
ESTUDIOS / RESEARCH STUDIES

Análisis de la implementación Schema.org en el repositorio RODERIC e impacto en el posicionamiento en Google y Google Scholar

Nuria Nevado-Chiné*, Rubén Alcaraz-Martínez*, José Ángel Navalón**

*Departament de Biblioteconomia, Documentació i Comunicació Audiovisual.
Universitat de Barcelona

Correo-e: nnevadoc@gmail.com ORCID iD: <https://orcid.org/0000-0002-2595-8747>

Correo-e: ralcaraz@ub.edu ORCID iD: <https://orcid.org/0000-0002-2595-8747>

**Servei d'Informàtica. Universitat de València

Correo-e: Jose.A.Navalon@uv.es ORCID iD: <https://orcid.org/0000-0003-3707-8787>

Recibido: 07-06-20; 2ª versión: 07-08-20; Aceptado: 10-08-20. Publicado: 01/07/2021

Cómo citar este artículo/Citation: Nevado-Chiné, N.; Alcaraz-Martínez, R.; Navalón, J. A. (2021) Análisis de la implementación Schema.org en el repositorio RODERIC e impacto en el posicionamiento en Google y Google. *Revista Española de Documentación Científica*, 44 (3), e300. <https://doi.org/10.3989/redc.2021.3.1797>

Resumen: Se presenta la implementación de Schema.org en el repositorio RODERIC de la Universitat de València. Para el análisis del impacto de la implementación se han definido ocho indicadores que se han analizado en Google Search y Google Scholar según el caso: visitas, visitas a registros bibliográficos, documentos descargados, impresiones, clics, CTR, posición media en la SERP y posición en la SERP que fueron analizados durante dos períodos consecutivos de un año, antes y después de la implementación. Los resultados obtenidos muestran resultados desiguales para ambos buscadores. En el caso de Google Search, a pesar de conseguirse un incremento considerable en el número de impresiones (21,05%), tanto los clics (10,38%), como el número de sesiones (15,03%) descienden. En el caso de Google Scholar, las sesiones se incrementan ligeramente (6,25%). El número de registros visualizados y de descargas de documentos del repositorio mejora en un 16,21% y 12,18%, respectivamente.

Palabras clave: Schema.org; repositorios; posicionamiento web; SEO académico; web semántica, JSON-LD, RODERIC

Analysis of the Schema.org implementation in the RODERIC repository and impact on search engine optimization in Google and Google Scholar

Abstract: The implementation of Schema.org in the RODERIC repository of the Universitat de València is presented. For the analysis of the impact of the implementation, eight indicators have been defined that have been analyzed in Google Search and Google Scholar according to the case: visits, visits to bibliographic records, downloaded documents, impressions, clicks, CTR, average position in the SERP and position in the SERP; they were analyzed for two consecutive one-year periods, before and after implementation. The results obtained show uneven results for both search engines. In the case of Google Search, despite achieving a considerable increase in the number of impressions (21.05%), both clicks (10.38%) and the number of sessions (15.03%) decreased. In the case of Google Scholar, the sessions increased slightly (6.25%). The number of records viewed and document downloads from the repository improved by 16.21% and 12.18%, respectively.

Keywords: Schema.org; repositories; search engine optimization; academic SEO, semantic web, JSON-LD, RODERIC

Copyright: © 2021 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

1. INTRODUCCIÓN

De acuerdo con la consultora MOZ (2020), el SEO (Search Engine Optimization) es la práctica encaminada al aumento de la cantidad y calidad de tráfico hacia un sitio web a través de los resultados de búsqueda orgánicos procedentes de los buscadores de Internet.

El objetivo del SEO es, por lo tanto, conseguir aumentar el tráfico hacia un sitio web, logrando que sus páginas sean más relevantes que las de sus competidores para determinadas palabras clave utilizadas por los usuarios de los motores de búsqueda. El trabajo SEO se centra en satisfacer ciertos factores de posicionamiento que tienen en cuenta los motores de búsqueda mediante todo un conjunto de estrategias y aspectos técnicos con el objetivo de mejorar su visibilidad en las páginas de resultados de los motores de búsqueda (Injante y Mauricio, 2019). Estos factores se pueden clasificar en dos grandes grupos, según si se trata de aspectos que se pueden acometer trabajando el contenido o la configuración técnica del sitio web (factores internos o *SEO on page*), o si se trata de factores que, siendo ajenos a éste, presentan una influencia en el posicionamiento (factores externos o *SEO off page*) (Serrano-Cobos, 2015). Si bien Google nunca ha dado a conocer abiertamente cuáles son todos los factores involucrados en el cálculo de la relevancia utilizado para ordenar los resultados de búsqueda, justificando su falta de transparencia como una medida necesaria para la lucha contra el *spam* (Beel y Gipp, 2010), lo cierto es que algunos de sus más de 200 factores (Google, 2020a) han trascendido fruto del trabajo de profesionales del sector, como resultado de procesos de ingeniería inversa (Rovira y otros, 2018).

La dura competencia por estar bien posicionado en los motores de búsqueda académicos ha propiciado que, desde hace algunos años, las técnicas SEO hayan traspasado las fronteras de los sitios web y los motores de búsqueda generalistas para empezar a aplicarse también en el contexto de buscadores como Google Scholar (GS) (Martín-Martín y otros, 2017). Si bien algunos de los factores de posicionamiento mantienen una cierta correspondencia entre los dos tipos de buscadores, GS cuenta con un algoritmo propio adaptado al ámbito específico de la literatura científica (Rovira y otros, 2018). Al conjunto de estrategias y técnicas asociadas con la mejora de la visibilidad de los sitios web que albergan literatura científica se le conoce con el nombre de ASEO (Academic SEO) (Muñoz-Martín, 2015) (Codina, 2019). Su objetivo principal, como en el caso del SEO, es atraer tráfico a estos sitios web, aumentando así la probabilidad

de que sus documentos sean encontrados, posibilitando así también a sus autores conseguir una mayor visibilidad, que se puede acabar traduciendo en un incremento posterior en el número de citas y otras métricas personales (Farhadi y otros, 2013).

Entre los factores internos que han despertado un mayor interés entre los profesionales del SEO en los últimos años, destaca el uso de Schema.org. También en el ámbito de la biblioteconomía y documentación. Prueba de ello es el peso que ha adquirido año tras año en la encuesta sobre datos enlazados de OCLC (2018), cuyos resultados muestran incrementos desde el 2014 del 30% y el 46% de los proyectos que lo utilizan de acuerdo con el análisis de Smith (2018) para los años 2015 y 2018, respectivamente. También en el caso de los repositorios de datos de investigación, ámbito en el que Fenner y otros (2018) identificaron un total de 32 repositorios que ya lo habían implementado para, en un trabajo más reciente, proponer una hoja de ruta que incluye el uso de esta tecnología (Fenner y otros, 2019).

Schema.org es un vocabulario que, a través de diversos esquemas, propone una serie de atributos para la descripción normalizada de diferentes tipos de entidades presentes en un sitio web (personas, artículos, eventos o recetas, entre otras) (Sulé, 2015). El proyecto, iniciativa de los principales motores de búsqueda de Internet (Google, Bing, Yahoo! y Yandex) tiene como principal objetivo facilitar la comprensión del contenido presente en una página web, posibilitando a los motores de búsqueda ofrecer tanto unos resultados más relevantes, como nuevos servicios enriquecidos como el Google Knowledge Graph, o visualizaciones más informativas en sus páginas de resultados, conocidas como *rich snippets* (Hilliker y otros, 2013). Aplicado al ámbito que nos ocupa, Schema.org cuenta con esquemas específicos para la descripción de artículos científicos, tesis, libros o trabajos de fin de máster, entre otros.

El repositorio RODERIC (Repositorio de Objetos Digitales para la Enseñanza, la Investigación y la Cultura) fue lanzado en 2010 con el objetivo reunir, preservar y visibilizar la producción científica de la Universitat de València. RODERIC funciona como una ventana única desde la cual da acceso a materiales de diferente índole como la producción científica de la institución, el fondo antiguo procedente de su Biblioteca Histórica (Somni), materiales docentes, contenidos de carácter institucional como los derivados de la celebración de actos y reuniones oficiales, así como otros documentos derivados de la acción cultural de la universidad. Cumplidos los diez años de existencia, el principal reto al que

RODERIC se enfrenta en el contexto actual es el de competir con garantías con el resto de los repositorios nacionales e internacionales en las páginas de resultados de los motores de búsqueda. Es por este motivo que, en 2018, RODERIC se plantea la implementación de Schema.org como estrategia para mejorar sus métricas de posicionamiento.

Este trabajo muestra el proceso de implementación del vocabulario Schema.org, así como de otros ajustes enmarcados dentro de los considerados factores internos, realizados en el repositorio RODERIC de la Universitat de València. También recoge un análisis del impacto de esa implementación pasado un año en los resultados de búsqueda de Google Search y GS. La implementación de Schema.org en el repositorio persigue el objetivo general de incrementar la cantidad y la calidad del tráfico a sus documentos tanto desde motores de búsqueda generalistas, como desde otros especializados en literatura científica.

2. JUSTIFICACIÓN

El acceso abierto tiene como uno de sus principales objetivos facilitar el acceso a los resultados de la investigación financiada con fondos públicos. Este movimiento depende del consentimiento del autor o titular de los derechos de autor, pero también, en gran medida, de Internet y las tecnologías que la sustentan (Suber, 2010), entre las cuales, los buscadores.

En los últimos años, diversos estudios han analizado el comportamiento de los investigadores a la hora de buscar información poniendo como eje central de sus investigaciones qué fuentes utilizan. Brown y Swan (2007) revelan un escaso uso por parte de los investigadores de los repositorios de sus instituciones como fuente de acceso a la información científica, siendo sólo el 2% de los encuestados los que lo utilizan con frecuencia, mientras que un 4% afirman hacerlo ocasionalmente. Por otro lado, más del 90% de los investigadores encuestados afirman utilizar Google Search como fuente para localizar literatura científica en acceso abierto; más del 70% en el caso de GS. Cifras muy por encima de otros servicios especializados como repositorios institucionales (algo más del 30%), BioMed Central (casi el 30%), sitios web de bibliotecas públicas (20%) o DOAJ (alrededor del 15%), por citar algunos de los más destacados. Inger y Gardner (2013) redundan en el mismo resultado, identificando a los motores de búsqueda académicos como GS, como el principal punto de partida para la localización de artículos de revista. También destacan una mayor popularidad de GS en áreas como las ciencias sociales, la psicología y la educa-

ción, mientras que, en el resto de las áreas, Google Search es más popular. El estudio de Arshad y Ameen (2017) sobre la forma de acceso a revistas electrónicas también muestra un uso preferente de buscadores generalistas y GS, frente a bases de datos especializadas o los servicios suscritos por sus bibliotecas de referencia, entre otros. Los resultados del estudio de Tenopir y otros (2019) muestran que un 28,6% de los artículos consultados por investigadores se descubren mediante el uso de buscadores. El 56,6% de estos artículos se encuentran gracias a buscadores como Google Search y GS, un porcentaje que supera con creces a otros servicios como Academic Search Premier, Web of Science, JSTOR, o a servicios de preprints como arXiv.org y otros. Los resultados de trabajos anteriores ponen de evidencia la importancia de aumentar la visibilidad de los repositorios institucionales y de la literatura científica en motores de búsqueda como Google Search y GS.

3. TRABAJOS RELACIONADOS

Son varios los trabajos que han reunido y analizado diferentes aspectos que deben abordarse para conseguir una mayor visibilidad de los repositorios en los motores de búsqueda (Arlitsch y O'Brien, 2011) (Kelly y Nixon, 2013) (Tmava y Alemneh, 2013) (Tonkin y otros, 2013) (Hilliker y otros, 2013) (Sexton y Aery, 2013) (Wallis y otros, 2017). También desde el punto de vista del trabajo que atañe a los propios investigadores en las diferentes fases de la elaboración y posterior diseminación de un trabajo científico, como parte esencial del SEO académico (Beel y Gipp, 2010) (Muñoz-Martín, 2015) (Codina, 2019). Por su parte, Google también ha venido publicando tanto guías de optimización SEO de sitios web en Google Search (Google, 2020b), como pautas específicas para asegurar el rastreo e indexación de repositorios y otras aplicaciones de software para la gestión de publicaciones científicas en GS (Google, 2020c). Si a estos trabajos le sumamos otros orientados a la mejora del posicionamiento de cualquier tipo de sitio web en buscadores (Killoran, 2013) (Su y otros, 2014), así como el trabajo divulgativo de compañías especializadas en SEO (MOZ, 2015) (Searchmetrics, 2018), podemos afirmar que en la actualidad contamos con un conjunto de referencias suficientes como para abordar esta tarea.

El principal déficit de los trabajos anteriores es que no evalúan el impacto en el algoritmo de los factores que recogen. Entre los estudios que sí evalúan el rendimiento de estos factores destacan los trabajos de Beel y Gipp (2009a) (2009b), Martín-Martín y otros (2014) (2016) (2017), Rovira y otros (2018), Pekala (2018) y Macgregor (2019).

De ellos, los únicos que se centran en el uso de Schema.org son Pekala (2018) y Macgregor (2019), mientras que el resto de los trabajos se centran en otros aspectos relevantes para el ASEO enmarcados dentro de los denominados factores externos.

El trabajo de Beel y Gipp (2009a) muestra una correlación directa entre el número de citas recibidas por un artículo y su posición en el ranking. Por otro lado, los mismos autores (2009b) analizan el impacto de la antigüedad de los artículos como factor de posicionamiento. Si bien, sus resultados muestran que los documentos más antiguos se encuentran con una mayor frecuencia en mejores posiciones, esto puede deberse al hecho de que también se trate de trabajos con un mayor número de citas acumuladas. Una conclusión similar a la de Martín-Martín y otros (2016). Por su parte, Martín-Martín y otros (2014) analizan la correlación existente entre el número de versiones de un documento en GS y el número de citas recibidas, en relación con la posición en el ranking. Sus resultados no muestran una correspondencia entre el número de versiones del documento y la posición, aunque sí detectan un ligero efecto positivo en este sentido en un trabajo posterior (Martín-Martín, 2017). Por lo que respecta a la relación entre citas y posicionamiento, los datos obtenidos por los autores permiten confirmar que el recuento de citas es el factor de posicionamiento más importante en GS. En la misma línea, Rovira y otros (2018) también analizan el impacto de las citas recibidas como factor de posicionamiento en GS. Sus resultados también muestran que la cantidad de citas recibidas por un artículo es el factor SEO externo más importante, muy por encima del resto.

Por lo que respecta a los trabajos que analizan el impacto de la implementación de Schema.org, Pekala (2018) muestra los resultados del uso de esta tecnología en el repositorio DigitalGeorgetown. Los resultados de su análisis muestran un impacto pequeño, pero visible en los resultados de búsqueda de Google Search que se traduce en un 26,67% de resultados que mejoraron en el ranking, un 18,33% que empeoraron y un 55% que no mostraron ningún cambio. Por otro lado, la tasa de crecimiento del tráfico de referencia desde esta fuente aumentó en un 63,3% en comparación a la misma métrica en los últimos 3 años que se situaba en torno a una media del 23,27%. Otra aportación de interés de este trabajo es un mapeo que contempla equivalencias entre los elementos Dublin Core (DC) utilizados en el repositorio DigitalGeorgetown y los atributos de Schema.org equivalentes.

Por su parte, Macgregor (2019) recoge los resultados de ciertas mejoras realizadas sobre Stra-

thprints, el repositorio institucional de la Universidad de Strathclyde. Los resultados muestran un aumento del 63% en el tráfico de referencias desde el conjunto de todos los buscadores, incrementándose significativamente en los casos de Google Search (64,63%) y GS (94,63%), mientras que descendió en el caso del buscador Bing. Por su parte, también se observaron incrementos en la cantidad de impresiones (145%) y clics (53%) obtenidos por los resultados de búsqueda del repositorio en Google Search para el periodo analizado.

Tanto Pekala (2018), como Macgregor (2019), coinciden en reconocer límites evidentes en sus metodologías de evaluación. Uno de los límites más importantes tiene que ver con la imposibilidad de aislar el impacto de la implementación de Schema.org del resto de más de 200 factores que Google dice tener en cuenta en su algoritmo para determinar la relevancia de una página web y ordenar los resultados de búsqueda. Por otro lado, en los resultados obtenidos por Macgregor (2019), debe tenerse en cuenta que la implementación de Schema.org se acompaña de otras mejoras técnicas que también pueden haber tenido un impacto positivo en los resultados, algunas de las cuales se han recogido en la tabla I junto con el resto de los factores contemplados en la literatura analizada en este trabajo.

Por su parte, otros profesionales del sector (MOZ, 2019) (Price, 2019) destacan la inexistencia de evidencias concluyentes sobre una hipotética mejora en el ranking tras la implementación de Schema.org. No obstante, sí existe un cierto consenso en destacar el valor de los fragmentos enriquecidos en la SERP generados con estos vocabularios para la mejora de métricas como el CTR (Click-Through Rate) (Goodsell, 2015), lo que puede derivar en una mayor cantidad de tráfico.

Finalmente, también cabe destacar la importancia de otros factores que, si bien no se encuentran exclusivamente relacionados con los repositorios, ni con GS, si son factores con un peso importante en el algoritmo de Google Search. El primero de estos factores consiste en disponer de un sitio adaptado a dispositivos móviles, un requisito imprescindible desde que Google anunciará la actualización *mobile-first index* de su algoritmo (Fan, 2018). Hasta la llegada de esta actualización, Google contaba con dos rastreadores diferenciados (Googlebot y Googlebot mobile), para rastrear el contenido de las versiones de escritorio y móvil de los sitios web, respectivamente, dando lugar a dos índices diferenciados. La puesta en marcha del *mobile-first index* implica el fin del rastreo con Googlebot y la unificación de los dos índices en uno

Tabla I. Mejoras recogidas en la literatura y las directrices de Google para aumentar la visibilidad de un sitio web en los motores de búsqueda

ID	Mejora	Tipo de factor	Buscador	Fuente
1	Utilizar esquemas de metadatos estandarizados.	Factor interno	Google Scholar	Tmava y Alemneh (2013)
2	Exponer los datos bibliográficos como metaetiquetas siguiendo alguno de los siguientes estándares: Highwire Press tags, Eprints tags, BE Press tags o PRISM tags.	Factor interno	Google Scholar	Arlitsch y O'Brien (2011); Google (2020c)
3	Usar preferentemente ficheros en formato HTML o PDF, con texto indexable y con tamaños que no superen los 5 MB.	Factor interno	Google Scholar	Muñoz-Martín, B. (2015); Google (2020c)
4	Incluir en la primera página el título del documento con un tamaño de fuente mayor e inmediatamente a continuación el nombre de los autores. Evitar el uso de portadas generalistas sin esta información.	Factor interno	Google Scholar	Tonkin y otros (2013); Google (2020c)
5	Optimización de título (entre 6 y 16 palabras), resumen y palabras clave. Las palabras clave a posicionar deben aparecer en estos campos, así como en el inicio del artículo, a lo largo del cuerpo con una densidad suficiente, en los títulos de tablas y gráficos y en las discusiones y resultados. Si el documento se ofrece en formato HTML se deben marcar las secciones con encabezados <h1>... <h6>	Factor interno	Google Search Google Scholar	Beel; Gipp y Wilde (2010); Killoran (2013) Su y otros (2014); MOZ (2015); Muñoz-Martín (2015); Codina (2019); Google (2020b)
6	Informar los campos de metadatos disponibles en los ficheros PDF.	Factor interno	Google Scholar	Beel; Gipp y Wilde (2010); Muñoz-Martín, B. (2015)
7	Incluir la palabra clave en el nombre del fichero.	Factor interno	Google Scholar	Beel; Gipp y Wilde (2010); Muñoz-Martín, B. (2015)
8	Implementar el repositorio con alguna de las siguientes aplicaciones de software: Eprints, Digital Commons o DSpace. Utilizar Open Journal Systems en el caso de publicaciones en serie.	Factor interno	Google Scholar	Google (2020c)
9	Implementación de Schema.org	Factor interno	Google Search Google Scholar	Hilliker y otros (2013); Sexton y Aery (2013); MOZ (2015); Wallis y otros (2017); Searchmetrics (2018); Pekala (2018); Macgregor (2019); Google (2020b)
10	Dotar al repositorio de una arquitectura de la información que permita a los rastreadores acceder a los documentos desde la página de inicio, mediante enlaces a un índice de autores, materias, colecciones...	Factor interno	Google Scholar Google Search	Arlitsch y O'Brien (2011); Google (2020b); Google (2020c)
11	Identificar cada documento con un URL único.	Factor interno	Google Scholar	Arlitsch y O'Brien (2011); Google (2020c)
12	Configuración de ficheros sitemap en formato XML.	Factor interno	Google Search Google Scholar	Sexton y Aery (2013); Google (2020b)
13	Incluir las palabras clave en el URL. Evitar la creación de URL demasiado largos.	Factor interno	Google Search	Killoran (2013); MOZ (2015); Su y otros (2014)
14	Implementación del protocolo HTTPS.	Factor interno	Google Search	MOZ (2015); Searchmetrics (2018); Google (2020b)
15	Optimizar la velocidad de carga.	Factor interno	Google Search	MOZ (2015); Google (2020b)
16	Utilizar un tamaño de texto igual o superior a los 16px.	Factor interno	Google Search	Google (2020b)

ID	Mejora	Tipo de factor	Buscador	Fuente
17	Separar suficientemente los elementos en los que se puede hacer clic	Factor interno	Google Search	Google (2020b)
18	Definir la ventana gráfica (viewport)	Factor interno	Google Search	Google (2020b)
19	Evitar el scroll horizontal (contenido más ancho que la pantalla)	Factor interno	Google Search	Google (2020b)
20	Evitar el uso de complementos incompatibles u obsoletos como Flash.	Factor interno	Google Search	Google (2020b)
21	Incluir la palabra clave a posicionar en el atributo alt de las imágenes	Factor interno	Google Search	Google (2020b)
22	Contar con un sitio web adaptado a dispositivos móviles, preferentemente responsivo.	Factor interno	Google Search	Far (2012); Alcaraz (2016); Google (2020b)
23	Proveer al repositorio de enlaces entrantes desde fuentes como blogs y sitios de referencia con gran autoridad como Wikipedia o redes sociales y plataformas como YouTube. Preferentemente, los enlaces que apunten al repositorio contendrán en su texto ancla la palabra clave para la cual se desea posicionar.	Factor externo	Google Search Google Scholar	Kelly y Nixon (2013); Killoran (2013); Su y otros (2014); MOZ (2015); Codina (2019); Google (2020b)
24	Aumentar el número de citas recibidas por los documentos.	Factor externo	Google Scholar	Beel y Gipp (2009); Martín-Martín y otros (2014); Rovira y otros (2018)
25	Conseguir un CTR adecuado en la SERP.	Factor externo	Google Search	Killoran (2013)
26	Mantener un porcentaje de rebote bajo.	Factor externo	Google Search	Killoran (2013)

sólo, el móvil. La preferencia de Google por los sitios web responsivos frente a otras aproximaciones técnicas para el cumplimiento de este requisito es sobradamente conocida desde 2012 (Far, 2012). Esta aproximación presenta otros beneficios desde el punto de vista del SEO, frente al desarrollo de versiones móviles independientes. Entre estos destaca una autoridad de dominio que se mantiene para ambas versiones, una total equidad en los enlaces entrantes recibidos, ya que todos apuntan al mismo dominio, así como un mantenimiento más sencillo y barato, al sólo contar con una única versión del código fuente (Alcaraz, 2016).

Estrechamente relacionado con la adaptación a dispositivos móviles, la velocidad de carga se ha convertido en uno de los factores internos más importantes para los sitios web, en un contexto en el que la mayoría ya reciben más visitas desde dispositivos móviles, que desde equipos de escritorio, tal y como recogen los datos globales de Statcounter para 2019, en los que el tráfico móvil (teléfonos y tabletas) supone un 56,73% del total, frente al 43,27% correspondiente al tráfico desde equipos de escritorio.¹ Desde 2010, Google viene anunciando diferentes actualizaciones del algoritmo en las que esta señal se ha incorporado o reforzado (Singhal y Cutts, 2010) (Wang y Phan 2018).

El último de estos factores tiene que ver con la usabilidad del sitio web. En concreto con un pe-

queño conjunto de métricas que Google evalúa y sobre las que informa en Google Search Console: tamaño de texto demasiado pequeño para leerlo (menos de 16 px), elementos en los que se puede hacer clic demasiado cerca unos de otros, no configurar la ventana gráfica (*viewport*), y el uso de complementos como Flash, incompatibles (Google, 2020d).

4. METODOLOGÍA

Con el objetivo de analizar el impacto de la implementación de Schema.org en el repositorio RORDERIC, se estableció una metodología formada por 4 fases: definición de métricas de visibilidad; análisis de los metadatos utilizados en el repositorio y mapeo con sus esquemas correspondientes; implementación del vocabulario schema.org y análisis del impacto de la implementación en el posicionamiento en Google Search y GS.

4.1. Definición de métricas de visibilidad

De acuerdo con los trabajos de Espadas y otros (2008) y Caro y otros (2010) es posible evaluar la visibilidad de un sitio web a partir de métricas que se centran en tres aspectos clave: visitas, rendimiento en motores de búsqueda y enlaces entrantes. Todas ellas determinan en su conjunto la capacidad de un sitio web para ser encontrado en la Web. Por su parte, Aguillo (2009) propone diferen-

tes indicadores aplicables al análisis de repositorios. En el caso que nos ocupa, nos interesan especialmente los indicadores de uso, para los cuales el autor presenta una taxonomía organizada en tres grupos: popularidad, comportamiento y descargas.

En la tabla II, se recogen los indicadores utilizados para medir el impacto de Schema.org en el repositorio RODERIC obtenidos a partir del análisis de la literatura anterior, así como las herramientas de analítica con las que se obtuvieron los datos. Los indicadores propuestos se clasifican en 3 grupos: popularidad, rendimiento y descargas.

4.2. Análisis de los metadatos utilizados en RODERIC y mapeo con Schema.org

En esta fase se llevó a cabo un mapeo entre los elementos Dublin core (DC) utilizados en RODERIC y las propiedades Schema.org equivalentes. Más allá de los 15 elementos básicos DC, más los 40 calificadores existentes, RODERIC implementa 58 calificadores adicionales propios para la descripción de algunas colecciones especiales, lo que supuso un total de 113 propiedades a mapear. Por su parte, RODERIC contempla 31 tipologías documentales diferentes para la descripción de los documentos que se ingestan en el repositorio. El análisis se acotó a las tres tipologías documentales más representativas (artículos de revista, libros y tesis en ese orden). Para la formalización del mapeo se tuvieron en cuenta los trabajos previos de Mixter y otros (2014), Wallis y otros (2017), Feener y otros

(2019), así como la documentación técnica publicada por las bibliotecas de Hesburgh (2018).

4.3. Implementación del vocabulario Schema.org en el repositorio

De los diferentes estándares existentes para la implementación del vocabulario Schema.org en el repositorio, se optó por la codificación en formato JSON-LD (JavaScript Object Notation for Linked Data). La elección de este estándar frente al resto de posibilidades (microdatos o RDFa) se justifica por ser la sintaxis preferida y recomendada por Google en sus directrices para desarrolladores, así como por el hecho de no tener que intercalarlo con el texto visible del documento, pudiendo de esta manera expresar con mayor facilidad ciertos datos en el código fuente que quizá no interese mostrar en la interfaz. Finalmente, tal y como destacan Wallis y otros (2017) y Fenner y otros (2019), otro beneficio de JSON-LD es que se trata de la sintaxis más sencilla de implementar de todas. En la figura 1 se muestra un ejemplo de implementación de esta sintaxis para la descripción de una tesis disponible en el repositorio.

La interfaz XMLUI de DSpace utilizada por RODERIC para mostrar los datos del repositorio se basa en el uso de Apache Cocoon, un entorno de programación basado en la expresión de elementos XML y su posterior transformación a través de hojas de estilo XSLT (eXtensible Stylesheet Language Transformations).

Tabla II. Indicadores para la evaluación de la visibilidad del repositorio, descripción y herramienta de analítica utilizada para obtenerlo

Indicador	Tipo	Descripción	Herramienta
Visitas	Popularidad	Número total de visitas al repositorio (sesiones en la terminología de Google Analytics).	Google Analytics
Visitas a registros bibliográficos	Popularidad	Número total de visitas a registros bibliográficos del repositorio.	DSpace SOLR Statistics
Documentos descargados	Descargas	Número total de documentos PDF descargados.	DSpace SOLR Statistics
Impresiones	Rendimiento	Número de veces que se ha mostrado un resultado de RODERIC en la SERP Google Search.	Google Search Console
Clics	Rendimiento	Número de clics recibidos por los resultados de RODERIC mostrados en la SERP de Google Search.	Google Search Console
CTR (Click Through Rate)	Rendimiento	Número de clics recibidos por los resultados de RODERIC respecto al número de impresiones en la SERP de Google Search. Se calcula como: clics/impresiones x 100.	Google Search Console
Posición media en la SERP	Rendimiento	Posición media obtenida por los resultados de RODERIC en la página de resultados de Google Search.	Google Search Console
Posición en la SERP	Rendimiento	Posición obtenida por los resultados de RODERIC en la SERP cuando se realiza una búsqueda por una palabra clave determinada.	SERP de Google Search y Google Scholar

Figura 1. Ejemplo de implementación de la sintaxis JSON-LD de Schema.org para la descripción de una tesis de RODERIC

```

1  {
2    "@context":"https://schema.org",
3    "@id":"https://hdl.handle.net/10550/72548",
4    "@type":"thesis",
5    "name":"Anxiety and self-related constructs in learners who stutter in the
        learning of English as a foreign language",
6    "headline":"Anxiety and self-related constructs in learners who stutter in the
        learning of English as a foreign language",
7    "alternatename":"",
8    "author":[
9      {
10         "@type":"Person","name":"Miller, Ronan"
11      }],
12   "contributor":[
13     {
14       "@type": "Person",
15       "name":"García Pastor, María Dolores"
16     }],
17   "keywords":["stammering","foreign language teaching","foreign language
        learning","TEFL","SLA","foreign language anxiety","stuttering","self related
        constructs","stuttering","tartamudez","UNESCO::PEDAGOGÍA","UNESCO::CIENCIAS DE
        LAS ARTES Y LAS LETRAS"],
18   "description":"La Ansiedad en una Lengua Extranjera (ALE) se ha considerado uno
        de los factores más importantes que afectan el grado de éxito del aprendizaje de
        idiomas (Daubney, Dewaele y Goknou, 2017). Sin embargo, la investigación sobre
        ALE en Alumnos Con Tartamudez (ACT) es escasa. Esta tesis doctoral aborda este
        problema explorando la relación entre ALE y la tartamudez en el aprendizaje del
        inglés como lengua extranjera en el contexto español.",
19   "publisher":"",
20   "datePublished":"2020",
21   "citation":{"
22     "@id":"https://hdl.handle.net/10550/72548"
23   }
24 }

```

XSL (eXtensible Stylesheet Language) es un lenguaje que permite transformar datos que provienen de documentos XML en otros datos en formato XML, o en otro formato, como es el caso de JSON-LD. Posteriormente mediante la aplicación de CSS (Cascading Style Sheet) se les puede dar el aspecto final que veríamos por pantalla en forma de página web. Para conseguir incorporar los datos de Schema.org se implementaron varias reglas en XSLT que transforman los metadatos propios del objeto digital en un elemento <script> con las propiedades de Schema.org en formato JSON-LD, quedando insertado dinámicamente para cada documento en su cabecera.

El Servei d'Informàtica de la Universitat de València implementa en un entorno de pruebas el código anterior para, posteriormente, llevarlo a producción en el mes de abril de 2019. El código se revisa en dos ocasiones (12 de abril y 30 de mayo) con el objetivo de corregir algunos pequeños errores de sintaxis detectados. Una vez implementado el código en el entorno de producción, se solicita, a través de Search Console, la reindexación del repositorio a través del envío de un fichero sitemap en formato XML.

4.4. Análisis del impacto de la implementación de Schema.org

La implementación de Schema.org en el repositorio RODERIC se llevó a cabo en abril del 2019. Para la recogida de datos, se utilizaron las siguientes herramientas: Google Analytics, Google Search Control y DSpace SOLR Statistics. Los periodos de análisis comprenden del mes de abril de 2018 a marzo de 2019, y un segundo periodo que abarca del mes de abril de 2019 al mes de marzo de 2020.

Adicionalmente y, con el objetivo de poder determinar si los documentos del repositorio pasaron a obtener mejores posiciones en el ranking de resultados tras la implementación de Schema.org, se escogió una muestra de 20 documentos (10 artículos de revista, 5 libros antiguos y 5 tesis) (disponible en el Apéndice 1). Se realizó un total de tres búsquedas en Google Search y GS los días 01/04/2019, 01/06/2019 y 01/06/2020 para cada uno de los documentos, que se corresponden con el día anterior a la implementación, tras seis meses de la implementación y tras doce meses de la implementación, respectivamente.

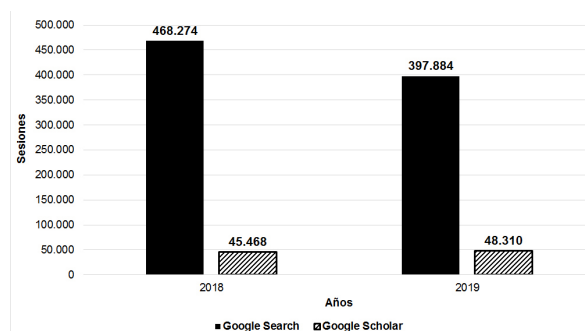
Las búsquedas consistieron en palabras clave extraídas de los títulos de los documentos, suficientemente específicas como para aparecer en la primera página de resultados, pero evitando reproducir exactamente los títulos con el objetivo de observar la competitividad de RODERIC con otras fuentes. Todas las búsquedas se realizaron con el navegador Chrome, en modo incógnito y sin iniciar sesión en ningún perfil de Google. Si el documento no se encontró en las primeras cinco páginas de resultados, no se contó. El número de páginas de resultados consultadas se limitaron a cinco por la limitación de resultados ofrecidos por el buscador en determinados casos para consultas tan específicas.

5. RESULTADOS

5.1. Indicadores de visibilidad del repositorio

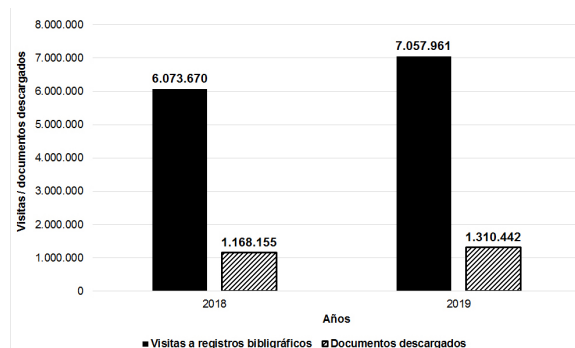
Las sesiones procedentes del tráfico derivado de Google Search descendieron en el periodo 2019-2020 un 15,03%, mientras que en el caso de GS se observó un ligero incremento de un 6,25%. Google Search es la fuente de referencia más importante para todo el periodo analizado, suponiendo un 65,25% y un 63,51% del total para los años 2018 y 2019, respectivamente. En el caso de GS, este buscador supone el 7,82% y el 7,39% del total para 2018 y 2019, lo que lo sitúa como la segunda fuente de tráfico orgánico más importante de RODERIC (figura 2).

Figura 2. Evolución de las sesiones procedentes desde Google Search y Google Scholar para los periodos de abril de 2018 a marzo de 2019 y de abril de 2019 a marzo de 2020.



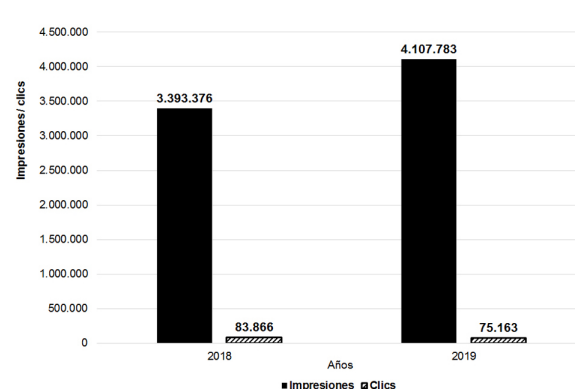
Por lo que respecta a las visitas recibidas a registros bibliográficos, se observa un aumento del 16,21%. Una tendencia al alza que también se da en el caso del número total de documentos descargados del repositorio que aumenta un 12,18% (figura 3).

Figura 3. Evolución de las visitas a registros bibliográficos y de los documentos descargados para los periodos de abril de 2018 a marzo de 2019 y de abril de 2019 a marzo de 2020.



Si bien las impresiones aumentan un 21,05% en Google Search en el periodo posterior a la implementación de Schema.org, con respecto a las obtenidas el año anterior, los clics descienden en un 10,38% (figura 4), implicando así también un descenso del CTR que pasa de un 2,69% de media en el primer periodo, a un 1,84% en el segundo. La posición media en la página de resultados de búsqueda también empeora, pasando de la posición 26 a la 32.

Figura 4. Evolución de las impresiones y clics obtenidos en Google Search para los periodos de abril de 2018 a marzo de 2019 y de abril de 2019 a marzo de 2020.



5.2. Resultados del análisis del ranking de resultados

Los resultados globales muestran un 50% de búsquedas que han acabado en mejores posiciones al final del periodo analizado, respecto a la búsqueda inicial antes de la implementación de Schema.org. En un 25% de los casos, los resul-

tados mantienen la misma posición en el ranking, y en el 25% restante de los casos pasan a ocupar peores posiciones. El motor de búsqueda en el que se aprecian mejores resultados globales es GS, con un 60% de búsquedas en mejores posiciones, un 25% que se mantienen y un 15% que descienden. En el caso de Google Search, el 40% de las búsquedas mejora en el ranking, el 25% se mantiene y el 35% restante desciende de posición.

En sólo 6 de los casos, el documento de RODE-RIC no se ha encontrado entre las cinco primeras páginas de resultados en ninguna de las tres búsquedas realizadas a lo largo del año analizado. Esta situación se da en 2 ocasiones en Google Search y 4 en GS. En sólo 1 ocasión, un resultado que no aparecía entre las primeras cinco páginas de Google Search en la búsqueda anterior a la implementación, pasó a posicionarse en la primera página de resultados (posición 3) para, en la última búsqueda, volver a caer a posiciones inferiores a las de la quinta página o desaparecer del índice. En 20 de las 40 búsquedas (12 en el caso de GS y 8 en el caso de Google Search) los resultados pasan a ocupar posiciones más elevadas en el ranking. En 6 ocasiones, la página de resultados muestra una URL de la versión móvil del repositorio, a pesar de realizarse las búsquedas desde un equipo de escritorio. En 3 de esas ocasiones, en la SERP conviven resultados de la versión de escritorio y la versión móvil del repositorio. La lista completa con los términos de búsqueda y los cambios en las posiciones del ranking se pueden consultar en el Apéndice 2.

6. LÍMITES

Uno de los grandes retos que plantea el análisis de la repercusión de la implementación de Schema.org como factor de posicionamiento en los motores de búsqueda es la imposibilidad de poder aislar este factor de los centenares de factores contemplados por los algoritmos de los buscadores, más aún cuando estos se actualizan constantemente. En este sentido, los resultados están condicionados por el resto de los factores contemplados en la literatura (tabla I), así como por otros no conocidos por los autores debido a la opacidad de los algoritmos.

Finalmente, tanto el volumen de la muestra, como el periodo analizado tras la implementación de Schema.org pueden no haber sido suficientes para estudiar el impacto de esta implementación. No cabe duda de que una mejora como ésta, necesita ser observada en un espectro temporal más amplio y sobre un mayor número de casos o

documentos analizados. En este sentido, Pekala (2018) apunta que los cambios en los resultados de búsqueda no son inmediatamente visibles, ni tampoco consistentes en el tiempo, aspecto este último que hemos podido comprobar con la fluctuación de la posición en el ranking de diversas búsquedas.

7. DISCUSIÓN

Aunque la implementación y el análisis de su impacto se han realizado en un repositorio implementado con la versión 4.1 de DSpace, técnicamente, la metodología expuesta en este trabajo es compatible con versiones más actuales del software. También se antoja posible su implementación en otras plataformas como EPrints, Fedora Commons o Invenio, por citar algunas de las más utilizadas en la actualidad dentro de la oferta de aplicaciones de software libre. La diferencia entre una y otra plataforma la encontramos en el uso del estándar XSLT para la salida web de los datos, que puede variar según la plataforma, por otros estándares. No obstante, y, aunque sería preciso un estudio más riguroso, los autores de este trabajo no conocemos ningún software para la creación de repositorios que en la actualidad incorpore en su núcleo la capacidad para exponer de manera nativa datos en formato Schema.org.

Los resultados obtenidos en este trabajo coinciden en parte con las afirmaciones de los profesionales del SEO que destacan la inexistencia de evidencias en la mejora del ranking derivadas del uso de Schema.org (MOZ, 2019) (Price, 2019). Adicionalmente, entre los factores con un impacto en el posicionamiento que también pueden haber influenciado en la obtención de peores resultados en relación con la cantidad de sesiones desde Google Search, así como en la posición media en la SERP, tras la implementación de Schema.org en el repositorio, destacan: la no utilización del protocolo HTTPS en el nombre de dominio; el hecho de contar con un sitio relativamente lento cuando se accede desde dispositivos móviles;² o el utilizar una versión específica del repositorio cuando se accede desde un dispositivo móvil (<http://mobiroderic.uv.es>), práctica que si bien en los primeros años de existencia de los teléfonos inteligentes se popularizó, en actualidad se ha visto superada en beneficios por el diseño web responsivo, una aproximación centrada en adaptar la interfaz de manera dinámica, sin necesidad de mantener dos sitios web diferentes en paralelo, cosa que puede perjudicar métricas esenciales como la cantidad de enlaces entrantes que se reparten entre el dominio principal y el subdominio en el que se encuentra la versión

móvil, o incluso en la generación de contenido duplicado, una práctica perseguida por los buscadores.

Por otro lado, la implementación por parte de Google del *mobile-first index*, centrado en la indexación de la versión móvil de los sitios web –aplicado sobre RODERIC desde septiembre de 2018– puede haber empeorado la situación antes descrita en el contexto de Google Search. A estos factores debemos sumarles otros factores externos, entre los que destaca el buen trabajo que haya podido realizar la competencia durante el periodo analizado, así como la cantidad de citas recibidas por los trabajos de otros repositorios que podrían ser superiores a las de los trabajos depositados en RODERIC, aspecto que no se ha analizado. En este sentido, parece claro que, en la actualidad, si bien la implementación de Schema.org podría suponer ciertos beneficios tanto a nivel de posicionamiento, como de interoperabilidad, parece que estos otros factores presentan un peso mayor en el algoritmo.

En relación con la capacidad de Schema.org para describir los objetos digitales de un repositorio, el estándar muestra una capacidad suficiente para describir con un alto nivel de precisión el contenido de tesis, libros y artículos de revista, si bien en el caso del libro antiguo, se encuentran a faltar diversos campos con un nivel de especificidad equivalente.

8. CONCLUSIONES Y TRABAJO FUTURO

Recoger, preservar y difundir la producción académica de una institución (Abadal, 2012) son los tres procesos fundamentales que dan sentido a un repositorio. En la actualidad, la difusión del contenido que albergan pasa ineludiblemente por gozar de un buen posicionamiento en los motores de búsqueda generalistas y especializados en literatura académica, dada la importancia de Google y Google Scholar en la recuperación y uso de la información académica (Orduña-Malea y Delgado, 2014). Una máxima que adquiere aún más sentido si atendemos a los resultados de la encuesta conducida por MOZ (Ray, 2019), en la que sólo el 7% de los usuarios dicen navegar más allá de la primera página de resultados, mientras que los dos primeros resultados de la primera página se llevan el 75% de los clics. En este sentido, la implementación de Schema.org, si bien minoritaria en este ámbito, ya es una prioridad en la agenda de organizaciones como OCLC (2018), y una realidad en repositorios de diferente índole como los analizados en los trabajos relacionados de este trabajo, así como en repositorios de da-

tos ampliamente utilizados como Zenodo o Data-Cite Search.

La implementación de Schema.org en RODERIC ha servido también para poner en evidencia diferentes déficits en materia de SEO que deben abordarse con urgencia para aumentar la visibilidad del repositorio en buscadores. En este sentido, una de las prioridades es la actualización de la plantilla de visualización, que deberá adaptarse a un diseño web responsivo, abandonando así la actual duplicidad de contenidos entre las dos versiones del repositorio. Una nueva plantilla que deberá satisfacer también aquellos requisitos de usabilidad móvil no alcanzados por la actual, así como ofrecer un tiempo de carga acorde con los requisitos de los buscadores. Todo ello, en vistas a una mejor adaptación al *mobile first index* de Google.

Por lo que respecta a la implementación de Schema.org, una vez consolidada la incorporación del código para un primer conjunto de tipologías documentales, en una segunda fase se espera abordar la implementación del resto de tipologías del repositorio.

Finalmente, en trabajos futuros también se prevé aumentar la muestra de búsquedas realizada, así como incluir un nuevo indicador de visibilidad en el índice con el objetivo de hacer seguimiento de la cantidad de páginas del repositorio disponibles en cada uno de los motores de búsqueda analizados.

9. AGRADECIMIENTOS

Los autores agradecen el apoyo de la Universitat de València y del Servei de Biblioteques i Documentació, así como de José Manuel Barrueco por sus aportaciones sobre el conocimiento y experiencia con RODERIC.

ACKNOWLEDGEMENTS

The authors acknowledge the support by the Universitat de València and the Servei de Biblioteques i Documentació as well as Jose Manuel Barrueco for their contributions on knowledge and experience with RODERIC.

10. NOTAS

¹ Datos disponibles en: <https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/worldwide/#-monthly-201901-201912>

² La puntuación obtenida por RODERIC en Google PageSpeed Insights, herramienta oficial de Google para evaluar la rapidez de una página web es de 61 sobre 100 para consultas móviles (consulta: 7/5/2020). Google recomienda alcanzar o superar una nota de 90 sobre 100.

11. REFERENCIAS

- Abadal, E. (2012). *Acceso abierto a la ciencia*. Barcelona: Editorial UOC.
- Aguillo, I. (2009). Measuring the institution's footprint in the web. *Library hi tech*, 27 (4), 540-556. <https://doi.org/10.1108/073788309>
- Alcaraz-Martínez, R. (2016). Webs i aplicacions mòbils: una aproximació des del punt de vista de l'experiència d'usuari. *Item: Revista De Biblioteconomia I Documentació*, 61, 21-42.
- Arlitsch, K.; O'Brien, P.S. (2011). Invisible institutional repositories addressing the low indexing ratios of IRs in Google Scholar. *Library Hi Tech*, 30 (1), 60-81. <https://doi.org/10.1108/07378831211213210>
- Arshad, A.; Ameen, K. (2017). Scholarly communication in the age of Google: exploring academics' use patterns of e-journals at the University of the Punjab. *The Electronic Library*, 35, 167-184. <https://doi-org.sire.ub.edu/10.1108/EL-09-2015-0171>
- Beel, J.; Gipp, B. (2009a). Google scholar's ranking algorithm: the impact of citation counts (an empirical study)". En: *Third international conference on research challenges in information science*, RCIS 2009, 439-446. <https://doi.org/10.1109/RCIS.2009.5089308>
- Beel, J.; Gipp, B. (2009b). Google scholar's ranking algorithm: the impact of articles' age (an empirical study). En: *Sixth international conference on information technology: new generations, ITNG'09*, 160-164. <https://doi.org/10.1109/ITNG.2009.317>
- Beel, J.; Gipp, B. (2010). Academic search engine spam and google scholar's resilience against it. *The Journal Of Electronic Publishing*, 13 (3), 1-28. <https://doi.org/10.3998/3336451.0013.305>
- Beel, J.; Gipp, B.; Wilde, E. (2010). Academic Search Engine Optimization (ASEO): optimizing scholarly literature for Google Scholar and Co. *Journal Of Scholarly Publishing*, 41 (2), 176-190. <https://doi.org/10.3138/jsp.41.2.176>
- Brown, S.; Swan, A. (2007). *Researchers' use of academic libraries and their services: a report commissioned by the Research Information Network and the Consortium of Research Libraries*. Disponible en: <https://eprints.soton.ac.uk/263868/> [Fecha de consulta: 26/03/2020].
- Caro, A.; Calero, C.; Moraga, M.A. (2010). Are web visibility and data quality related concepts? *IEEE Internet computing*, 15 (2), 43-49. <https://doi.org/10.1109/MIC.2010.126>
- Codina L. (2019). SEO académico: significación, componentes y fases. En: *EDICIC 2019*, Barcelona, p. 48. Barcelona: Universitat de Barcelona. <http://hdl.handle.net/10230/41996> [Fecha de consulta: 18/03/2020].
- Espadas, J.; Calero, C.; Piattini, M. (2008). Web site visibility evaluation. *Journal of the American Society for Information Science and Technology*, 59 (11), 1727-1742. <https://doi.org/10.1002/asi.20865>
- Fan Z. (2018). Rolling out mobile-first indexing. *Google webmaster central blog: official news on crawling and indexing sites for the Google index*. Disponible en: <https://webmasters.googleblog.com/2018/03/rolling-out-mobile-first-indexing.html> [Fecha de consulta: 25/03/2020].
- Far, P. (2012). Recommendations for building smart-phone optimized websites. *Google Webmaster Central Blog*. Disponible en: <https://webmasters.googleblog.com/2012/06/recommendations-for-building-smartphone.html> [Fecha de consulta: 26/03/2020].
- Farhadi, H.; Salehi, H.; Yunus, M.M.; Aghaei Chadegani, A.; Farhadi, M.; Fooladi, M.; Ale Ebrahim, N. (2013). Does it matter which citation tool is used to compare the h-index of a group of highly cited researchers?. *Australian Journal Of Basic And Applied Sciences*, 7 (4), 198-202. <https://ssrn.com/abstract=2259614>
- Fenner, M.; Crosas, M.; Durand, G.; Wimalaratne, S.; Gräf, F.; Hallett, R.; Bernal Llinares, M.; Schindler, U.; Clark, T. (2018). *Listing of data repositories that embed schema.org metadata in dataset landing pages*. <https://doi.org/10.5281/zenodo.1263942>
- Fenner, M.; Crosas, M.; Grethe, J.S.; Kennedy, D.; Hermjakob, H.; Rocca-Serra, P.; Durand, G.; Berjon, R.; Karcher, S.; Martone, M.; Clark, T. (2019). A data citation roadmap for scholarly data repositories. *Scientific Data*, 6 (28). <https://doi.org/10.1038/s41597-019-0031-8>
- Goodsell, Ben (2015). SEO for featured snippets leads to big gains. *Search engine land*. Disponible en: <https://searchengineland.com/seo-featured-snippets-leads-big-gains-236212> [Fecha de consulta: 06/08/2020].
- Google (2020a). ¿Cómo funciona la Búsqueda de Google? Disponible en: <https://developers.google.com/search/docs/beginner/how-search-works?hl=es>
- Google (2020b). *Guía de optimización en buscadores (SEO) para principiantes*. Disponible en: <https://developers.google.com/search/docs/beginner/seo-starter-guide?hl=es>
- Google (2020c). *Inclusion guidelines for webmasters*. Disponible en: <https://scholar.google.com/intl/en/scholar/inclusion.html> [Fecha de consulta: 19/03/2020].
- Google (2020d). *Informe "Usabilidad móvil"*. Disponible en: <https://support.google.com/webmasters/answer/9063469?hl=es> [Fecha de consulta: 19/03/2020].
- Hesburgh Libraries (2018). *Core elements: condensed view*. Disponible en: http://ndlib.github.io/metadata_application_profile/elements/ [Fecha de consulta: 30/03/2020].
- Hilliker, R.J.; Wacker, M. Nurnberger, A.L. (2013). Improving discovery of and access to digital repository contents using semantic web standards: Columbia University's Academic Commons. *Journal Of Library Metadata*, 13 (2-3), 80-94. <https://doi.org/10.1080/19386389.2013.826036>
- Inger, S.; Gardner, T. (2013). Library technology in content discovery: evidence from a large-scale reader survey. *Insights*, 26 (2), 120-127. <http://doi.org/10.1629/2048-7754.75>
- Injante, R.; Mauricio, D. (2019). Método para recomendar factores de posicionamiento personalizados en el motor de búsqueda de Google. *Revista Española De Documentación Científica*, 43 (1), 1-13. <https://doi.org/10.3989/redc.2020.1.1628>
- Kelly B.; Nixon W. (2013). SEO analysis of institutional repositories: what's the back story? En: *Open Repositories 2013*. University of Bath. <http://opus.bath.ac.uk/35871/> [Fecha de consulta: 19/03/2020].

- Killoran, J.B. (2013). How to use search engine optimization techniques to increase website visibility. *IEEE Transactions On Professional Communication*, 56 (1), 50-66. <https://doi.org/10.1109/TPC.2012.2237255>
- Macgregor, G. (2019). Improving the discoverability and web impact of open repositories: techniques and evaluation. *Code4Lib journal*, 43. <https://journal.code4lib.org/articles/14180> [Fecha de consulta: 18/03/2020].
- Martín-Martín, A.; Orduña-Malea, E.; Ayllón, J.M.; López-Cózar, E.D. (2014). Does Google Scholar contain all highly cited documents (1950-2013)? *EC3 working papers*. Granada: Universidad de Granada. <https://arxiv.org/abs/1410.8464>.
- Martín-Martín, A.; Orduña-Malea, E.; Ayllón, J. M.; Delgado López-Cózar, E. D. (2016). Back to the past: on the shoulders of an academic search engine giant. *Scientometrics*, 107 (3), 1477-1487. <https://doi.org/10.1007/s11192-016-1917-2>
- Martín-Martín, A.; Orduña-Malea, E.; Harzing, A.W.; Delgado López-Cózar, E. (2017). Can we use Google Scholar to identify highly-cited documents?. *Journal of Informetrics*, 11 (1), 152-163. <https://doi.org/10.1016/j.joi.2016.11.008>
- Mixter, J.; O'Brien, P.; Arlitsch, K. (2014). Describing theses and dissertations using Schema.org. *Proc. Int'l Conf. on Dublin Core and Metadata Applications 2014*, 138-146.
- MOZ (2015). *Search Engine Ranking Factors 2015: expert survey and correlation data*. Disponible en: <https://moz.com/search-ranking-factors> [Fecha de consulta: 26/03/2020].
- MOZ (2019). *Schema.org Markup*. Disponible en: <https://moz.com/learn/seo/schema-structured-data> [Fecha de consulta: 26/07/2020].
- MOZ (2020). *What is SEO*. Disponible en: <https://moz.com/learn/seo/what-is-seo> [Fecha de consulta 22/07/2020]. no lo veo en texto
- Muñoz-Martín, B. (2015). Incrementa el impacto de tus artículos y blogs: de la invisibilidad a la visibilidad. *Revista de la Sociedad Otorrinolaringológica de Castilla y León, Cantabria y La Rioja*, 4, s6-s32. <http://hdl.handle.net/10366/126907>.
- OCLC (2018). *International linked data survey*. Disponible en: <https://www.oclc.org/research/themes/data-science/linkedata/linked-data-survey.html> [Fecha de consulta: 30/03/2020].
- Orduña-Malea, E.; Delgado López-Cózar, E. (2014). The dark side of open access in Google and Google Scholar: the case of Latin-American repositories. *Scientometrics*, 102, 829-846. <https://doi.org/10.1007/s11192-014-1369-5>
- Pekala, S. (2018). Microdata in the IR: a low-barrier approach to enhancing discovery of institutional repository materials in Google. *Code4Lib journal*, 39. <https://journal.code4lib.org/articles/13191> [Fecha de consulta: 18/03/2020].
- Price, C. (2019). What is Schema markup & why it's important for SEO. *Search engine journal*. Disponible en: <https://www.searchenginejournal.com/technical-seo/schema/>. [Fecha de consulta: 06/08/2020].
- Ray, L. (2019). We surveyed 1400 searchers about Google: here's what we learned. *Moz blog*. Disponible en: <https://moz.com/blog/new-google-survey-results> [Fecha de consulta: 01/06/2020].
- Rovira, C.; Guerrero-Solé, F.; Codina, L. (2018). Received citations as a main SEO factor of Google Scholar results ranking. *El Profesional De La Información*, 27 (3), 559-569. <https://doi.org/10.3145/epi.2018.may.09>
- Searchmetrics (2018), *Ranking factors 2018: targeted analysis for more success on Google and in your online market*. Disponible en: <https://www.searchmetrics.com/knowledge-base/ranking-factors/> [Fecha de consulta: 26/03/2020].
- Serrano-cobos, J. (2015). *SEO: introducción a la disciplina del posicionamiento en buscadores*. Barcelona: UOC.
- Sexton, W.; Aery, S. (2013). *Using Schema.org & Google Site Search with library digital collections*. Washington, DC: CNI. Disponible en: <https://www.cni.org/news/video-using-schema-org-google-site-search-with-library-digital-collections> [Fecha de consulta: 25/03/2020]. <https://vimeo.com/65144172>
- Singhal, A.; Cutts, M. (2010). Using site speed in web search ranking, *Google webmaster central blog: official news on crawling and indexing sites for the Google index*. Disponible en: <https://webmasters.googleblog.com/2010/04/using-site-speed-in-web-search-ranking.html> [Fecha de consulta: 26/03/2020].
- Smith-Yoshimura, K. (2018). Analysis of 2018 international linked data survey for implementers. *Code4Lib journal*, 42. <https://journal.code4lib.org/articles/13867>
- Su, A.J.; Hu, Y.C.; Kuzmanovic, A.; Koh, C.K. (2014). How to improve your Google ranking: myths and reality. *ACM transactions on the Web*, 8 (2), article 8. <https://doi.org/10.1145/2579990>
- Suber, P. (2010). *Open access overview: focusing on open access to peer-reviewed research articles and their preprints*. Disponible en: <http://legacy.earlham.edu/~peeters/fos/overview.htm> [Fecha de consulta: 08/05/2020].
- Sulé-Duesa, A. (2015). Schema.org, la mejora de la visualización de los resultados en los buscadores y mucho más. *BiD: textos universitaris de biblioteconomia i documentació*, 34. <http://bid.ub.edu/es/34/sule.htm>
- Tenopir, C.; Christian, Lisa; Kaufman, J. (2019). Seeking, reading, and use of scholarly articles: an international study of perceptions and behavior of researchers. *Publications*, 7 (1), 1-23. <https://doi.org/10.3390/publications7010018>
- Tmava A.M.; Alemneh D.G. (2013). Enhancing content visibility in institutional repositories: overview of factors that affect digital resources discoverability. En: *iConference, Fort Worth*. <https://digital.library.unt.edu/ark:/67531/metadc146566/> [Fecha de consulta: 18/03/2020].
- Tonkin, E.L.; Taylor, S.; Tourte, G.J.L. (2013). Cover sheets considered harmful. *Information services & use*, 33, 129-137. <https://doi.org/10.3233/ISU-130705>
- Wallis, R.; Isaac, A.; Charles, V.; Manguinhas, H. (2017). Recommendations for the application of Schema.org to aggregated cultural heritage metadata to increase relevance and visibility to search engines: the case of Europeana. *Code4Lib journal*, 36. <https://journal.code4lib.org/articles/12330>
- Wang, Z.; Phan, D. (2018). Using page speed in mobile search ranking. *Google webmaster central blog: official news on crawling and indexing sites for the Google index*. Disponible en: <https://webmasters.googleblog.com/2018/01/using-page-speed-in-mobile-search.html> [Fecha de consulta: 26/03/2020].

Apéndice 1. Muestra de documentos de RODERIC.

1. <http://hdl.handle.net/10550/67558>
2. <http://hdl.handle.net/10550/69823>
3. <http://hdl.handle.net/10550/29512>
4. <http://hdl.handle.net/10550/69000>
5. <http://hdl.handle.net/10550/66201>
6. <http://hdl.handle.net/10550/70118>
7. <http://hdl.handle.net/10550/69681>
8. <http://hdl.handle.net/10550/69668>
9. <http://hdl.handle.net/10550/69000>
10. <http://hdl.handle.net/10550/68793>
11. <http://hdl.handle.net/10550/40281>
12. <http://hdl.handle.net/10550/41039>
13. <http://hdl.handle.net/10550/41822>
14. <http://hdl.handle.net/10550/40826>
15. <http://hdl.handle.net/10550/39868>
16. <http://hdl.handle.net/10550/68339>
17. <http://hdl.handle.net/10550/67671>
18. <http://hdl.handle.net/10550/70137>
19. <http://hdl.handle.net/10550/69843>
20. <http://hdl.handle.net/10550/26369>

Apéndice 2. Términos de búsqueda, buscador en el que se ha realizado la búsqueda y posición en el ranking para cada una de las fechas. El valor "No" indica que no se ha encontrado el resultado de RODERIC entre las cinco primeras páginas de resultados. Cuando el resultado corresponde a una página de la versión móvil del repositorio se indica: "(móvil)".

ID	Término de búsqueda	Buscador	Posición 1/4/2019	Posición 1/11/2019	Posición 1/6/2020
1	actively fiber laser modulation acousto-optic tunable bandpass filter	Google Search	No	No	No
		Google Scholar	No	2	2
2	redes sociales cartotecas universitarias	Google Search	2	2	1
		Google Scholar	3	2	No
3	cerámica ibérica albacete iconografía	Google Search	2	No	No
		Google Scholar	1 2 (móvil)	1 2 (móvil)	1
4	Strongyloides stercoralis hyperinfection syndrome pancreas transplantation	Google Search	6	3	3
		Google Scholar	No	No	38
5	profesor educación en valores retos	Google Search	10	2	12
		Google Scholar	No	30	31
6	historia educación musical España caso valenciano	Google Search	4	7	7
		Google Scholar	1	3	2
7	Influence of substrate design for in vitro mechanical testing	Google Search	No	3	No
		Google Scholar	2	2	1
8	muerte provincia de Santa Cruz	Google Search	No	No	No
		Google Scholar	No	No	No
9	Strongyloides stercoralis hyperinfection syndrome	Google Search	22	22	32
		Google Scholar	No	No	No
10	Tratamiento defectos óseos sépticos masivos femorales	Google Search	No	36	3
		Google Scholar	1	10	2
11	Discurso pronunciado Academia Científico-Literaria Juventud Católica Valencia	Google Search	1	3 4 (móvil)	1
		Google Scholar	8	12 13 (móvil)	1
12	Alimentos adulterados defunciones	Google Search	1 (móvil)	1 (móvil)	2 (móvil)
		Google Scholar	No	1 (móvil)	1
13	Discurso leído Academia Científico-Literaria Juventud Católica de Valencia	Google Search	1 (2 móvil)	1 (2 móvil)	1 (2 móvil)
		Google Scholar	8	1	1
14	variedades vid común Andalucía	Google Search	3	4	7
		Google Scholar	No	No	14
15	alcornoque industria taponera	Google Search	17	11	5
		Google Scholar	1	1	No
16	multimodal analysis British television commercials	Google Search	2	2	No
		Google Scholar	No	No	6
17	Spinal Muscular Atrophy Reporter System	Google Search	4	1	1
		Google Scholar	No	No	No
18	Complicaciones infecciosas en el trasplante de sangre de cordón umbilical	Google Search	No	7	1
		Google Scholar	No	No	2
19	Exposición laboral factores psicosociales daño musculoesquelético personal sanitario enfermería	Google Search	No	1	1
		Google Scholar	No	No	3
20	trasplante de sangre de cordón umbilical	Google Search	No	No	2
		Google Scholar	No	No	No