Apagón digital de la producción científica española en *Google Scholar*

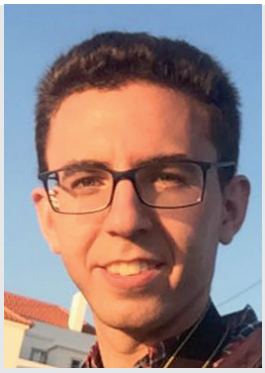
Digital blackout of Spanish scientific production in *Google Scholar*

Emilio Delgado-López-Cózar y Alberto Martín-Martín

Delgado-López-Cózar, Emilio; **Martín-Martín, Alberto** (2018). "Apagón digital de la producción científica española en *Google Scholar*". *Anuario ThinkEPI*, v. 12, pp. 265-276.

https://doi.org/10.3145/thinkepi.2018.40

Publicado en *IweTel* el 5 de diciembre de 2017



Resumen: En la última edición de *Google Scholar Metrics* (2012-2016) se detectó una brusca caída del número de revistas científicas españolas indizadas. Su número pasó de 1.101 en la edición 2011-2015 a 599 en la actual. Se realiza un análisis por disciplinas para comprobar cuáles han sido las más afectadas. Después de repasar varias hipótesis que puedan explicar este descenso, se concluye que ha sido la desaparición súbita de *Dialnet* en *Google Scholar* la responsable del incidente. Dado que *Dialnet* es, en muchos casos, el único lugar donde se puede encontrar información sobre un gran número de revistas españolas que carecían de exposición directa en la web académica, su eclipse ha implicado directamente la invisibilidad de todas estas revistas españolas que venían siendo indizadas en *Google Scholar Metrics*.

Palabras clave: *Google Scholar*; *Google Scholar Metrics*; *GS*; *GSM*; Repositorios institucionales; *Dialnet*; Revistas; Indización; Cobertura; Acceso abierto.

Abstract: In the most recent edition of *Google Scholar Metrics* (2012-2016) there was an abrupt drop in the number of Spanish scientific journals; the total was 1,101 in the 2011-2015 edition but fell to 599 in the 2012-2016 edition. We conducted an analysis by discipline in order to find which journals were affected. After considering several hypotheses to explain this phenomenon, we conclude the main cause was the sudden disappearance of the Spanish bibliographic database *Dialnet* from *Google Scholar*. Because *Dialnet* is the only online source that provides information about an important number of Spanish journals (which do not have any other online exposure), being dropped by *Google Scholar* meant that all these Spanish journals, which had previously been indexed in *Google Scholar Metrics,* became suddenly invisible.

Keywords: *Google Scholar*; *Google Scholar Metrics*; Institutional repositories; *Dialnet*; Journals; Indexing; Coverage; Open access.

Preparando nuestro ya tradicional Índice H de las revistas científicas españolas según Google *Scholar Metrics (2012-2016)* nos hemos topado con una desagradable sorpresa: el número de revistas únicas localizadas ha descendido casi a la mitad pasando de 1.101 a 599. Se rompe bruscamente la tendencia ascendente en la cobertura de revistas desde que nació el producto. Cada año el número de revistas españolas cubiertas por *GSM* (*Google Scholar Metrics)* se fue incrementando, desde las aproximadamente 900 revistas incluidas en la edición 2007-2011 hasta las cerca de 1100 revistas de la edición 2011-2015.

El hundimiento es generalizado –sólo en una

disciplina no se ha producido reducción- aunque ha afectado a unas especialidades más que otras (tabla 1). Las más perjudicadas han sido Derecho, Química, Economía y Empresa, Ciencia Política y de la Administración, Urbanismo e Ingenierías.

Inmediatamente nos pusimos a buscar explicaciones a este extraño fenómeno. Varias podían ser las causas:

- Que los robots de *Google Scholar* hubieran fallado estrepitosamente en su tarea. Es evidente que el buscador comete errores, pero es poco probable que yerre masivamente en la identificación de sitios web de revistas anteriormente ya procesadas.
- Que los criterios de inclusión de *GSM* hubieran cambiado drásticamente. También era poco probable, porque si ese hubiera sido el caso, los responsables de *Google Scholar* habrían informado de este cambio y habría sido reflejado en la documentación de la plataforma. De todas formas, para asegurarnos, contactamos con ellos y nos confirmaron que no había habido ningún cambio en los criterios de inclusión de *GSM*.
- Que las revistas hubieran incumplido los dos requisitos de indización que impone *GSM*. A saber: haber publicado 100 trabajos en los últimos cinco años o poseer al menos una cita en dicho período. Poco probable que multitudinariamente se incumplan estos requisitos de un año a otro y, menos aún, teniendo en cuenta la utilización de series quinquenales con variación de sólo un año (Ej.: 2011-2015 frente a 2012-2016). Es un sistema que propicia la estabilidad en los indicadores. Ni siquiera la no indización del último año provocaría el incumplimiento de estos requisitos por parte de muchas revistas.
- Que los sitios web donde estuvieran alojadas las revistas hubieran cambiado su arquitectura de un año a otro, incumpliendo los requisitos técnicos exigidos por el buscador; o bien que hubieran quedado deshabilitados temporalmente (fuera de servicio) en el preciso momento en que los robots pasaran a rastrearlos. Poco probable que estos cambios se produzcan simultáneamente en más de 500 revistas.

Tabla 1. Comparación del número de revistas españolas indizadas en las ediciones de *Google Scholar Metrics* de 2011-2015 y 2012-2016

Disciplinas	Número de revistas		% Reducción
	2011-2015	2012-2016	
Derecho	156	35	78
Urbanismo	16	5	69
Química	6	2	67
Multidisciplinar (Humanidades)	36	12	67
Economía y Empresa	87	30	66
Ingenierías	49	19	61
Ciencia Política y de la Administración	48	19	60
Ciencias agrarias	27	11	59
Arte	61	27	56
Documentación	13	6	54
Ciencias de la tierra	12	6	50
Ciencias de la vida	26	13	50
Física	4	2	50
Filosofía	51	29	43
Sociología	64	37	42
Filología Hispánica	58	34	41
Comunicación	37	24	35
Multidisciplinar (Ciencias)	3	2	33
Educación	110	79	28
Historia	81	58	28
Ciencias del Deporte	22	16	27
Filologías modernas	15	11	27
Geografía	23	17	26
Matemáticas	12	9	25
Ciencias de la Salud	177	134	24
Antropología	13	10	23
Psicología	50	40	20
Lingüística	21	18	14
Filología Clásica	7	6	14
Multidisciplinar (Ciencias Sociales)	5	5	0
Estudios hebreos, arabes y orientales	5	5	0

Tras rechazar todas estas hipótesis creemos haber descubierto la fuente del problema. Este se llama *Dialnet*. La desaparición súbita de *Dialnet* de *Google Scholar* ha implicado directamente la invisibilidad de todas las revistas españolas que venían siendo indizadas en *GSM*. Pensamos que se trata de la hipótesis más plausible porque *Dialnet* es, en muchos casos, el único lugar

donde se puede encontrar información sobre un gran número de revistas españolas que carecían de exposición directa en la web académica.

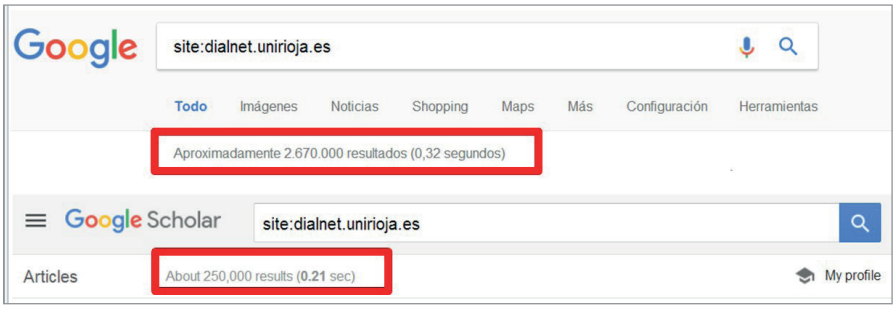


Figura 1. Número de items de *Dialnet* indizados por *Google* y *Google Scholar* el 29/11/2017.

Desde hace ya muchos años, la base de datos bibliográfica/recolector de texto completo *Dialnet* (desarrollada por la *Universidad de la Rioja*) ha sido una de las fuentes más completas para la identificación de publicaciones científicas editadas en España. Así lo hemos constatado reiteradamente en diversas búsquedas que con motivos distintos (seminarios, talleres, cursos, investigaciones, etc.) hemos venido realizando a lo largo de estos años. La última medida que habíamos tomado de la cobertura de *Dialnet* fue en marzo de 2017 para un capítulo de libro sobre *Google Scholar* publicado por *Springer* (**Delgado-López-Cózar** *et al.*, 2017). El número de items para la consulta "site:dialnet.unirioja.es" fue de 2.280.000. En el pasado, y en otras mediciones de la misma naturaleza realizadas con otros propósitos, siempre habíamos recuperado más de 2 millones para esta consulta, por lo que los resultados habían sido siempre muy coherentes.

Aunque sabemos que este comando "site" no es del todo fiable para medir con precisión la cobertura, ya que, entre otras limitaciones, sólo tiene en cuenta la versión primaria de los documentos (y no todas las versiones secundarias), sí es válido para obtener una idea aproximada del volumen de datos indizados por un sitio web. Los datos que proporciona son una prueba palmaria de que ha habido un cambio significativo en la indización que *Google Scholar* tiene de los documentos disponibles en *Dialnet*.

Hoy, sin embargo (29/11/2017) la misma consulta en *Google Scholar* sólo devuelve 250.000 resultados (figura 1, abajo). De pronto se han esfumado más de 2 millones de items. Si hacemos la consulta en *Google* para comprobar qué es lo que tiene el buscador general sobre *Dialnet* y para saber si el problema le ha afectado también, nos encontramos con 2.670.000 items (figura 1, arriba), cifra coincidente con lo que históricamente ha significado *Dialnet* para *Google Scholar*.

Para profundizar un poco más sobre qué es lo que *Google Scholar* tiene actualmente indizado de Dialnet, quisimos analizar los documentos que siguen estando vaciados actualmente en el buscador. El 1/12/2017 se realizaron búsquedas utilizando el comando "*site*" similares a las mostradas en la figura 1, pero filtrando por años de publicación, desde 1990 hasta 2017 (28 búsquedas en total). Todos los resultados que se mostraron para cada búsqueda fueron extraídos y analizados para identificar su tipología documental. Los resultados obtenidos se pueden observar en la figura 2. Los datos y el procesamiento que se realizó sobre los mismos están disponibles en el material complementario (**Martín-Martín**; **Delgado-López-Cózar**, 2017).

Como se puede apreciar en la figura 2, en ninguna de las búsquedas por año se llegaron a recuperar siquiera 1.000 items, a pesar de que el número de resultados estimado para cada búsqueda era siempre superior a 1.000. Es especialmente notable la diferencia entre el número de resultados estimado para los años 2015 y 2016 (en ambos casos eran superiores a 10.000), cuando el número de resultados reales que se

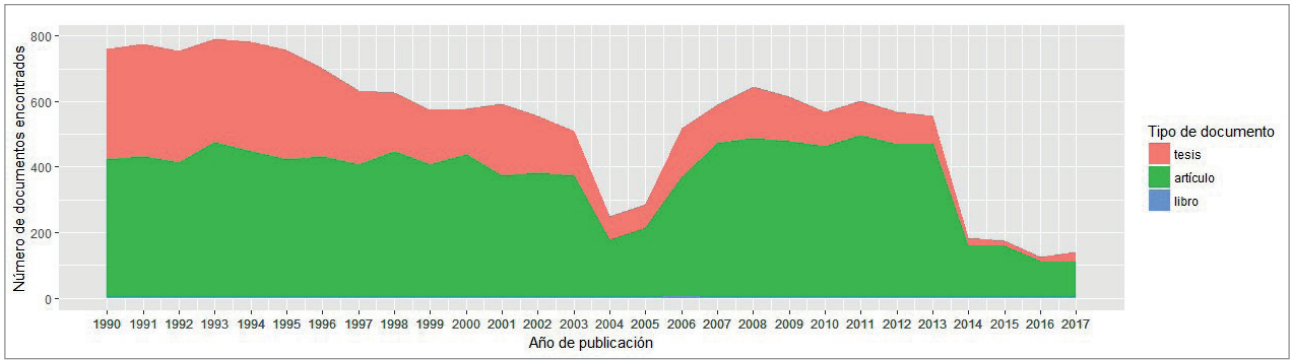


Figura 2. Número y tipología documental de documentos indizados actualmente en *Google Scholar* provenientes de *Dialnet* (1990-2017).

pudieron extraer en esos años no llegó a 200.

Respecto a los tipos de documentos que siguen estando indizados, aproximadamente un 69% (10.486) son artículos de revista cuyos textos están alojados en los servidores de *Dialnet*, y el 31% (4.709) restante son tesis doctorales. Hay un muy pequeño número de libros que también están alojados en los servidores de *Dialnet* (36, el 0,2% del total).

"El número de revistas únicas localizadas en *Google Scholar Metrics* ha descendido casi a la mitad pasando de 1.101 a 599 en los datos del periodo 2012-2016, frente a 2011-2015"

Para finalizar también quisimos averiguar si seguía habiendo entrada de nuevos registros en *Google Scholar* provenientes de *Dialnet*. Para ello realizamos de nuevo la búsqueda con comando *site*, y ordenamos por fecha. Comprobamos que *Google Scholar* sí sigue indizando documentos de Dialnet. Descargamos los 999 resultados que ofrecía el buscador y los procesamos para su análisis. Según la información declarada en la interfaz, estos registros habían sido incluidos hacía menos de un mes desde que se realizó la búsqueda (también el 1/12/2017). El análisis de la tipología de estos documentos reveló que el 70% de ellos (703) eran tesis, el 29% (286) artículos de revista cuyos textos están alojados en *Dialnet*, y el 1% libros, también alojados en *Dialnet*.

De estos datos se desprende que este fenómeno parece afectar sobre todo a los documentos (artículos, libros, o tesis) sobre los que *Dialnet* no tiene conocimiento de la existencia de una versión en abierto (alojada en sus propios servidores, o en otras partes), que son la gran mayoría de los documentos indizados por *Dialnet*.

Es evidente que por alguna razón, *Google Scholar* ya no cubre esta fuente adecuadamente, aunque *Google* sí lo haga. No sabemos a ciencia cierta por qué ha podido pasar esto. Hemos revisado los permisos de *crawling* declarados en el archivo *robots.txt* de *Dialnet*, y lo hemos comparado con el *robots.txt* que tenían en mayo de 2016 gracias a la información disponible en *la Wayback Machine* de *Internet Archive* (figura 3), que fue aproximadamente un mes antes de que se recogieran los datos para la edición 2016 de *Google Scholar Metrics* (la última edición en la que se registró una amplia cobertura de revistas españolas).

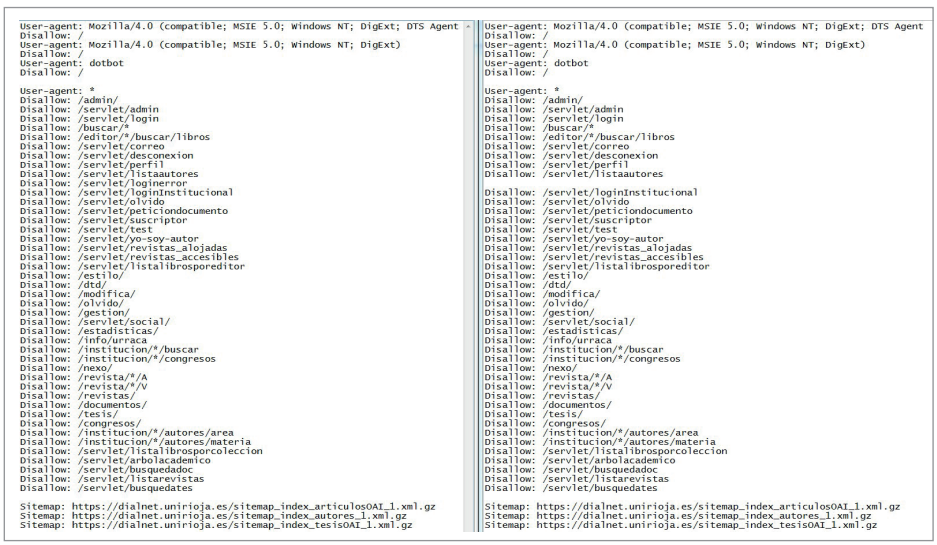
https://web.archive.org/web/20171129161225/https:/dialnet.unirioja.es/robots.txt

https://web.archive.org/web/20160515011218/https:/dialnet.unirioja.es/robots.txt

Los dos archivos son esencialmente los mismos, y por lo que podemos ver, ninguno de ellos impide a *Google Scholar* indexar las partes del sitio web que contienen los registros de los documentos.

Una causa del problema puede ser que, aunque no hayan modificado el *robots.txt*, sí se haya modificado la estructura interna de la web de forma que los robots de *Google Scholar* ya no sepan interpretarla, y por tanto se haya considerado que la información que antes estaba disponible ya no lo está. Como consecuencia, *Google Scholar* ya no muestra esa información en el buscador (pues sólo le interesa mostrar información relativa a documentos que están disponibles en la web).

El problema no es ya que muchas de las revistas españolas que anteriormente eran visibles a los robots de *Google Scholar* ahora han dejado de serlo total o parcialmente, sino que buena parte de los documentos españoles cubiertos por *Dialnet* (libros, capítulos de libros, tesis, etc…) pudieran haberse quedado fuera de circulación sin posibilidad de ser buscados y encontrados en la herramienta más empleada hoy por los científicos en el mundo para buscar información.



```
User-agent: Mozilla/4.0 (compatible; MSIE 5.0; Windows NT; DigExt; DTS Agent
Disallow: /
User-agent: Mozilla/4.0 (compatible; MSIE 5.0; Windows NT; DigExt)
Disallow: /
User-agent: dotbot
Disallow: /

User-agent: *
Disallow: /admin/
Disallow: /servlet/admin
Disallow: /servlet/login
Disallow: /buscar/*
Disallow: /editor/*/buscar/libros
Disallow: /servlet/correo
Disallow: /servlet/desconexion
Disallow: /servlet/perfil
Disallow: /servlet/listaautores
Disallow: /servlet/loginerror
Disallow: /servlet/loginInstitucional
Disallow: /servlet/olvido
Disallow: /servlet/peticiondocumento
Disallow: /servlet/suscriptor
Disallow: /servlet/test
Disallow: /servlet/yo-soy-autor
Disallow: /servlet/revistas_alojadas
Disallow: /servlet/revistas_accesibles
Disallow: /servlet/listalibrosporeditor
Disallow: /estilo/
Disallow: /dtd/
Disallow: /modifica/
Disallow: /olvido/
Disallow: /gestion/
Disallow: /servlet/social/
Disallow: /estadisticas/
Disallow: /info/urraca
Disallow: /institucion/*/buscar
Disallow: /institucion/*/congresos
Disallow: /nexo/
Disallow: /revista/*/A
Disallow: /revista/*/V
Disallow: /revistas/
Disallow: /documentos/
Disallow: /tesis/
Disallow: /congresos/
Disallow: /institucion/*/autores/area
Disallow: /institucion/*/autores/materia
Disallow: /servlet/listalibrosporcoleccion
Disallow: /servlet/arbolacademico
Disallow: /servlet/busquedadoc
Disallow: /servlet/listarevistas
Disallow: /servlet/busquedates

Sitemap: https://dialnet.unirioja.es/sitemap_index_articulosOAI_1.xml.gz
Sitemap: https://dialnet.unirioja.es/sitemap_index_autores_1.xml.gz
Sitemap: https://dialnet.unirioja.es/sitemap_index_tesisOAI_1.xml.gz
```

```
User-agent: Mozilla/4.0 (compatible; MSIE 5.0; Windows NT; DigExt; DTS Agent
Disallow: /
User-agent: Mozilla/4.0 (compatible; MSIE 5.0; Windows NT; DigExt)
Disallow: /
User-agent: dotbot
Disallow: /

User-agent: *
Disallow: /admin/
Disallow: /servlet/admin
Disallow: /servlet/login
Disallow: /buscar/*
Disallow: /editor/*/buscar/libros
Disallow: /servlet/correo
Disallow: /servlet/desconexion
Disallow: /servlet/perfil
Disallow: /servlet/listaautores

Disallow: /servlet/loginInstitucional
Disallow: /servlet/olvido
Disallow: /servlet/peticiondocumento
Disallow: /servlet/suscriptor
Disallow: /servlet/test
Disallow: /servlet/yo-soy-autor
Disallow: /servlet/revistas_alojadas
Disallow: /servlet/revistas_accesibles
Disallow: /servlet/listalibrosporeditor
Disallow: /estilo/
Disallow: /dtd/
Disallow: /modifica/
Disallow: /olvido/
Disallow: /gestion/
Disallow: /servlet/social/
Disallow: /estadisticas/
Disallow: /info/urraca
Disallow: /institucion/*/buscar
Disallow: /institucion/*/congresos
Disallow: /nexo/
Disallow: /revista/*/A
Disallow: /revista/*/V
Disallow: /revistas/
Disallow: /documentos/
Disallow: /tesis/
Disallow: /congresos/
Disallow: /institucion/*/autores/area
Disallow: /institucion/*/autores/materia
Disallow: /servlet/listalibrosporcoleccion
Disallow: /servlet/arbolacademico
Disallow: /servlet/busquedadoc
Disallow: /servlet/listarevistas
Disallow: /servlet/busquedates

Sitemap: https://dialnet.unirioja.es/sitemap_index_articulosOAI_1.xml.gz
Sitemap: https://dialnet.unirioja.es/sitemap_index_autores_1.xml.gz
Sitemap: https://dialnet.unirioja.es/sitemap_index_tesisOAI_1.xml.gz
```

Figura 3. Archivos *robots.txt* de *Dialnet* del 15 de mayo de 2016 (izquierda) y del 29 de noviembre de 2017 (derecha)

El 21 de noviembre escribimos de nuevo a los ingenieros de *Google Scholar* solicitando explicaciones para este raro fenómeno. La respuesta no ha llegado, todavía… Sólo ellos pueden aclararlo.

En cualquier caso este incidente viene a recordarnos dos cosas:

- El carácter guadianesco de un *search engine*. El sino de un buscador es la inestabilidad. Es tan cambiante como la web que quiere representar.
- La dependencia que las revistas y editoriales pueden tener de los intermediarios en el proceso de transferencia de información (agregadores, recolectores, bibliotecas digitales…). Si el espejo se rompe ningún contenido será visible o accesible. La única forma de resolverlo es que los sitios web presentaran arquitecturas transparentes cumpliendo los estándares y los requisitos técnicos que promueven la transferencia de la información (**Orduña-Malea**; **Delgado-López-Cózar**, 2014; **Orduña-Malea** *et al.*, 2016).

Si este problema persiste no cabe duda que perjudicará enormemente la visibilidad de la ciencia publicada en España.

Referencias

Delgado-López-Cózar, Emilio; **Orduña-Malea, Enrique**; **Martín-Martín, Alberto**; **Ayllón, Juan-Manuel** (2017). "*Google Scholar*: The big data bibliographic tool". In: Cantu-Ortiz, Francisco J. (ed.). *Research analytics : boosting university productivity and competitiveness through scientometrics* (pp. 59–80). Boca Raton, FL: CRC Press. ISBN: 978 1498785426

Martín-Martín, Alberto; **Delgado-López-Cózar, Emilio** (2017). *Dialnet blackout in Google Scholar (data and code)*.
https://osf.io/bxjn7

Orduña-Malea, Enrique; **Delgado-López-Cózar, Emilio** (2014). "Low visibility of Latin American repositories in Google Scholar: technical incompatibility or lack of web strategy?". *LSE Research online*, 31 Jul.
http://eprints.lse.ac.uk/71410/

Orduña-Malea, Enrique; **Martín-Martín, Alberto**; **Ayllón, Juan-Manuel**; **Delgado-López-Cózar, Emilio** (2016). *La revolución Google Scholar : Destapando la caja de Pandora académica*. Granada: Universidad de Granada. ISBN: 978 84 338 5941 9

Emilio Delgado López-Cózar
Universidad de Granada
Facultad de Comunicación y Documentación
edelgado@ugr.es

Alberto Martín-Martín
Universidad de Granada
Facultad de Comunicación y Documentación
albertomartin@ugr.es

* * *

Información preocupante

Miguel Navas-Fernández

Esta información me parece de gran interés:

"… creemos haber descubierto la fuente del problema. Este se llama *Dialnet*. La desaparición súbita de *Dialnet* de *Google Scholar* ha implicado directamente la invisibilidad de todas estas revistas españolas que venían siendo indizadas en *GSM* (*Google Scholar Metrics*)".

Quizás resulte de interés seguir debatiendo sobre este problema reportado por Emilio Delgado-López-Cózar y Alberto Martín-Martín.

Aunque *Google Scholar* no lo es todo y ni *Dialnet* ni ningún actor del sistema científico español es responsable del problema, la última frase del mensaje resulta bastante preocupante:

> "Si este problema persiste no cabe duda que perjudicará enormemente la visibilidad de la ciencia publicada en España".

Museo de Ciències Naturals de Barcelona
Centre de Documentació
mnavas@bcn.cat

* * *

El problema está en vías de solución

Emilio Delgado-López-Cózar

Hola Miguel, empiezo por el final: mensaje de tranquilidad: el problema está en vías de solución. La respuesta que esperábamos de *Google Scholar* llegó a última hora del día 5:

> "Estamos trabajando con los colegas de *Dialnet* para encontrar una solución adecuada. Ellos han realizado un ajuste en su sistema y actualmente estamos volviendo a rastrear las páginas actualizadas. La indexación de los artículos alojados en *Dialnet* debería volver gradualmente a ser como antes".

Ciertamente *Google Scholar* no lo es todo, pero según demuestran todos los estudios empíricos sobre hábitos de búsqueda de los científicos, *GS* es la principal puerta de entrada a la búsqueda de información, y la herramienta más usada en estas tareas en casi todos los campos científicos, especialmente en las Ciencias Sociales y Humanidades que, como bien sabes, son los agujeros

negros de los sistemas de información científica.

Y sobre la afirmación "…ni *Dialnet* ni ningún actor del sistema científico español es responsable del problema…", no estoy en condiciones de corroborarla. En nuestro documento arbitramos varias hipótesis: sólo los protagonistas (*Google scholar* y *Dialnet*) pueden aclarar qué ha ocurrido realmente.

La moraleja de toda esta historia tiene que ver con la pregunta que formulábamos hace tres años en un post en el blog de la *London School of Economics*: *Low visibility of Latin American repositories in Google Scholar: technical incompatibility or lack of web strategy?* (**Orduña-Malea**; **Delgado-López-Cózar**, 2014a).

"Si los repositorios no están bien construidos los "search engine" no harán visibles sus contenidos y eso condenará al ostracismo a las publicaciones científicas y a los que las generan y a los almacenes que las guardan y difunden"

En ese texto, referido a una investigación que habíamos realizado sobre repositorios latinoamericanos ("The dark side of open access in Google and Google Scholar: the case of Latin-American repositories"), planteábamos crudamente el problema: si los repositorios no están bien construidos los "search engine" (sea *Google*, *Google Scholar* o *Bing*, *Microsoft Academic*, *Semantic Scholar*…) no harán visibles sus contenidos y eso condenará al ostracismo a las publicaciones científicas y a los que las generan (autores e instituciones en las que trabajan) y a los almacenes que las guardan y difunden (repositorios). No hay peor castigo para la libre circulación del conocimiento.

Orduña-Malea, Enrique; **Delgado-López-Cózar, Emilio** (2014). "Low visibility of Latin American repositories in Google Scholar: technical incompatibility or lack of web strategy?". *LSE Research online*, 31 Jul.
http://eprints.lse.ac.uk/71410/

Orduña-Malea, Enrique; **Delgado-López-Cózar, Emilio** (2014). "The dark side of open access in Google and Google Scholar: The case of Latin-American repositories". *Scientometrics*, v. 102, n. 1, pp. 829-846.
https://arxiv.org/abs/1406.4331
https://doi.org/10.1007/s11192-014-1369-5

edelgado@ugr.es

* * *

¿El problema ha tenido incidencia también sobre los datos de citación?

Luis Rodríguez-Yunta

Hay que agradecer a Emilio y al grupo *EC3* el continuo seguimiento que mantienen sobre la evolución de *Google Scholar* como fuente, señalando sus virtudes y sus defectos. Sin duda, los datos de *GSM* siempre debían tomarse con reservas, tanto antes como ahora.

Imagino que la caída de fuentes en *Google Scholar* además de afectar al número de revistas, debería haber afectado al conjunto de citas que reciben las revistas que sí están. ¿Habéis apreciado igualmente una caída de datos en la citación? O si en este aspecto se percibe menor bajada ¿cómo debe interpretarse?

CCHS-CSIC
luis.ryunta@cchs.csic.es

* * *

No se aprecia una caída relevante de datos de citación

Emilio Delgado-López-Cózar

Luis, interesantes reflexiones y preguntas. *Google Scholar Metrics* es un producto que a decir de sus creadores tenía la función prioritaria de que los científicos fueran capaces de identificar cuáles son las revistas más influyentes en una disciplina o sobre un tema. Esta es la base en la que se asientan todas sus limitaciones de búsqueda (ofrecer listas de 100 revistas más citadas en un período de cinco años o sólo 20 títulos por consulta). Y de ahí nuestro empeño a la hora de construir nuestros índices: aflorar lo oculto, a pesar de lo que tenían previsto sus diseñadores.

Lo interesante del producto para la ciencia en lengua no anglosajona y realizada en países periféricos, por encima de los problemas de precisión en las mediciones, es que es capaz de exhibir revistas que quedaban totalmente ocultas en los sistemas de información científica tradicionales (léase *WoS*, *Scopus*, etc…), mostrando a su vez los artículos con más repercusión.

Y ya contestando a tus preguntas directamente, te diré que:

"¿Habéis apreciado igualmente una caída de datos en la citación?"

Sin haber realizado un análisis estadístico exhaustivo, parece que los indicadores bibliomé-

tricos de las revistas españolas que figuran este año en *GSM* no se han resentido. Por dos razones:

- por el período de cálculo empleado por *GSM*: la utilización de series quinquenales con variación de sólo el último año (Ej.: 2011-2015, 2012-2016), propicia la estabilidad en los indicadores;
- el hecho de que una revista no aparezca en *GSM*, no quiere decir que los títulos y pdfs de sus artículos hayan desaparecido totalmente de los almacenes de *Google Scholar*. Y estos son la base para los recuentos de citas. Es bien conocido, por otra parte, que la mayoría de los documentos poseen varias versiones, producto de diversas localizaciones.

Ya se sabe que los caminos que llevan a Roma son diversos y, a veces, como ocurre con los documentos de *Google Scholar*, insondables. Y afortunadamente que ocurra esto en *Google Scholar* es extraordinario. Porque si falla un espejo (como ha podido ocurrir ahora con *Dialnet*) siempre puede haber uno alternativo. Desde este punto de vista podemos decir que los daños se minimizan.

edelgado@ugr.es

* * *

Adaptación de *Dialnet* para solucionar el problema

Eduardo Bergasa

Desde *Dialnet* estamos al tanto de este tema desde junio. *Google Scholar* hizo cambios en sus criterios de indexación lo que provocó una importante disminución del contenido de *Dialnet* en su índice.

Acordamos con *GS* unos cambios en la plataforma para evitarlo y aunque por nuestra parte los implementamos en la primera semana de julio, estamos sometidos al próximo ciclo de refresco del índice que, ellos afirman, será a finales de diciembre o enero.

Fundación Dialnet. Universidad de la Rioja
eduardo.bergasa@unirioja.es

* * *

Saber dónde ha estado el problema ayudará a mejorar a todos

Emilio Delgado-López-Cózar

Hola Eduardo, ante todo alegrarme de que el tema esté en vías de solución. Me gustaría aprovechar tu mensaje para pedirte que nos expliques cuales han sido esos "cambios en la plataforma" que acordasteis con *Google Scholar* y ya habéis implementado. Estoy seguro que pueden ayudar a otros responsables de repositorios que pudieran tropezarse con problemas parecidos. Si es un problema de arquitectura o de metadata sería muy bueno saberlo.

Lo que me extraña es que la causa de la disminución de registros de *Dialnet* obedezca a que "*Google Scholar* hizo cambios en sus criterios de indexación". Nuestra primera hipótesis cuando detectamos el problema fue esta… Y así se lo comunicamos a *Google Scholar*. Transcribo el mensaje remitido:

On Mon, Nov 20, 2017 at 3:02 AM, Emilio Delgado-López-Cózar <*edelgado@ugr.es*> wrote:

"Querido Anurag, (…) Le escribimos porque hemos notado un fenómeno confuso sobre *Scholar Metrics*. Como sabe, desde que se lanzó *Scholar Metrics* por primera vez, nuestro grupo ha reunido y publicado una lista de todas las revistas editadas en España (todas las que pudimos encontrar) con cada nueva edición del producto. (…) Queremos preguntarle si ha habido algún cambio en la forma de calcular *Scholar Metrics*, o si este cambio sólo ha sido causado porque algunas revistas españolas ya no cumplen con sus criterios de inclusión".

Su contestación fue rápida y muy clara:

017-11-21 03:50, Anurag Acharya escribió:

"Querido Emilio: (…) No ha habido cambios en la indexación de *Scholar* o en cómo se calcula *Scholar Metrics*. Si hubiera habido algún cambio en el cálculo de *Scholar Metrics*, habríamos actualizado las páginas de ayuda".

En fin ¿dónde ha estado el problema? Creo que la respuesta a esta pregunta ayudará a mejorar a todos y a entender la relación entre los rastreadores de *Google Scholar* y los repositorios, entre el barco y los puertos del conocimiento… Perdona la metáfora pero creo que viene al caso.

edelgado@ugr.es

* * *

Aclaración sobre la política de indexación de Google

Eduardo Bergasa

No entré en detalle por no aburrir a los lectores de *IweTel* y no desviar la atención. Quise sobre todo frenar esa sensación de alarma y transmitir que el tema ya está solucionado. Por eso te ofrecía hacer las aclaraciones técnicas que necesitaseis fuera de la lista.

En cualquier caso, intentaré responder a tus dudas.

Google Scholar introdujo nuevas comprobaciones en los metadatos y en ellas detectaron que unos pocos de nuestros artículos no respetaban el orden de los autores de la fuente original. Esto es debido a que antes de 2011 se hizo alguna carga de revistas en inglés. Los metadatos de estas revistas, que llevan siglos sin actualizarse, no tienen los altos niveles de calidad y revisión que se exige a la información que introducen las universidades actualmente.

Google detectó el problema al comparar las versiones de estas referencias con las encontradas en otras bases de datos, y en lugar de únicamente ignorar este contenido en su índice, decidieron eliminar casi todo *Dialnet*. Nosotros lo descubrimos porque notamos un bajón enorme del tráfico de usuarios recibido de GS.

Rápidamente hicimos ajustes para que ignorasen esos artículos en lengua inglesa que no les gustaban, pero una vez excluidos nos hacen esperar un ciclo completo de reindexación.

eduardo.bergasa@unirioja.es

* * *

Los repositorios deben cuidar la calidad de los metadatos

Emilio Delgado-López-Cózar

Gracias por las explicaciones. Son sintéticas y muy reveladoras. Deduzco de ellas importantes mensajes para nuestra comunidad profesional:

- *Google Scholar*, acusado con razón de que está plagado de errores (para mí pequeños dada la magnitud de su empresa: indizar todo, en todos los lugares y lenguas y en todos los formatos), parece que tiene procedimientos de chequeo de la calidad de los metadatos que engulle. Lo que resulta sorprendente es que dejen de indizar todo *Dialnet* (un gigante informativo) por la detección de errores en el campo autor de un puñado de artículos.
- Los repositorios deben cuidar en extremo la calidad de sus metadatos, no sólo por su compromiso con disponer de información bibliográfica calidad, sino por sus posibles consecuencias en la deficiente indización de sus productos.

En fin, gracias por la información. Seguro que nos ayuda a todos a mejorar.

edelgado@ugr.es

* * *

A vueltas con el apagón digital de la producción científica en *Google Schoolar*

Emilio Delgado-López-Cózar

19 de diciembre

Navegando por la Red, me he topado, a mi pesar, otra vez con el problema del que alertábamos estos días. En este caso el apagón digital afecta al repositorio de mi universidad. En este caso más que apagón hay un eclipse total: ningún registro indizado en *Google Scholar* a pesar de que *Google* informa de que contiene 46.900 items.

Aunque ya sabemos que el comando *site* no es fiable para realizar la comprobación, es preciso verificar documento a documento, sí nos permite saber si algo raro pasa. Así lo hicimos en el caso de *Dialnet* y, efectivamente, detectamos el problema. He realizado un sondeo de urgencia sobre otros repositorios según podéis comprobar en la siguiente tabla. Los resultados son variables, pero prácticamente todos son del mismo tenor.

Repositorios	*Google*	*Google Schoolar*
site:digibug.ugr.es	46.900	0
site:digitum.um.es	43.600	3.030
site:eprints.ucm.es	69.700	8.570
site:zaguan.unizar.es	103.000	846
site:idus.us.es	686.000	40.500
site:minerva.usc.es	44.500	11.300
site:digibuo.uniovi.es	34.100	15.300
site:roderic.uv.es	63.400	29.100
site:uvadoc.uva.es	55.300	22.400
site:repositorio.uam.es	57.000	23.600
site:gredos.usal.es	79.000	50.500

Os pido a todos los bibliotecarios que os ocupéis de estas tareas que comprobéis si está ocurriendo algo parecido con vuestros repositorios... y lo comentéis, para que lo pongamos en conocimiento de los responsables de *Google Scholar*.

Este eclipse en *Google Scholar* estaría afectando a la accesibilidad de muchos documentos almacenados en nuestros repositorios. Aunque, afortunadamente siempre podremos decir que ahí está *Google*... en la casa madre los

documentos sí pueden ser localizados, accedidos y descargados sin problemas... También minimizarán el contratiempo aquellos que tengan a buen recaudo sus materiales además en otros pósitos… ¿Este es el problema de poner todos los huevos en la misma cesta? Una nueva llamada de atención…

edelgado@ugr.es

* * *

Universidad de Alcalá de Henares

Juana Frías-Fernández
19 de diciembre

Efectivamente, en nuestro caso ocurre lo mismo:

Repositorios	*Google*	*Google Schoolar*
site:ebuah.uah.es	64.500	518

juana.frias@UAH.ES

* * *

Universidad Católica de Chile

Javiera Bravo
19 de diciembre

En nuestro caso: país Chile
Repositorio de la Universidad Católica de Chile

Repositorios	*Google*	*Google Schoolar*
site:repositorio.uc.cl	75.100	10.100

ibravoc@UC.CL

* * *

No se deben comparar los mecanismos de indización de *Google* y *Google Scholar*

Jordi Prats-Prat
19 de diciembre

Al margen del problema de indización del repositorio de la *Universidad de Granada*, que espero que se pueda solucionar pronto, me ha sorprendido tu mensaje.

¿De verdad estás comparando los mecanismos de indización de *Google* con los de *Google Scholar*? No soy experto en la materia, pero me atrevería a decir que son muy distintos.

No lo veo nada claro y quizás habría que hilar más fino. Pero siguiendo los mismos parámetros que indicas (y también con un sondeo de urgencia) debo indicar que, en el caso de otras plataformas, como puede ser *Researchgate*, el porcentaje de indización en *Google Scholar* respecto al buscador de *Google* creo que es aproximadamente del 3%, muy inferior a los datos que facilitas de algunos de los repositorios institucionales de las universidades que mencionas. No lo considero significativo por las dudas que me sugiere el indicador, es sólo para seguir tu argumentación.

Quizás se trate de políticas de *Google Scholar* y podríamos considerar que, en este sentido, se está realizando un muy buen trabajo global por parte de las universidades en la gestión y evolución de sus repositorios institucionales, en comparación con otros servicios. Sin dormirnos, hay que seguir trabajando.

Por otra parte, considerar que los repositorios institucionales recogen sólo la producción de investigación de sus instituciones (la que considera *Google Scholar*) tampoco me parece acertado. Lo veo un poco superado.

La producción académica de una universidad abarca muchos otros ámbitos, como pueden ser las colecciones patrimoniales, institucionales, docentes. Se trata de optimizar y sacar rendimiento a las infraestructuras disponibles y hay muy buenos ejemplos. No son menores. Posiblemente estas se recojan en *Google,* pero no en *Google Scholar*. Ha habido una evolución importante en la idea de repositorio institucional que cabría contemplar.

Que los repositorios institucionales sean visibles en *Google Scholar* me parece básico, pero los datos que presentas no me parecen correctos.

Y si podemos ayudar a los compañeros de la *Universidad de Granada* en facilitar su indización en *Scholar,* que cuenten con ello.

jordi.prats@UPC.EDU

* * *

Universidad Politécnica de Madrid

José-Ignacio González-González
20 de diciembre

En la *Universidad Politécnica de Madrid* estuvimos comparando la indexación de nuestro repositorio en *Google* y en *Google Scholar* durante todo el 2015 y siempre dio resultados muy distintos.
Yo no soy tampoco experto, pero como dice Jordi, creo que no tiene nada que ver la forma de indexar de *Google* con la de *Google Scholar*.
Estos son los datos:

site:oa.upm.es	*Google*	*Google Schoolar*
03/02/2015	155.200	14.200
16/04/2015	164.000	15.400
11/05/2015	184.000	15.600
03/09/2015	169.000	16.700
21/10/2015	165.000	17.200
08/01/2016	307.000	18.700
20/12/2017	239.000	27.000

joseignacio.gonzalez@UPM.ES

* * *

Universidad de Huelva

José-Carlos Morillo-Moreno
20 de diciembre

En la *Universidad de Huelva* también estamos teniendo problemas en la indexación por parte de *Google Scholar*, de hecho, un técnico de *Google Scholar* contactó con nosotros hace unos días para informarnos que tenían problemas en la indexación de nuestro repositorio *Arias Montano* y algunas páginas web de la institución. Estamos intentando resolverlo pero aún sin éxito.

¿Han contactado también desde *Google Scholar* con otros repositorios por problemas de indexación?, no sé si relacionar el problema específico surgido con nosotros con el problema general que se está relatando en esta conversación.

jcarlos.morillo@BIBLIO.UHU.ES

* * *

Universidad de Granada

María-Ángeles García-Gil
20 de diciembre

En primer lugar, hoy aparecemos indizados en *Google Scholar*, con 21.800 resultados. Teniendo en cuenta que tenemos 38.000 documentos aproximadamente en *Digibug*, la proporción de indexación es excelente, ya que según *Google Scholar* los archivos que exceden de 5 Mb se consideran libros, y por tanto, no aparecen reflejados.
https://scholar.google.es/intl/es/scholar/inclusion.html#content

Llevamos un tiempo constatando que tenemos un problema en los resultados en la búsqueda de *site*, debido a motivos estructurales de los sistemas de suministro eléctrico, habiendo sufrido apagones puntuales del servidor.

De todas formas echamos de menos en *Google Scholar* un sistema de validación, como los existentes en *Recolecta*, *OpenAire*,…

Aprovechamos la ocasión para animaros a que visitéis *Digibug*, por el contenido de la producción científica del PDI de la *Universidad de Granada*, por el nivel de los trabajos académicos de nuestros estudiantes, por la importancia de los fondos culturales y la calidad en la descripción de nuestros contenidos.
http://digibug.ugr.es

digibug@UGR.ES

* * *

Mal negocio si *Google Scholar* no reúne todos los contenidos académicos

Miguel Navas-Fernández
20 de diciembre

Sea como fuere, si uno va a *Google Scholar* a buscar un contenido académico, y luego resulta que éste no se encuentra allí sino en *Google*, mal negocio, ¿no?

¿Se sabe si los repositorios anglosajones sufren el mismo problema?

mnavas@BCN.CAT

* * *

Hay que saber interpretar los resultados

Emilio Delgado-López-Cózar
20 de diciembre

Gracias a todos por los mensajes remitidos que sirven para que aprendamos más sobre lo que es *Google*, *Google Scholar* y el comando *site*. Los hechos hasta ahora son:

1. El repositorio de la *UGR* estuvo literalmente apagado. Yo utilicé metafóricamente el término "apagón", pero resulta que efectivamente era un apagón eléctrico el que había ocultado al repositorio en *Google*. Tuve mala suerte: navegué a oscuras y me di un trompazo. Estupendo que haya vuelto la luz.

2. En Huelva, el repositorio ha tenido problemas de indización y hasta los ingenieros de *Google Scholar* se han puesto en contacto con sus responsables. Estupenda y sorprendente noticia… No sabía que los de Google hacían eso…

3. En la UPM venían haciendo un seguimiento comparando los items incluidos en *Google* y *Google Scholar*. Me parece loable por el interés que demuestra en seguir el grado de visibilidad del sitio web en los dos buscadores. Ahora bien,

como señalaremos a continuación los datos deben ser correctamente interpretados.

Independientemente de estos hechos, Jordi lanzó un mensaje planteándose los fundamentos de los datos ofrecidos, su fiabilidad y validez. Muchas gracias por tu mensaje que me sirve de pretexto para ofrecer explicaciones metodológicas y detalles técnicos sobre *Google*, *Google Scholar* y el comando *site*, sin los cuales las interpretaciones de los datos y de los mensajes que puedan elaborarse al respecto podrían ser torcidas. Ahora hilaré más fino como reclamabas acertadamente en tu mensaje.

Antes de ello permíteme que me remita al capítulo 4 del libro *La revolución Google Scholar: destapando la caja de pandora académica*, titulado "Capturando la web académica: funcionamiento general" (**Martín-Martín** *et al.*, 2016) donde damos abundante información de cómo funciona *Google Scholar*, centrando especialmente el tiro en su forma de indizar la Web y los documentos académicos. El abundante espacio de un libro permite contar cosas que no es posible hacer en el tiránico estrecho margen de un tweet o correo electrónico. Alguien como yo, autor de un libro y de otros trabajos del mismo calado, difícilmente puede confundir los mecanismos de indización de *Google* con los de *Google Scholar*… Y desde ya proclamo, nada tienen que ver entre sí y, evidentemente, ello debe tenerse en cuenta para interpretar correctamente los datos como bien señalas en parte de tu mensaje y ahora yo remarcaré y ampliaré con otros puntos a considerar que tú no has contemplado.

Dicho esto, y antes de proseguir, debo indicar que los problemas de indización de *Dialnet* y de algunos repositorios no son una entelequia. Con toda la cautela que pude afirmé:

> "Aunque ya sabemos que el comando site no es fiable para realizar la comprobación, es preciso verificar documento a documento, sí nos permite saber si algo raro pasa".

Me reitero en la aseveración. Por tanto, objetivo más que cumplido: mi mensaje de alerta iba en ese sentido… Si además generamos debate y afloramos información que ayude a comprender los datos, mejor que mejor. Y paso ya a ofrecer explicaciones que permitan interpretar adecuadamente los datos de comparación.

Cuando se hace una consulta con el comando *site* deben contemplarse los siguientes extremos:

- Nos devuelve todos los "ítems" que encuentra en el dominio especificado (digibug.ugr.es). Items no es igual a documentos sino a documentos y páginas web. Para comprobarlo pido a todos los lectores de la lista que repliquéis esta consulta site:digibug.ugr.es en *Google*. Observaréis que los primeros enlaces corresponden a las páginas generales del repositorio (inicial, ayuda…). En cambio *Google scholar* encuentra solo documentos, salvo error, que siempre puede ocurrir. Por consiguiente, no deben compararse literalmente sin más cifras de *Google* con *Google Scholar*…
- En *Google Scholar* el comando *site* nos devuelve sólo la versión principal de los documentos (la que *Google Scholar* ha tomado como entrada primaria), no contemplando las versiones que figuren anidadas al registro y que podrían ser documentos alojados en el repositorio. Por tanto, es muy posible que los resultados de la búsqueda estén infravalorando la tasa de indización de documentos. Y es por esta razón por lo que ya advertimos en nuestro primer mensaje sobre *Dialnet* y repetí en este que

  > "el comando *site* no es fiable para realizar la comprobación, es preciso verificar documento a documento".

  Ahora bien,

  > "sí es válido para obtener una idea aproximada del volumen de datos indizados por un sitio web".

- *Google* indiza todas las páginas web y los "documentos", entendiendo este concepto en sentido amplio; lo que sus robots son capaces de capturar en la web. Aspira a todo, pero obviamente no lo tiene todo. Me remito a los ya antiguos trabajos de Gilles al respecto. En principio no pone restricciones técnicas. En cambio la política de selección de *Google Scholar* es distinta: pretende indizar los documentos que cuelgan de la web académica (universidades, editoriales, bibliotecas, etc…). Impone una serie de condiciones a los documentos a indizar. Entre otros destaco los que más pueden afectar a los documentos incluidos en los repositorios:

  Documentos menores de 5Mb
  Documentos que contengan resúmenes

Por consiguiente, hay en los repositorios documentos (sobre todo tesis doctorales y otros trabajos académicos) que no van a ser indizadas en *Google Scholar*. Y eso es fácil comprobarlo.

Respecto a los repositorios universitarios españoles o de otros países, ya sabemos, como muy bien se ha señalado en varios de los mensajes, que en ellos figuran documentos de diversa naturaleza: académica, científica, cultural, institucional. Y así debe ser, yo soy partidario de que el repositorio sea un fiel reflejo de la institución a la que representa y sirve. El problema es que esto no ocurre y depende de las políticas y prácticas institucionales que rigen en cada organización. Por tanto, la búsqueda en *Google* contiene todos

los documentos albergados en el repositorio mientras que la realizada en *Google Scholar* se circunscribe fundamentalmente a los académicos y científicos. Ergo, no se deben comparar los datos sin tener presente esta circunstancia

Y, por último, ¡ojo!: siempre hay que hablar de datos aproximados. Si *Google* y *Google Scholar* ante una búsqueda nos responde que hay "Aproximadamente…" no vamos a ser nosotros los que hablemos de datos exactos... Entonces ¿qué es lo que debe hacerse para tener una idea aproximada del grado de visibilidad de los documentos en *Google Scholar*?:

- Tomar como referencia para la comparación siempre los documentos académicos (tesis, TFG, TFM…) o científicos indizados en el repositorio.
- Elegir una muestra representativa de documentos del repositorio y buscarlos *en Google Scholar* y *Google*. Estos datos son los realmente representativos de la visibilidad.
- Seguir cotejando con el comando *site* los "ítems" que figuran en *Google* y *Google Scholar*, como hace José-Ignacio en la *UPM*, para saber "si pasan cosas raras" y, además, comprobar si va creciendo nuestra indización en *Google Scholar*. Yo, interpretando los datos de la *UPM,* lo que colijo es que el número de documentos indizados de la *UPM* en el período de año (2015-2016), y tomados como entrada principal en *Google Scholar*, es de 4.500 documentos. Pues sólo resta cotejar, comprobar e interpretar acertadamente.

En definitiva, no es que los datos que suministré sean incorrectos, lo que es incorrecta es la interpretación que se hace o se puede hacer de los mismos. Por eso me parece muy oportuna la intervención de Jordi y mi respuesta para aclararlos. A partir de ahí, espero que ahora nadie pueda decir lo que yo no he dicho ni pretendido decir… y lo más importante que todos sepamos usar los comandos de búsqueda e interpretar adecuadamente los resultados que proporcionan.

Sobre lo que comentas de *researchgate*, el tema es de mayor calado… requeriría un mensaje aparte. Y este es ya demasiado largo.

Martín-Martín, Alberto; **Delgado-López-Cózar, Emilio**; **Orduña-Malea, Enrique**; **Ayllón, Juan M.** (2016). *La revolución de Google Scholar. Destapando la caja de pandora académica*. Granada: Universidad de Granada. ISBN: 978 84 338 5985 3
http://www.unebook.es/es/ebook/la-revolucion-google-scholar_E0002614676

edelgado@ugr.es

* * *

Hay que seguir cuidando la indización de nuestros repositorios en *Google Scholar*

Jordi Prats-Prat

20 diciembre

Creo que era importante aclarar algunas dudas que quizás hubieran llevado a alguna confusión.

Diría que a día de hoy es esencial para los repositorios institucionales cuidar su indización en *Google Scholar* y por las respuestas a tu mensaje se está haciendo. Posiblemente se trata de uno de los servicios de búsqueda de documentación científica más importantes a los que se puede acceder actualmente (aparte de otros servicios que ofrece, como páginas personales, institucionales, métricas…).

Hay que cuidar la buena presencia de los repositorios en esta plataforma para garantizar al PDI de la universidad que, si depositan sus trabajos en los repositorios institucionales, estos van a estar disponibles en *Google Scholar* (y en otros servicios). También por el fuerte impacto que tiene en los accesos y visibilidad de los repositorios institucionales y sus contenidos. Destacaría en este sentido la presencia en *Google Scholar* del amplio abanico que se considera literatura gris, más consultada posiblemente de lo que pueda parecer y en la que los repositorios institucionales están haciendo un buen trabajo de difusión.

Sobre la referencia al libro *La revolución Google Scholar: destapando la caja de pandora académica* no sólo recomendaría la lectura del capítulo 4, sino el libro en su totalidad. Se trata de un buen texto para introducirse en la plataforma, sus políticas, sus criterios, como explotarla… Muy entendedor para conocer el potencial de *Google Scholar*.

No es la primera vez que oigo por compañeros de otras universidades que *Google Scholar* cuida a sus proveedores. Les comunica cuándo hay problemas de indización, indicando también (creo) los problemas identificados. Si es así es de agradecer, tratándose de un servicio que debe indizar millares de fuentes. En nuestro caso no hemos detectado problemas de indización del repositorio institucional de la UPC, crucemos los dedos, pero sí que contactamos con ellos cuando estábamos implementando la nueva versión del repositorio para resolver dudas y las respuestas fueron muy buenas.

Sobre los posibles problemas con repositorios anglosajones, no tengo información muy precisa. Sí que la indización de los contenidos de los repositorios institucionales en *Google Scholar* era uno de los indicadores que se contemplaban en el ranking de repositorios de *Webometrics* (a día de hoy cancelado). Por lo que recuerdo, en este indicador aparecían repositorios no anglosajones bien posicionados, por lo que deduzco que la indización no debe depender del país de origen del repositorio.
http://repositories.webometrics.info

jordi.prats@UPC.EDU