendo que ese es el criterio de ordenación empleado cuando las búsquedas son por autor o fuente de publicación (revista, congreso).

Los ingenieros de GS están convencidos que nada mejor para presentar los resultados de una búsqueda que el número de citas de los documentos, ya que estas son la mejor manera de fijar la relevancia, importancia, significación, repercusión de un trabajo. Los documentos más citados lo son porque la propia comunidad científica y profesional libremente lo ha decidido; son, por tanto, esos documentos imprescindibles que no deben ignorar los lectores. No olvidéis, que la meta de GS es encontrar documentos relevantes para un lector. Las citas no son un fin en sí mismo, como para los bibliómetras que ven en ellas un medio para medir la calidad, sino que tienen un carácter instrumental: sirven para identificar los documentos más notables y valiosos sobre un tema, de un autor o de una revista. De ahí que los documentos más citados se sitúen preferentemente. Repito, su posición se basa en un principio muy democrático pero al mismo tiempo muy meritocrático, pues es expresión objetiva y cuantitativa de la opinión de la comunidad científica en su conjunto, que es la que se expresa a través de ellas.

Prueba de este papel tan extraordinario que GS otorga a las citas es que todos los productos que han generado, aparte del propio buscador GS, son ordenados por defecto según su citación:

- Los documentos en los perfiles bibliográficos de un autor (*Google Scholar Citations*).
- Las revistas en los rankings por lenguas y disciplinas de *Google Scholar Metrics*. Y dentro de las revistas, los documentos que contribuyen a su índice h.
- Los documentos incluidos en los *Classic Papers*, esos documentos altamente citados en las distintas disciplinas y especialidades.

Si es de una lógica aplastante que el número de citas sea el criterio de ordenación de resultados dominante en búsquedas por fecha, autor, fuente de publicación, lo es también en el caso de las búsquedas temáticas pero con matices, pues en este tipo de consultas existen otros muchos factores que juegan un papel trascendente;

próximamente desvelaremos cuales son cuando llegue a buen puerto la investigación en la que nos encontramos ahora. Obviamente, dentro de estos factores desempeña una función primordial la presencia de las palabras empleadas en una búsqueda en títulos, resúmenes, palabras clave, títulos de los ficheros y urls. No olvidemos que GS es un buscador basado en palabras clave. Qué contestaríais a las siguientes preguntas si tuvierais que diseñar vosotros un algoritmo de búsqueda y recuperación de documentos:

- ¿Es razonable otorgar un peso especial a aquellos vocablos que figuren en el título, en el resumen y las palabras clave de un documento?
- ¿Es razonable favorecer los documentos que contengan esos vocablos en la misma url o en los títulos de los ficheros o en los metadatos incrustados en el propio pdf?

Si estos elementos se emplean por los autores para denominar sus documentos, representarlos y etiquetarlos temáticamente hablando, para resumirlos de manera sintética y precisa, es lógico pensar que los vocablos que figuren en ellos encarnen y personifiquen a los documentos. De ahí que se les otorgue un peso especial en las búsquedas. Para GS el resumen es tan importante, que es un requisito previo y general para que un documento pueda ser indizado.

De todo lo dicho, y a modo de recomendación para autores y revistas, se colige que debe extremarse el cuidado en la redacción de títulos, resúmenes, palabras clave:

- Deben ser claros, precisos y significativos. Deben reflejar los objetivos, la metodología y los resultados (en el caso de trabajos de investigación), y las conclusiones obtenidas.
- Deben estar cuajados de palabras clave que sean representativas del contenido de los documentos y, en la medida de lo posible, redactados de acuerdo a los vocabularios y tesauros que sean de uso corriente en las distintas disciplinas
- Deben ser siempre traducidos al inglés, en tanto en cuanto, este idioma se ha convertido en lengua franca de la ciencia.

Si estas recomendaciones dependen exclusivamente de autores y revistas no ocurre lo mismo con las citas. Estas dependerán de la relevancia y calidad del trabajo. Cosa no menor... realmente el quid de la cuestión

No quisiera terminar esta nota sin reflexionar sobre las consecuencias no de orden práctico sino de fondo, filosóficas diría yo, que tiene este modo de funcionamiento. El que el algoritmo de GS privilegie en las páginas de resultados a los documentos más citados tiene un efecto inmediato. Los documentos más antiguos por el hecho de haberse producido antes tienen mayor potencialidad de citación que los documentos recientes. Aparecer en esta primera página de resultados les otorga una mayor probabilidad de ser leídos y, por ello, de ser citados de nuevo. De ahí que se genere un bucle o círculo virtuoso permanente. ¿Tendrá que ver este fenómeno con el incremento de citas de los documentos más antiguos que viene observándose en los últimos tiempos? Esto es lo que insinuamos en 2016 en un trabajo publicado en [Scientometrics](#).

Por otra parte, que GS ofrezca al usuario tan pocas posibilidades de modificar el orden en que quiere visualizar los resultados (solo por fecha podemos reordenarlos) plantea un problema de hondo calado: ¿estamos determinados en nuestras lecturas por el propio buscador? Aunque siempre el lector tiene la decisión última sobre lo que lee o no lee, conociendo las prácticas de comportamiento de los usuarios que se acomodan a las propuestas del buscador puede estarse condicionando la percepción del científico, en particular, y de la ciencia, en general.

El éxito de Google no sólo estribó en encontrar un método fiable para ordenar las páginas web de interés para el usuario sino que al ser capaz de escudriñar lo que el usuario hace en cada momento adoptó una perspectiva subjetiva en la información. Ahí realmente radica su propuesta revolucionaria y su éxito. No existe realidad informativa objetiva (la que sirvió para montar todos los sistemas tradicionales de información) sino que la información solo existe en la mente de un sujeto que la produce y la consume. Al conocer de manera precisa su comportamiento informativo (el de él personalmente cuando tiene una cuenta Gmail activada o el de la IP del ordenador desde donde se trabaja y consulta) sabe qué es lo que demandamos, y por elevación, lo que necesitamos en cada momento. Es por eso por lo que Google parece adelantarse a nuestros deseos cuando empezamos a escribir en la caja de búsqueda... nos pro-pone... y acierta. Él como, como otros sistemas de recomendación, parece que nos da lo que nos gusta, porque sabe lo que nos gusta. Cuanto más le damos, más sabe de nosotros, y más nos gusta lo que nos ofrece. Una atracción fatal.

Si esta filosofía anida también en Google Scholar, algo que parece niegan los ingenieros de GS, nuestro ser está condicionado.

Ser o no Ser el drama hamletiano con el que comenzaba este mensaje... Caminamos a **hombros de un gigante**

4. LOS ESCAPARATES DE GS: GOOGLE SCHOLAR METRICS & GOOGLE SCHOLAR CITATIONS

Si en los mensajes anteriores me consagré a dilucidar sobre como estar en *Google Scholar* (indexación) y como ocupar, dentro de él, un buen sitio (posicionamiento), ahora me centraré en los productos derivados del buscador: GSM (*Google Scholar Metrics*), la herramienta destinada a identificar las revistas científicas de mayor impacto según las citas recopiladas de GS, y GSC (*Google Scholar Citations*) el portal que facilita a los autores crear un perfil bibliográfico-bibliométrico de sus publicaciones indexadas en GS. Hablaremos, por tanto, no del estar, no del ser, sino del figurar, aunque sólo los que son y están pueden figurar.

No se me ocurre una mejor manera de definir estos productos sino como los escaparates de GS. Todos los días cuando transitamos por las calles de nuestras ciudades nos topamos con cientos de tiendas de todo tipo, mejor dicho, con los escaparates de las mismas donde se exhibe una selección muy bien adornada de las mercancías que los comercios atesoran en sus estantes o guardan en sus almacenes. ¿Qué es si no GSC?

el escaparate que *Google* ofrece a los autores para que muestren sus bienes más preciados, esto es, sus obras; ¿qué es si no GSM? el escaparate donde se presentan las mejores revistas científicas con sus mejores galas, quiero decir, con sus mejores artículos.

Con GSC y GSM, *Google* constituyó una auténtica [familia](#). Espoleado seguramente por la competencia, *-Microsoft Academic Search* había lanzado su novedoso producto en 2009 donde permitía buscar documentos, autores, instituciones revistas, congresos; *ResearchGate* y *Academia.edu* ofrecían la posibilidad de crear perfiles donde los autores pudieran compartir y difundir sus ideas y publicaciones-, *Google* decidió lanzar en 2011 GSC y en 2012 GSM. En su afán de abrir nuevos mundos, *Google* nos brindó nuevos medios de encontrar información relevante. Quiero subrayar, por enésima vez en esta serie de mensajes, que GS no es ni más ni menos que un buscador de información científica; GSC y GSM vinieron para enriquecer esta misión, facilitando la identificación de los científicos y de su producción, en un caso, y de las revistas, congresos (solo en Computación), repositorios (solo arxiv, nber) y sus artículos y comunicaciones más relevantes.

Fueron ideados con una aproximación original y diferente a los de la competencia pero con la marca de la casa: a imagen y semejanza del buscador general son rápidos, simples y sencillos de usar y entender, así como de acceso libre y gratuito. Siempre guiados a que los usuarios sean capaces de encontrar los documentos más relevantes, de autores o de revistas, entendiendo que estos son los más citados, como ya expliqué en el mensaje anterior. Los indicadores bibliométricos como el índice h, introducidos por primera vez en estos productos por *Google* y adoptados como medio de detectar el lustre de publicaciones y autores, solo tienen la función de mostrar qué autores o revistas han sido capaces de publicar muchos documentos y muy citados. GSC y GSM no son productos bibliométricos *per se* a pesar del empeño que se toman algunos en verlos como tales u otros en “explotarlos”.

4.1 Google Scholar Metrics: un buscador de revistas y artículos minimalista; una “herramienta bibliométrica low cost”

Renuncio a volver a describir las características, minimalistas funcionalidades y limitaciones que presenta este producto. Desde su nacimiento, que recibimos con entusiasmo, duras críticas ([EPI](#), [Cybermetrics](#)) y esperanzas ([Learned Publishing](#)), todos los años hemos saludado la aparición de cada una de las ediciones de GSM ([2007-2011](#), [2008-2012](#), [2009-2013](#), [2010-2014](#), [2011-2015](#)) a excepción del año pasado (2012-2016), con un informe resaltando novedades y recordando virtudes y, sobre todo, defectos; una última sistematización de sus carencias se puede encontrar en el [apéndice 7 de esta obra](#). Los errores en cada edición, cuando eran de bulto, fueron subsanados inmediatamente; algunas de las ideas sugeridas fueron adoptadas por *Google*, lo cual lógicamente nos llenó de satisfacción, y otras muchas no, como puede apreciarse en la lectura de dichos informes.

Y fuimos más allá: intentando testar la cobertura de GSM y forzando su “explotación bibliométrica” elaboramos el [Journal Scholar Metrics](#), donde se listan más de 9.000 revistas publicadas en caracteres latinos de Arte, Humanidades y Ciencias Sociales. Asimismo, y con igual propósito aunque con diferente factura hemos venido publicando el [Índice h de las revistas científicas españolas según Google Scholar Metrics \(2007-2011, 2008-2012, 2009-2013, 2010-2014, 2011-2015, 2012-2016\)](#), donde se demuestra la amplia cobertura de GSM, a pesar de sus restrictivos criterios de selección de revistas. Hemos sido capaces de identificar en torno a 1.000 revistas españolas en GSM, la mitad de la población de revistas circulantes en España. Ahora bien, si se compara con las 547 SJR-Citescore o las 117 de JCR, es todo un mundo.

A nivel global estimamos que la cifra de fuentes indizadas en GSM rondaría las 40.000. Ciertamente [no están todas las que son](#) pero son bastantes. Y, utilizando la metáfora que reza en el título del mensaje, de ellas en torno a 5.500 figuran directamente en el escaparate de GSM (las 100 con más índice h en 12 lenguas, excepto en inglés donde se muestra las 20 revistas con mayor h-5 index de 262 categorías temáticas) o en la propia tienda (buscando en el cajón de búsqueda por palabras clave se pueden recuperar 20 revistas en cada consulta). Para localizar el resto de revistas hay que acudir al almacén, es decir, a *Google Scholar*, donde figuran con mayor o menor exhaustividad prácticamente la totalidad de revistas científicas circulantes.

4.2 Indexación de revistas en GSM

A lo largo de estos años he contestado bastantes e-mails de directores de revistas intentando explicar el por qué sus revistas no figuraban en GSM. Creo que es fácil de explicar, aunque parece que difícil de entender para muchos. Tres son los requisitos, uno de ellos realmente es un prerrequisito: estar indizado en GS. Por favor, si usted no está en GS no va a estar en GS-M. Los dos criterios específicos de GSM son:

- Contar con al menos 100 artículos publicados en el quinquenio cubierto.
- Contar con al menos 1 cita recibida para los artículos publicados en el quinquenio cubierto.

Cuando digo que hay que estar en GS, el prerrequisito para figurar en GSM, conviene recalcar que hay que estar todo el año y las 24 horas cada día. No vale descansar los fines de semana, como hacen algunos repositorios, o caerse de la web por distintos problemas técnicos. Si los robots pasan por allí en ese momento y no encuentran la puerta (sitio web) o se encuentran la puerta cerrada, la revista quedará excluida. Para la alimentación de GSM, los robots rastrean la web en mayo-junio, un mes o dos meses antes del lanzamiento de GSM, que suele ser en el mes de julio. Repito, debe comprenderse esto muy bien: si el sitio web de la revista no está accesible en ese momento, la revista quedará marginada. Esto ocurrió el año pasado con *EPI* que se aloja en el repositorio de la *FECYT*. El sitio estaba en obras, en el momento en que el robot pasó, con la fatalidad de que *Dialnet*, donde *EPI* figura convenientemente representada,

había sido también apartado de GS por lo que ya explicamos en el [apagón digital](#) y el mensaje anterior.

Por tanto, lo que se ve en GSM de una revista es lo que los robots encontraron en GS a fecha de junio de cada año. Lo que no es razonable es comparar los indicadores de GSM con los que muestra la revista en cada búsqueda que se haga de ella en GS o con los que figuren en los perfiles creados en GSC, que están actualizándose continuamente. Los datos no pueden coincidir en ningún caso. GS es un buscador web y refleja lo que sus robots son capaces de encontrar en la web en cada momento. Es dinámico como la web misma. Esta es un hecho diferencial de un motor de búsqueda y una base de datos tradicional. En cambio, GSM mostrará lo que es la revista en junio del año en que se recogieron los datos. Una vez publicados los listados de GSM se mantendrán todo el año, hasta que sean sustituidos en la próxima edición por la actualización correspondiente. En fin, una foto fija al igual que ocurre con JCR, SJR, CiteScore.

Por tanto, ojo con las comparaciones infundadas pues dan lugar a interpretaciones incorrectas. Aunque no hay que descartar pequeñas diferencias producto de errores. Ya he dicho en otras ocasiones que la empresa de GS es titánica (encontrar cualquier tipo de documento, en cualquier lugar y en cualquier formato y lengua) como para no [cometer errores](#). Seguro que los humanos con el pico y la pala, quiero decir la tecla y el cajón de búsqueda, siempre pueden encontrar en GS artículos o versiones de los mismos que el buscador ha sido incapaz de identificar o vincular. Pueden ser frecuentes pero, en muchas ocasiones, son reflejos de errores en las fuentes que no afectan, en lo sustancial, a los indicadores bibliométricos.

Se ponía como ejemplo de pobreza de indicadores y de revista casi invisible a *Els Marges: revista de llengua i literatura*. Empecemos por señalar dos aspectos cruciales para una correcta representación de esta revista en GS:

- Cuando se busca en la web posee dos sitios (<https://www.elsmarges.cat>) y (<https://www.raco.cat/index.php/Marges>), bien es verdad que en la primera url cuando se pincha en *Hemeroteca* se redirige al segundo... Pero no dejamos de tener dos puertas.
- Lo peor: el título de la revista varía según la web (*Les Marges* o *Marges, Els*) y dentro de la misma revista.

Está claro que los problemas identitarios pasan factura: las consecuencias son que podemos encontrar artículos en GS sobre

“Marges, Els”

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=source%3A%22Marges%2C+Els%22&btnG=

“Els marges”

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=source%3A%22els+marges%22&btnG=

Es evidente que en GS “curar” los contenidos se convierte en “currar”. Pero mi lema es “prevenir mejor que curar” para evitar que el “curar” se convierta en el “currar”. Y en cualquier caso, me remito a la metáfora de mi primer mensaje: espejo, espejito...

Afortunadamente GS es muy listo, cuando se busca *Els marges* sin entrecomillado el resultado es

https://scholar.google.com/scholar?start=220&q=source:els+marges&hl=en&as_sdt=0,5

En términos de indicadores bibliométricos... pasar de un índice h global de 14 a 16... es bueno... pero no cambia radicalmente la posición de la revista...

No obstante, la existencia de errores en GSM, me remito a los informes publicados todos los años donde pueden encontrar una muestra de ellos, puedo afirmar que globalmente considerados son minúsculos; seguro que lacerantes para quien los sufre, pero estadísticamente insignificantes para poder calificar al producto como sesgado o incompleto. Llevamos ya cinco ediciones del *Índice h de revistas científicas españolas según Google Scholar Metrics* y los errores detectados, bien por presencia o ausencia injustificada bien por errónea representación bibliográfica son despreciables (no llegan al 1%)

Resumiendo, lo primero que debe hacer un editor cuya revista no figura en GSM es saber si cumple con estos tres criterios. Me resulta frustrante que después de explicarlo no se entienda; aunque realmente lo que creo es que no se quiere entender. Y, a partir de ahí lo lógico en estos casos: descalificación del producto. Las posiciones en los debates suscitados por estos temas son previsibles: si estoy bien colocado, magnífico producto; si estoy mal colocado, producto malo; si no estoy, el producto es francamente pésimo y detestable. Recientemente asistíamos a un debate en INCYT sobre JCR, SJR et alii, donde se reproducían estos estereotipos. ¿Dónde está la declaración de conflictos de interés? En ciencia las cartas deben estar siempre boca arriba. Dime dónde y con quién andas y te diré quien eres...

La orientación selectiva de GSM (revistas con 100 artículos en cinco años y una cita) no casa con la política inclusiva de GS (cualquier revista que cuelgue de la web académica y que cumpla los requisitos a los que me referí en el primer mensaje). Pero es que son productos distintos con metas distintas. Si GS da a cualquier artículo publicado en cualquier lengua, formato y lugar del mundo la posibilidad de ser buscado, encontrado, leído y citado, GSM pretende identificar a los mejores artículos y revistas de entre los millones de documentos indizados en GS. Cuando GSM elimina las revistas pequeñas o no activas y con un índice h igual a 0 (irrelevantes incluso para los propios que publican en la revista) lo que pretende es eliminar ruido y ayudar a los lectores a encontrar las publicaciones científicas más citadas en fechas recientes, las que la comunidad considera más relevantes. Al mismo tiempo, se propone ayudar a los autores a identificar los canales (revistas, congresos, repositorios) que concentran la mayor atención y el mayor número de documentos relevantes. Quieren auxiliarles a la hora de elegir los mejores medios para comunicar sus ideas y hallazgos.

Hemos recordado a los ingenieros de GS que estos criterios pueden dejar fuera a revistas de Arte, Humanidades y Ciencias Sociales publicadas en lengua inglesa o no inglesa donde el volumen de publicaciones es reducido, sin que esto vaya en desdoro de su calidad. En concreto, pedimos que se redujera el umbral de 100 a 50. [La respuesta fue no](#).

Por la misma razón no están dispuestos a mostrar directamente en sus escaparates más que las 100 revistas de más impacto de las lenguas más activas científicamente hablando. Si el inglés puede representar en torno al 70% de las publicaciones científicas circulantes, solo en este caso, están dispuestos a montar escaparates para cada una de las 262 categorías temáticas en que organizan el conocimiento. Sé que puede ser descorazonador para las lenguas minoritarias; desde luego no hay que ver demonios o fantasmas donde no los hay. Para Google las letras, en este caso, son números... números de documentos circulantes en una lengua y atención recibida por los mismos en función del número de citas que atraen.

¿Qué puede hacer una revista para mejorar su difusión y visibilidad si no figura en GSM?

Sin dudarlo, crear un perfil en GSC.

4.3 Creando perfiles de revistas en Google Scholar Citations

¿Qué problema hay en que una revista pueda crear un perfil en GSC?

Técnico, ninguno, puesto que los perfiles pueden ser creados con total libertad por los usuarios. Desventajas para la revista no se me alcanzan. Ciertamente GSC fue concebido como una plataforma para que los autores crearan sus perfiles. Esto es lo que tiene ser un producto *laissez faire, laissez passer*. Lo evidente y palpable es que muchas revistas lo están usando como escaparate. Y Google, que lo sabe, no va a cerrar el producto por ello. Lo que ha hecho es filtrar los perfiles de revistas de forma que no se puedan buscar desde la ventana de búsqueda de GSC y tampoco sean recuperables con un hiperenlace en GS.

Una revista debería crear el perfil en GSC por dos razones:

1. La primera porque aumenta su visibilidad como revista y la de sus artículos. Si no es directamente localizable en GSC, ni en GS, sí lo es en Google (<https://goo.gl/iYUjV6>). Y no olvidéis que Google es el principal buscador general y también científico, sobre todo en muchas áreas de humanidades y ciencias sociales, donde es tan usado o más que GS para búsquedas científicas. Considero, por tanto, un error lo que han hecho algunas revistas ocultando su perfil; tiran piedras contra su tejado
2. Monitorizar en tiempo real el rendimiento de la revista comprobando el impacto de los artículos publicados y teniendo un retrato bibliométrico de la evolución de la revista. Mantener el perfil de una revista será un ejercicio de transparencia muy plausible.

Además, acceder al perfil público de una revista nos permite, por utilizar algunos ejemplos de nuestra disciplina:

- Saber que de los cuatro artículos más citados en [Anales de Documentación](#), tres son traducciones de artículos originales publicados en revistas extranjeras; esto no es un desdoro para la revista sino que indica el acierto de los editores a la hora de seleccionar estos interesantes artículos.

- Indagar sobre los efectos de las identidades lingüísticas de una revista como [BiD: textos universitaris de biblioteconomia i documentació](#), cuya lengua de publicación preferente es el catalán, aunque acepta artículos en otras lenguas y, procura versiones bilingües. Pues justamente si se compara el impacto en número de citas de las versiones en español/inglés y en catalán los datos hablan por sí solos. Sobre una muestra de los artículos más citados nos encontramos estas proporciones: 74 a 3, 62 a 1, 60 a 1, 44 a 13, 48 a 2, 32 a 7, 34 a 3... (versión española frente a catalana, excepto el primer caso que es un artículo con versión inglesa). El impacto de los artículos escritos en catalán es el que es. No se puede llegar más lejos de las fronteras que uno se ha autoimpuesto. Una apuesta por usar una lengua minoritaria, totalmente legítima y muy respetable, tiene sus consecuencias. Asumámosla y no pidamos peras al olmo.

Eso sí usar el perfil en GSC para una revista o una editorial requiere como se señalaba con fina ironía mucho “curro”. Porque el buscador interno de GSC está preparado para identificar automáticamente trabajos de autores, pero no tanto de editoriales y revistas. Incluir en el perfil todos los trabajos publicados por la revista y unir sus diferentes versiones puede requerir un penoso e ingrato trabajo de antiguo bibliógrafo. Y si no se está ojo a vizar fácilmente se pueden colar trabajos indeseados (véase el caso de esta revista (<https://goo.gl/9T15VR>) donde la mitad de las citas que recibe lo son por un artículo que no es suyo. No recomiendo la actualización automática de este tipo de perfiles. Google actualiza dos o tres veces a la semana el perfil, y puede si no estamos continuamente en alerta, introducir el error. Y no hay nada peor que un perfil sucio. Aprovechando que el Pisuerga pasa por Valladolid, señalar que esta recomendación es válida para todos los autores que se apelliden Martínez, Fernández, García, Pérez, Hernández... Si encima se llaman María, José, Antonio tienen todos los números de lotería para recibir asignaciones que no le corresponden. Evidentemente, si te apellidas Zunzunegui, no debes preocuparte.

En definitiva, recordad siempre que en GSC “el perfil retrata a quien lo elabora. Esta amplia autonomía que Google concede al autor trasluce la filosofía desde la que se ha diseñado este producto: un sistema de información orientado al usuario (...). Se deja total libertad a los autores para que puedan configurar el perfil a su entero gusto y beneficio desde su identificación personal, profesional y temática hasta la vinculación de la producción bibliográfica que estimen pertinente pasando por el modo de actualización del perfil (automático o no). El autor puede añadir, editar, y eliminar los registros bibliográficos tal como le venga en gana. Google se despreocupa de la posible veracidad y precisión de los datos y deja en manos del autor toda la responsabilidad sobre lo mostrado”.

Siendo conscientes de las potencialidades de GS hace un par de años propusimos a Google que al igual que creó un portal para los perfiles bibliográficos de autor (GSC), creara portales de revistas (GS Journals) y de instituciones (GS Institutions). Les dimos argumentos en favor de la propuesta que podrían mejorar la calidad global de los datos que manejan en GS. Obtuvimos la callada por respuesta; aunque a los pocos meses GS creó el hiperenlace institucional que permite agrupar a los autores que declaran en su perfil su afiliación a la institución. Una forma elegante y barata de no meterse en más honduras. En cualquier caso, creo que, aparte de desviarles de sus reales intereses como buscador, estas propuestas podrían generales más trabajo y costes. Hay que recordar que el equipo que mantiene GS está formado por una docena de personas. Increíble pero cierto.

Perdonad por la extensión; soy consciente de que un texto de esta naturaleza es muy pesado para un medio como la mensajería electrónica; para piar siempre está twitter. Los debates surgen espontáneamente y hay que responder a ellos donde se produzcan. Mi ánimo siempre es no dejar tinta en el tintero, aunque a veces este se derrama, así que las alternativas a GS “no son para el verano” ... Cuando el calor amaine y tomemos resuello puede que hablemos de ello.

BIBLIOGRAFÍA

Google Scholar Digest. Research on Google Scholar: Emprical Evidences. <http://googlescholar Digest.blogspot.com/p/bibliography.html>

Delgado-López-Cózar, E., Martín-Martín, A. (2018). Apagón digital de la producción científica española en Google Scholar. Anuario Think EPI, 12. 265-276. <https://osf.io/preprints/socarxiv/7taqn/>

Orduña-Malea, E., Martín-Martín, A., Ayllón, J. M., Delgado López-Cózar, E. (2016). La revolución Google Scholar : Destapando la caja de Pandora académica. Granada: Universidad de Granada, UNE: Unión de Editoriales Universitarias

Beel J, Gipp B,Wilde E. (2010). Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar andCo. Journal of Scholarly Publishing, 41 (2): 176–190, doi: 10.3138/jsp.41.2.176

Martín-Martín A., Orduña-Malea E., Ayllón, JM., Delgado López-Cózar E. (2014). Does Google Scholar contain all highly cited documents (1950-2013)? <https://arxiv.org/abs/1410.8464>

Rovira C., Guerrero-Solé F., Codina L. (2018). Las citas recibidas como principal factor de posicionamiento SEO en la ordenación de resultados de Google Scholar. El profesional de la información, 27(3), 559-569.

Martín-Martín A., Orduña-Malea E., Ayllón J.M., Delgado López-Cózar E. Back to the past: on the shoulders of an academic search engine giant. Scientometrics (2016) 107: 1477. <https://doi.org/10.1007/s11192-016-1917-2>

Delgado López-Cózar, E. The Google Scholar Family: ¿Is it an alternative for the evaluation of science? En: IV Seminario EC3: Altmetrics y Unidades de Bibliometría. Universidad de Granada, Facultad de Comunicación y Documentación, 14-15 marzo 2013. <http://hdl.handle.net/10481/24142>

Martín-Martín, A.; Ayllón, J.M.; Orduña-Malea, E.; Delgado López-Cózar, E. (2016). 2016 Google Scholar Metrics released: a matter of languages. EC3 Working Papers, 22. 21st of July, 2016. arXiv:1607.06260.

Delgado-López-Cózar E., Cabezas-Clavijo Á. (2012). Google Scholar Metrics: an unreliable tool for assessing scientific journals. *El Profesional de la Información*, 21(4), 419–427. <http://dx.doi.org/10.3145/epi.2012.jul.15>

Delgado López-Cózar E. Robinson-García N. (2012). Repositories in Google Scholar Metrics or what is this document type doing in a place as such? *Cybermetrics*, 16(1), paper 4. http://digibug.ugr.es/bitstream/10481/22019/1/repositorios_cybermetrics.pdf

Delgado López-Cózar E. Cabezas Clavijo A. (2013). Ranking journals: could Google Scholar Metrics be an alternative to Journal Citation Reports and Scimago Journal Rank? *Learned Publishing*, 26 (2): 101-113. <http://arxiv.org/ftp/arxiv/papers/1303/1303.5870.pdf>

Cabezas-Clavijo Á., Delgado López-Cózar E. (2012). Scholar Metrics: el impacto de las revistas según Google, ¿un divertimento o un producto científico aceptable? EC3 Working Papers 1. <http://eprints.rclis.org/16830/1/Google%20Scholar%20Metrics.pdf>

Delgado López-Cózar E, Cabezas-Clavijo Á (2012). Google Scholar Metrics updated: Now it begins to get serious. EC3 Working Papers 8: 16 de noviembre de 2012. <http://digibug.ugr.es/bitstream/10481/22439/6/Google%20Scholar%20Metrics%20updated.pdf>

Cabezas-Clavijo Á, Delgado López-Cózar E (2013). Google Scholar Metrics 2013: nothing new under the sun. EC3 Working Papers, 12: 25 de julio de 2013. <http://arxiv.org/ftp/arxiv/papers/1307/1307.6941.pdf>

Martín-Martín A., Ayllón J.M., Orduña-Malea E., Delgado López-Cózar E. (2014). Google Scholar Metrics 2014: a low cost bibliometric tool. EC3 Working Papers, 17: 8 July 2014. <http://arxiv.org/pdf/1407.2827>

Martín-Martín A., Ayllón J.M., Orduña-Malea E., Delgado López-Cózar E. (2015). 2015 Google Scholar Metrics: happy monotony. EC3 Google Scholar Digest, 26 Jun 2015. <https://doi.org/10.13140/RG.2.2.12729.47202>

Delgado López-Cózar E., Orduna-Malea E., Martín-Martín A. (2018). [Supplementary material to book chapter] Google Scholar as a data source for research assessment. In W. Glaenzel, H. Moed, U. Schmoch, M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators*. Springer. <https://osf.io/preprints/socarxiv/pqr53/>

Martín-Martín A., Orduña-Malea E., Delgado López-Cózar E. (2017). Scholar Mirrors: Integrating evidence of impact from multiple sources into one platform to expedite researcher evaluation. STI 2017 Conference Science, Technology and Innovation indicators: “Open indicators: innovation, participation and actor-based STI indicators” <https://osf.io/preprints/socarxiv/z4bwe/>

Martín-Martín, A., Orduna-Malea, E., Delgado López-Cózar E. (2017). Journal Scholar Metrics: building an Arts, Humanities, and Social Sciences journal ranking with Google Scholar data. STI 2017 Conference Science, Technology and Innovation indicators: “Open indicators: innovation, participation and actor-based STI indicators”. <https://osf.io/preprints/socarxiv/vxnw6/>

Delgado López-Cózar E., Ayllón JM, Ruiz-Pérez R. (2013). Índice H de las revistas científicas españolas según Google Scholar Metrics (2007-2011). 2ª edición. EC3 Informes, 3: 9 de abril de 2013. <http://digibug.ugr.es/handle/10481/24141>

Ayllón Millán J.M., Ruiz-Pérez R., Delgado López-Cózar, E. Índice H de las revistas científicas españolas según Google Scholar Metrics (2008-2012). EC3 Reports, 7 (2013). <http://hdl.handle.net/10481/29348>

Ayllón JM, Martín-Martín A., Orduña-Malea E., Ruiz Pérez R, Delgado López-Cózar E. (2014). Índice H de las revistas científicas españolas según Google Scholar Metrics (2009-2013). EC3 Reports, 17. Granada, 28 de julio de 2014. <http://hdl.handle.net/10481/32471>

Ayllón JM, Martín-Martín A., Orduña-Malea E., Delgado López-Cózar E. (2015). Índice H de las revistas científicas españolas según Google Scholar Metrics (2010-2014). Granada: EC3, 2015. (EC3 Reports; 13). <http://hdl.handle.net/10481/36998>

Ayllón JM, Martín-Martín A., Orduña-Malea E., Delgado López-Cózar E. (2016). Índice H de las revistas científicas españolas según Google Scholar Metrics (2011-2015). Granada: EC3, 2016. (EC3 Reports; 17). <http://hdl.handle.net/10481/42829>

Delgado López-Cózar E, Martín-Martín A., Orduña-Malea E. (2017). Índice H de las revistas científicas españolas según Google Scholar Metrics (2012-2016). Granada: EC3, 2016. (EC3 Reports; 20) <https://doi.org/10.13140/RG.2.2.29279.56484>

Cabezas-Clavijo Á, Delgado-Lopez-Cozar, E. (2012). Las revistas españolas de Ciencias Sociales y Jurídicas en Google Scholar Metrics ¿están todas las que son? EC3 Working Papers, 2. <http://eprints.rclis.org/16892>

Orduña-Malea E., Martín-Martín, A., Delgado López-Cózar, E. (2017). Google Scholar as a source for scholarly evaluation: A bibliographic review of database errors. Revista Española de Documentación Científica, 40(4): e185. <http://hdl.handle.net/10481/48295>