



UNIVERSIDAD
NACIONAL
DE LA PLATA

FACULTAD DE INFORMÁTICA

TESINA DE LICENCIATURA

TÍTULO: HERA: Herramienta para Enriquecimiento de Recursos Académicos

AUTORES: Juan Francisco Porto

DIRECTOR: Enzo Rucci

CODIRECTOR: Gonzalo Villarreal

ASESOR PROFESIONAL:

CARRERA: Licenciatura en Sistemas

Resumen

En esta tesina de grado describimos la problemática de la evaluación de calidad de contenido académico en tiempos contemporáneos. Mediante un análisis sobre medios de evaluación de calidad de recursos académicos, métricas, uso de las mismas, y bases de datos que las contienen, se propuso el diseño y desarrollo de una herramienta web que busca dar respuesta a este problema. HERA es una herramienta que apunta a simplificar, agilizar y apoyar el proceso de determinar la calidad y el impacto de un recurso académico.

Palabras Clave

Artículo científico, Revista científica, Paper, Journal, Bibliometría, Bases de datos académicas, Bases de dato bibliográficas, Cienciometría, Recuperación de información, Evaluación bibliográfica.

Conclusiones

HERA representa una herramienta que de manera ágil permite integrar y visualizar métricas de bases de datos relevantes y reconocidas en un único sitio. Esto es de vital importancia a la hora de evaluar contenido, no sólo por eximir a los investigadores de la tediosa tarea de buscar métricas en muchas páginas web, sino también por facilitar la visualización de métricas para un análisis integral de estas.

Trabajos Realizados

- Análisis de bases de datos bibliográficas y sistemas de evaluación particulares para la indexación de recursos académicos.
- Análisis de técnicas y herramientas para la recuperación, procesamiento e integración de métricas de revistas y artículos desde diferentes fuentes.
- Diseño y desarrollo de una aplicación web que permita enriquecer un recurso académico, a través de la integración de información asociada proveniente de bases de datos académicas.

Trabajos Futuros

- Expandir el banco de bases de datos académicas, buscando aumentar la diversidad y utilidad de HERA
- Implementar una API REST, permitiendo generar nuevas soluciones que integren y se nutran los datos de HERA
- Optimización de tiempo de búsquedas de métricas mediante la incorporación de una memoria caché



Universidad Nacional de La Plata
Facultad de Informática

Tesina de Licenciatura en Sistemas

HERA: Herramienta para Enriquecimiento de Recursos Académicos

Autor: Juan Francisco Porto

Director: Dr. Enzo Rucci
Co-Director: Dr. Gonzalo Luján Villarreal

Agradecimientos

A mis padres, que mediante trabajo duro y sacrificio, y tantas lecciones de la vida, incentivaron y motivaron mi educación para llegar hasta este momento, siendo pilares centrales en mi vida.

A mis otros padres, que incondicionalmente también me brindaron siempre su apoyo, en las buenas como en las malas, y les debo tantas alegrías.

A toda mi familia, por estar siempre pendiente de mis necesidades, cada uno brindando su grano de arena para construir este castillo.

A mis amigos, que con o sin quererlo forman parte de aquello que lo convierte a uno en el ser humano que es en el presente.

A mis directores, Enzo y Gonzalo, por la posibilidad de realizar este trabajo, su incansable tutela, y su paciencia férrea a la hora de corregir algunos horrores que pudieran haber aparecido en esta redacción.

Resumen

Mientras que la importancia de saber si un recurso académico es de buena calidad es innegable, los procesos de evaluación pueden resultar tediosos de realizar en profundidad. Un análisis de la diversa variedad de métricas sobre recursos y tendencias en su uso, demuestra que existen valoraciones imparciales de las mismas, se encuentran de manera dispersa, lo cual desalienta a buscarlas todas, y varias son utilizadas en mayor medida que otras, dejando algunas incluso ignoradas. Además, ciertos aspectos de los recursos que carecen de valor numérico, como el hecho de estar alojados en bases de renombre y/o acceso abierto, no suelen ser tomados en cuenta a la hora de realizar análisis breves sobre la calidad de un recurso.

En esta tesina se propuso una herramienta web que busca dar solución a esta problemática. Luego de analizar cuidadosamente los diversos tipos de métricas existentes, y numerosas bases de datos académicas (identificando sesgos, metodologías de evaluación y medios de acceso a los datos), se hizo una selección de los datos de aquellas consideradas apropiadas para extraer su información. HERA representa una herramienta que de manera ágil permite integrar y visualizar métricas de bases de datos relevantes y reconocidas en un único sitio. Esto es de vital importancia a la hora de evaluar contenido, no sólo por eximir a los investigadores de la tediosa tarea de buscar métricas en muchas páginas web, sino también por facilitar la visualización de métricas para un análisis integral de estas.

Índice

Índice	5
Capítulo 1 - Introducción	8
1.1 Motivación	8
1.1.1 Reseña histórica sobre la investigación y su publicación	8
1.1.2 La evaluación de la calidad e impacto en el presente	10
1.1.3 Acerca del proyecto	13
1.2 Objetivos y Metodología	13
1.3 Contribuciones	15
1.4 Organización del documento	16
Capítulo 2 - Marco referencial y estado del arte	17
2.1 Tendencias en el uso de los indicadores	17
2.1.1 Las citas y su número	18
2.1.2 La medición del impacto de un journal	21
2.2 Bases de datos académicas	25
2.2.1 Scopus	25
2.2.2 Web of Science	28
2.2.3 Directory of Open Access Journals	31
2.2.4 Latindex	34
2.2.5 Red Iberoamericana de Innovación y Conocimiento Científico	37
2.2.6 Crossref	40
2.2.7 Scimago Journal & Country Rank	42
2.2.8 Google Scholar	44
2.2.9 Microsoft Academic	45
2.2.10 Semantic Scholar	47
2.2.11 Servicios de métricas alternativas	49
2.3 Métricas bibliográficas y alternativas	53
2.3.1 Número de citas	54
2.3.2 Factor de impacto	54
2.3.3 Cuartil	55
2.3.4 Índice H	55
2.3.5 Índice G	56
2.3.6 Índice i10	56
2.3.7 Menciones y cálculos alternativos	57

2.3.8 Resumen de información de las bases analizadas	57
2.4 Recolección de datos en la web	60
2.4.1 OAI-PMH	60
2.4.2 Application Programming Interface	61
2.4.3 Web Scraping	62
2.5 Herramientas existentes para la evaluación de recursos académicos	63
2.5.1 ¿Dónde lo publico?	63
2.5.2 Matriz de Información para el Análisis de Revistas	64
Capítulo 3 - Propuesta	66
3.1 Propósito	66
3.2 Análisis y decisiones de diseño	67
3.2.1 Identificadores de los recursos	67
3.2.2 Modalidad de acceso a la información de las bases de datos	68
3.2.3 Medios de acceso a las bases de datos	69
3.2.4 Selección de bases de datos y métricas	71
3.2.4.1 Análisis sobre Scopus	71
3.2.4.2 Análisis sobre WoS	72
3.2.4.3 Análisis sobre DOAJ	73
3.2.4.4 Análisis sobre Latindex	73
3.2.4.5 Análisis sobre REDIB	74
3.2.4.6 Análisis sobre Crossref	75
3.2.4.7 Análisis sobre SCImago	75
3.2.4.8 Análisis sobre Google Scholar	76
3.2.4.9 Análisis sobre Microsoft Academic	76
3.2.4.10 Análisis sobre Semantic Scholar	77
3.2.4.11 Análisis sobre servicios de métricas alternativas	78
3.2.4.12 Decisiones sobre las métricas de citas y factor de impacto	78
3.3 Diseño y desarrollo	79
3.3.1 Diseño de aplicación back-end	80
3.3.2 Diseño de aplicación front-end	81
3.3.3 Componentes de la aplicación	82
3.4 Interfaz gráfica de usuario	86
3.4.1 Búsqueda de journal	87
3.4.1 Búsqueda de paper	89
Capítulo 4 - Pruebas y Resultados	92
4.1 Pruebas y resultados en búsquedas de journals	92

4.1.1 Resultados de Crossref	93
4.1.2 Resultados de DOAJ	95
4.1.3 Resultados de REDIB	96
4.1.4 Resultados de Web of Science	97
4.1.5 Resultados de Scopus y SJR	98
4.2 Pruebas y resultados en búsquedas de papers	101
4.2.1 Resultados de Crossref	101
4.2.2 Resultados de DOAJ	103
4.2.3 Resultados de Microsoft Academic	104
4.2.4 Resultados de Semantic Scholar	105
4.2.5 Resultados de Dimensions	106
4.2.6 Resultados de Altmetric	108
Capítulo 5 - Conclusiones y trabajos futuros	110
5.1 Conclusiones	110
5.2 Trabajos futuros	113
5.2.1 Expandir el banco de bases de datos académicas	113
5.2.2 Implementación de una API REST	114
5.2.3 Optimización de tiempo de búsquedas de métricas	114
Referencias	115

Capítulo 1 - Introducción

Inicialmente presentamos la motivación para el desarrollo de la herramienta, la cual se divide en dos partes: la primera de ellas es una reseña histórica sobre investigación académica mientras que la segunda analiza la calidad de su producción. A continuación, describimos los objetivos y la metodología utilizada para luego detallar las contribuciones de esta tesina. Por último, describimos la organización del resto del documento.

1.1 Motivación

1.1.1 Reseña histórica sobre la investigación y su publicación

En el ámbito académico-científico, múltiples investigadores de diversos campos de estudio se lanzan en la misión de la búsqueda del nuevo conocimiento y saber. Tanto para apoyar, fundamentar y complementar su trabajo, como para documentar sus hallazgos finales, producen lo que denominamos “artículo científico”.

Los artículos científicos, usualmente llamados *papers* (denominación que utilizaremos en adelante), son el medio de comunicación por excelencia de la comunidad científica. En términos formales, un *paper* se define como un informe escrito y publicado en el cual investigadores plasman sus desarrollos y hallazgos, poniéndolos a disposición no sólo para sus colegas sino para el mundo entero. Resulta importante aclarar que no existe un único tipo de *paper*. Mientras que algunos presentan una contribución original, otros refutan o comentan sobre hallazgos de otro investigador. En cualquier caso, el objetivo final es enriquecer el conocimiento humano (Díaz et al., [2016](#)).

Un *paper* ([DEFSA](#)) sólo será considerado válido si pasa por un proceso de revisión por pares, o *peer review* ([Elsevier](#)), el cual consiste en que uno o más árbitros revisen la relevancia y calidad del contenido de acuerdo con diferentes criterios. En caso de superar este proceso, el *paper* será publicado.

De este último paso, habitualmente se ocupan las revistas (o *journals*), aunque dependiendo del área de investigación, puede que se consideren otros medios alternativos de publicación. Por ejemplo, en las áreas de Ciencias Sociales y Humanidades se publican en libros (Howard D. White, et al., [2009](#)), (Engels et al., [2018](#)). También podemos observar preferencias particulares, como es la de investigadores en el área de Ciencias de la Computación, que frecuentemente optan por publicar sus hallazgos en conferencias (Godoy et al., [2015](#)).

Los primeros *journals* surgieron en 1665 (Rozemblum, [2020](#)) y desde ese momento cumplen una doble función: por un lado, difundir lo encontrado en materia científica por investigadores, y por el otro lado, dar crédito a los científicos por sus hallazgos o contribuciones. Esto, a su vez, les permite ganar renombre y popularidad, lo que es imprescindible para avanzar jerárquicamente en su carrera científica.

Desde entonces, la escritura de *papers* y la creación de *journals*, así como el acceso a los mismos, se ha ido incrementando de manera exponencial, producto de la rápida globalización causada por el desarrollo tecnológico. Mencionamos por ejemplo el artículo publicado en el sitio web Wired¹, donde destacan el avance en la velocidad del proceso de investigación gracias a la tecnología². Sumado a esto, el pase a la era digital ha permitido maneras más eficientes y prácticas de subir estos contenidos a diversos espacios digitales (repositorios digitales, redes académicas, etc., de los cuales se hablará posteriormente) y también de poseer perfiles donde se vuelca toda la información de un autor y sus trabajos en conjunto.

La rapidez con la que la cantidad y disponibilidad de información aumenta ha tenido fuerte impacto en el modo en que los investigadores la utilizan. Es habitual que un investigador lea, acceda y utilice información extraída de *papers* de otros colegas. Hasta hace algunas décadas atrás, el requisito de originalidad y el proceso de *peer review* eran elementos suficientes

¹ WIRED - <https://www.wired.com/>

² How Technology is Changing Academic Research - <https://www.wired.com/insights/2013/07/how-technology-is-changing-academic-research/>

para certificar la calidad del material con el que se trabajaba. Sin embargo, esto cambió de manera radical con el surgimiento de la Internet y de sistemas un poco más automatizados para la publicación y evaluación de los mismos (Rozemblum et al., [2015](#)). Un estudio realizado a finales del 2018 por International Association of Scientific, Technical and Medical Publishers sugiere que al menos 3 millones de *papers* fueron publicados en ese año, en aproximadamente 33100 *journals* académicos. Aún así, en el mismo reporte también se estima que el tiempo de lectura promedio ha disminuido, siendo más importante tener una visión general y horizontal de muchos *papers*, que leer en profundidad cada uno (Johnson et al., [2018](#)).

1.1.2 La evaluación de la calidad e impacto en el presente

En este nuevo océano de información los científicos han debido agudizar la elección de la información a la que acceden. Es muy importante evaluar cuidadosamente cada posible recurso a utilizar, asegurarse de que es de calidad y de que será de utilidad para nuevos trabajos de investigación. Al existir tanto contenido, no resulta viable que cada investigador tenga que hacer esto de manera individual. Así es como nace la necesidad de establecer criterios para determinar y diferenciar la calidad de las publicaciones (Gonzalez-Pardo et al., [2020](#)).

Una característica buscada intensamente y asociada estrechamente con la calidad tanto de un *journal* como de un *paper* es el impacto de estos. De hecho, es un error común confundir calidad con impacto y tomarlos como sinónimos. La calidad de un *journal* se asocia a los procesos que esta sigue y a los requisitos que impone para que un *paper* sea finalmente publicado. Por su parte, el impacto se asocia a la repercusión que el *journal* o el *paper* tiene en la comunidad científica, usualmente medido en número de citas (Repiso, [2015](#)).

La necesidad de determinar calidad e impacto de recursos científicos se traduce en lo que hoy vemos como sistemas de evaluación, indicadores y métricas de *journals* o *papers* . Si bien no son una herramienta automatizada

en su totalidad, ni mucho menos estandarizada, estos elementos nos permitirían tener valores concretos o numéricos que posiblemente darían respuesta a preguntas como ¿es este *paper* o este *journal* de buena calidad? ¿es este *paper* o este *journal* de alto impacto? ¿se puede predecir el éxito? ¿se puede entender si es de utilidad?. Lamentablemente, estos valores tienen un carácter subjetivo muy alto por no ser un estándar y por existir diferentes formas de calcularlos y valorarlos en los diferentes repositorios.

En el caso de los *journals*, la calidad suele determinarse a partir de su presencia en determinados sistemas de evaluación, conocidos usualmente como bases de datos bibliográficas. Cabe destacar que, en rigor de verdad, esto no se trata de una métrica por no poseer un valor numérico, con lo cual la interpretación del peso de esta información la hace individualmente cada investigador. Cada una de estas bases de datos tiene sus propios propósitos y criterios de inclusión, y existen en la actualidad una multiplicidad de ellas. Por su parte, la calidad de los *papers* habitualmente se asocia al *journal* en que fue publicado, por la necesidad de cumplir sus requisitos de publicación.

Intentando complementar las métricas clásicas, surgen soluciones alternativas, como lo son, por ejemplo las *altmetrics*. De manera concisa, ya que serán desarrolladas en un capítulo posterior, este tipo de métrica engloba e incorpora nuevos aspectos para evaluar contenido. El punto fuerte de ellas yace en su facilidad de interpretación visual, sus valores numéricos simples y su conveniente portabilidad. Mientras que su atractivo es indiscutible, no dejan de introducir a la mesa nuevos valores a ser interpretados por los investigadores.

Así, encontramos que los investigadores poseen mucha información a disposición, con más que variados métodos de análisis para las mismas. Surgen muchas preguntas respecto de los tipos de *métricas* ya mencionadas, y las problemáticas en general son el carácter natural subjetivo de las mismas, o de los sistemas de evaluación, su poca interrelación, y la posibilidad existente de encontrar indicadores bajo las mismas nomenclaturas, pero calculados de

formas levemente diferentes en las distintas bases de datos, produciendo valores poco similares. En la práctica, la evaluación de calidad por parte de un investigador que analiza estos valores bajo estos términos se vuelve dificultosa.

De este análisis de *journals* y *papers*, sus *métricas* y los problemas que conlleva analizarlas, entendemos que atacar su carácter subjetivo es un verdadero desafío. El carácter subjetivo no sólo lo otorga las métricas e indicadores, sino que por ser algo natural e inherente a la información, el mismo investigador que la lee y analiza le adjudica un valor personal que podría diferir del original.

En la actualidad, existen ya algunas aplicaciones que intentan dar solución a la problemática descrita, ofreciendo a los investigadores una recopilación de información que permite la evaluación de los recursos académicos con métodos variados. Por un lado, tenemos **¿Dónde lo Público?**³, que recopila información de *journals* de ciencias sociales y humanidades, que estén en español o portugués. Una opción superadora es la **Matriz de Información para el Análisis de Revistas (MIAR)**⁴, que da una vista ampliada e integral de métricas de un *journal* en particular que hayamos ingresado en su barra de búsqueda. Lamentablemente, MIAR no provee información a nivel de *paper*, y Dónde lo Público provee información a un nivel muy acotado, además de haber sido descontinuada su actualización.

Basándonos en que “a pesar del paso del tiempo, nuestra mejor medida válida de calidad continúa siendo el juicio humano” (Lindsey, [1989](#); Rhaiem, [2017](#)), y con el propósito de complementar las propuestas existentes, se desprende la necesidad de una herramienta automatizada que agilice el trabajo del investigador en su tarea de evaluación individual de contenidos científicos. En lo posible, deberá hacerlo de manera integral mostrando una visión rápida y concisa de todas las métricas mencionadas, sin que éste deba recorrer muchos sitios en la web por obligación, ni introduciendo sesgos adicionales. Creemos que con este panorama embebido en esta herramienta, los investigadores

³ <http://www.dondelopublico.com/>

⁴ <https://miar.ub.edu/>

podrán ejercer libremente su capacidad de análisis crítico, llevando a resultados genuinos en la determinación de la calidad de un recurso académico.

1.1.3 Acerca del proyecto

Este proyecto nace como una idea originada desde la Secretaría de Investigación y Transferencia de la Facultad de Ciencias Económicas (FCE) de la Universidad Nacional de La Plata. Además de las tareas de apoyo y seguimiento a la investigación en la FCE, la Secretaría lleva adelante diversas líneas de trabajo dedicadas a la formación de docentes, investigadores y becarios en áreas vinculadas a la evaluación científica y sistemas de publicación. Entre estas líneas de trabajo, surge la necesidad de contar con herramientas que agilicen la tarea de selección de revistas para publicar y la evaluación de publicaciones, y que a su vez colabore con la formación de las personas que las utilizan. La Secretaría se puso en contacto con el Instituto de Investigación en Informática (III-LIDI) de la Facultad de Informática de la UNLP y con la dirección PREBI-SEDICI de la Presidencia de la UNLP, y de allí surgió un proyecto conjunto que busca aprovechar la experiencia de cada uno de estos actores en áreas vinculadas a la evaluación científica, visibilidad y métricas, desarrollo de software, gestión de publicaciones, etc. La herramienta presentada en esta tesina de grado es un resultado concreto de esta colaboración interinstitucional.

1.2 Objetivos y Metodología

El objetivo general consiste en desarrollar una aplicación web que permita enriquecer un recurso académico (*paper* o *journal*), integrando información de diferentes bases de datos académicas. La aplicación propuesta estará destinada especialmente a miembros de la comunidad

académica-científica y apuntará a simplificar, agilizar y apoyar el proceso de determinar la calidad y el impacto de un recurso académico.

El objetivo general se descompone en los siguientes objetivos específicos:

- Describir el proceso de producción de publicaciones científicas y tendencias en el análisis de calidad, las bases de datos académicas y las métricas bibliográficas.
- Explicar los desafíos que enfrenta un investigador en la actualidad para determinar la calidad de un recurso académico.
- Diseñar y desarrollar una aplicación web que permita enriquecer un recurso académico, integrando información asociada que simplifique, agilice y apoye en el proceso de determinar su calidad e impacto.

Para cumplir con los objetivos específicos, se realizaron las siguientes actividades:

- Se estudió el proceso de generación de conocimiento y sus medios de publicación.
- Se relevaron las bases de datos académicas existentes y sus métricas bibliográficas.
- Se relevaron las propuestas existentes en la temática.
- Se analizaron diferentes alternativas para poder extraer información de las diferentes bases de datos académicas existentes.
- Se diseñó y desarrolló una aplicación web que permita enriquecer un recurso académico (artículo o revista), integrando información de bases de datos académicas que pueda ser de utilidad para determinar su calidad e impacto.
- Se analizaron opciones de visualización de la información integrada que ofrece la aplicación.

- Se verificó que la aplicación web desarrollada alcance la funcionalidad especificada.

1.3 Contribuciones

- Análisis del tipo de métricas, granularidad, información vinculada a la calidad, grado de estandarización, entre otros parámetros, que ofrecen distintos índices y bases de datos de publicaciones periódicas.
- Documentación de un conjunto de técnicas y herramientas capaces de recuperar, procesar e integrar métricas de revistas y artículos desde diferentes fuentes.
- Una herramienta web orientada especialmente a miembros de la comunidad académica-científica que colabore y sirva de soporte al proceso de determinar la calidad y el impacto de un recurso académico. HERA se encuentra disponible en <https://hera.sedici.unlp.edu.ar>

1.4 Organización del documento

A continuación, describiremos la organización de los capítulos restantes del documento.

En el capítulo 2 establecemos un breve marco histórico de tendencias en el uso de indicadores bibliométricos, mencionando fortalezas, debilidades y posturas sobre estos. A continuación, describimos algunas de las bases de datos académicas más reconocidas, detallando sus procesos de evaluación, métricas disponibles y métodos de acceso a las mismas. Finalmente, analizamos soluciones similares a la propuesta en esta tesina.

En el capítulo 3 presentamos el análisis preliminar realizado que condujo a definir los aspectos funcionales y no funcionales de la aplicación. Luego, describimos su diseño y desarrollo de manera breve, indicando fuentes de información y medios de procesamiento y transferencia de datos. Por último, mostramos la interfaz gráfica para los usuarios.

En el capítulo 4 presentamos la verificación de la herramienta desarrollada, mostrando los resultados de diferentes pruebas que consideran la búsqueda de *journals* y *papers*, e incluyendo casos donde se hallen numerosas métricas de los mismos como casos en los que no se encuentra el recurso.

Finalmente, en el capítulo 5, presentamos las conclusiones y posibles trabajos futuros que pudieran surgir a partir del desarrollo y de la utilización de la aplicación presentada.

Capítulo 2 - Marco referencial y estado del arte

En el presente capítulo se mencionan y explican las principales tendencias en tipos de indicadores utilizados vinculados a revistas científicas. Posteriormente, se realiza la introducción a las bases de datos académicas analizadas, mostrando qué datos aporta cada una a la valoración de calidad de un recurso académico. También se definen métricas comunes encontradas en la web, así como los métodos más habituales para su recopilación. Por último, se presentan algunas soluciones parciales a la problemática planteada en esta tesis, que sirven de inspiración en la decisión final del desarrollo de una herramienta.

2.1 Tendencias en el uso de los indicadores

Como mencionamos en el capítulo anterior en la Sección [1.1.2](#), existe una necesidad de evaluar el contenido académico. Particularmente hablaremos siempre del interés por saber su *calidad* y su *impacto*.

El *impacto* se suele medir en número de citas, o se pueden hacer cálculos con las mismas para medir *impacto* en un periodo de tiempo específico. Desde el surgimiento del primer indicador en aproximadamente 1960 ([Clarivate](#)), se han desarrollado una variedad de métricas e indicadores basados en citar para determinar el impacto de un *journal* o *paper*.

La *calidad* está relacionada con procesos de evaluación para la publicación de un *paper* o *journal*. Estos procesos de *peer review* intentan representar una evaluación crítica por profesionales que entienden del área de estudio del investigador. Evalúan aquello que “consideran” necesario en un *paper* o *journal* para así poder introducirlo en el mundo académico a disposición de aquellos que cumplan las “condiciones necesarias”.

No es de nuestro interés realizar una crítica sobre cada una al punto de decir cuál es mejor o cuál peor. Pero es de suma importancia que

comprendamos qué es lo que un investigador ve e interpreta de cada una de ellas, o al menos entender qué relevancia puede llegar a tener para él a la hora de evaluar contenido *él mismo*. De esta forma, lograremos realizar una selección educada de las mismas para ser utilizadas en la aplicación. En las próximas secciones de este capítulo, analizaremos tendencias del uso y significado de los indicadores, y justificaremos la elección de los mismos para su inclusión en la herramienta desarrollada.

2.1.1 Las citas y su número

Cuando hablamos de citas, nos referimos a las veces en que se hace referencia a un *paper* en otro recurso académico. El autor del recurso citador incorpora de alguna forma parte del contenido de una publicación de otro autor (o de sí mismo, en cuyo caso se denomina autocita). Esto puede hacerse compartiendo una idea, criticando algo que se ha dicho, o simplemente haciendo una cita directa, transcribiendo una oración o párrafo. Luego, se indica de dónde se ha extraído ese contenido, y se coloca en una lista de referencias toda la información necesaria para que los lectores puedan consultarlo. De esta forma, el número de citas es la cantidad de veces que un elemento es citado en otros. Si bien en este trabajo nos enfocaremos principalmente en citas a documentos tradicionales (por ej artículos científicos y tesis), el universo de objetos citables es cada vez más amplio, incluyendo a conjuntos de datos⁵, software⁶ y registros de propiedad intelectual⁷, y está generando nuevos debates en la comunidad científica respecto a cómo medir su impacto, la validez de las citas, técnicas de recolección de métricas, etc⁸.

Generalmente, es aceptado que un número alto de citas representa una buena medida de la calidad de un *paper*. Esto es porque podemos interpretar, como se piensa comúnmente (Lindsey, [1989](#)), que si un *paper* es muy citado,

⁵ <https://www.degruyter.com/document/doi/10.1515/ling-2017-0032/html>

⁶ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7805487.1/>

⁷ <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2019.jul.01>

⁸ <https://direct.mit.edu/dint/article/2/1-2/78/9999/FAIR-Data-Reuse-the-Path-through-Data-Citation>

significa que tiene una amplia contribución al campo de investigación que trata. Muchos autores lo habrían considerado “útil” o “válido”, y por ende lo utilizaron.

Debemos entender que en el mundo de los números de citas, esta no siempre es la verdad. Un número de citas alto puede pertenecer tanto a un *paper* bueno como a uno malo. Las citas pueden provenir no sólo de aquellos que valoran y aprecian la labor de un científico y la reutilizan, sino que puede darse el caso en que un científico realiza una crítica al trabajo de otro (Gruber, [2014](#)). En particular nos interesa remarcar que la crítica puede ser hacia aspectos negativos del trabajo que se cita, lo cual no nos da una buena imagen de la calidad del mismo. No obstante, el número sí nos da una buena noción de la capacidad de comunicación que tiene el *paper* y de los medios que el autor ha elegido para este propósito.

Otro punto a tener en cuenta, relacionado con la forma de comunicación, es que el tipo, la forma en la que se presenta un trabajo de investigación, afecta directamente a la cantidad de citas. Tal es, por ejemplo, el caso que padece la comunidad científica argentina en Ciencias de la Computación (Godoy et al., [2015](#)). Debido a que ellos tienen la preferencia de realizar congresos para exponer sus trabajos, tienden a tener números de citas y de factores de impacto más bajos que aquellos que publican en *journals* de manera tradicional. Sin embargo, también existe evidencia que indica que publicar en congresos tiene un mayor impacto que el que tiene la media de los *journals* que aparecen Journal Citation Ranking (Freyne et al., [2010](#)).

Por último, los números de citas podrán ser un fuerte indicador cuando se esté comparando recursos muy citados, pero no se puede decir lo mismo de aquellos que están en un espectro de citas bajas. Por ejemplo, resulta difícil comparar dos recursos cuando uno posee 2 citas y el otro posee 4. Además, el número de citas es algo que demora en crecer cuando un artículo es recientemente publicado. Es un hecho que cuando no hay un crecimiento notorio y veloz en el mismo, puede producirse un efecto en cadena en el cual los científicos evalúan el futuro de un trabajo y su impacto de manera negativa,

perdiendo la fe en el mismo y condenándolo a perecer. No podemos fiarnos completamente del número de citas, ya que existen varios mecanismos que benefician su crecimiento, hasta a veces de maneras ilícitas (Gruber, [2014](#)).

Entendemos que una forma de exponer en mayor medida un recurso, y así incrementar las posibilidades de ser más citado es la de publicar en un *journal* de gran renombre y prestigio, lo cuál será más difícil para aquellos científicos con menor cantidad de recursos económicos o antecedentes en el área. De esto debe señalarse cómo, de la misma manera, el publicar en *journals* menos prestigiosos o con políticas de acceso muy restringido puede tener influencias negativas en la cantidad de citas. También hay pequeños trucos para atraer más citas, como lo es el hecho de que publicar *review articles* genera un número mayor de citas que aquellos trabajos a los que referencia. Otros autores se han tomado el esfuerzo de enseñarnos cómo es posible manipular las citas de aquellos sitios “inteligentes” que indexan contenido académico mediante mecanismos automatizados, como por ejemplo Google Scholar (Delgado López-Cózar et al., [2012](#)).

Dejando de lado estos aspectos negativos, encontramos puntos que convierten a este indicador en digno de tener en cuenta. Es trabajo de los investigadores tener consciencia de lo que representa. Es vital entender que un número bajo puede significar simplemente que un trabajo es muy reciente, o que está publicado en revistas no muy conocidas o de bajo renombre, o que desafía la corriente de pensamiento dominante y por esto ha sido “castigado”. Debemos comprender, por un lado, que el carácter creativo del último ejemplo debe ser apreciado y evaluado de forma individual. Y por el otro, que los números de citas altos, si bien son imprecisos cuando se los compara, aún son de amplia utilidad para percibir factores importantes.

Finalmente, el número de citas también se relaciona fuertemente con factores como la localización del autor trabajando en universidades prestigiosas, su presencia en listados de importantes bibliografías de científicos, y la posesión de premios científicos y reconocimiento de la

comunidad (Lindsey, [1989](#)). El número podrá ser un valor objetivo, pero las citas son una medida que nace de la apreciación subjetiva de la que el autor citante no puede desprenderse. Es importante tenerlo en cuenta y, de manera imperativa, complementarlo con las demás métricas (Meyer et al., [2009](#)), y aspectos para tener un panorama amplio a la hora de evaluar la calidad de un recurso académico.

2.1.2 La medición del impacto de un *journal*

En el caso de los *journals*, la calidad suele determinarse a partir de su presencia en determinados sistemas de evaluación, conocidos usualmente como bases de datos bibliográficas. Cabe destacar que, en rigor de verdad, esto no se trata de una métrica por no poseer un valor numérico, con lo cual la interpretación del peso de esta información la hace individualmente cada investigador. Cada una de estas bases de datos tiene sus propios propósitos y criterios de inclusión, que detallaremos más adelante. Existen en la actualidad una multiplicidad de ellas. Por su parte, la calidad de los *papers* habitualmente se asocia al *journal* en que fue publicada, por la necesidad de cumplir sus requisitos de publicación.

Otra forma típica de evaluar la calidad de un *journal* es mediante su factor de impacto. Éste fue creado por Eugene Garfield en 1955, y tuvo un peso importante ya que ocurrió un traslado de métricas de evaluación que partía de lo subjetivo a lo objetivo. El factor de impacto nos da una idea de la repercusión que un *journal* tiene en el área de investigación para la que publica. Valores altos de éste generan efectos como la alta competencia de los autores por publicar en ella, alta viabilidad para la publicidad en la misma, obligación a las bibliotecas de poseer estas revistas en su catálogo, etc. (Repiso, [2015](#)). Este factor de impacto es una medida anual, que representa la media anual de citas de *papers* publicados en un determinado *journal* en los últimos dos años.

En el día de hoy, el factor de impacto es utilizado en combinación con rankings de *journals* por las universidades para comprobar el impacto de

investigación de aquellos individuos que trabajan en la misma (Gruber, [2014](#)). Los investigadores dependen del sistema de perfiles académicos, el cual se relaciona de forma directa con valores como las citas y de esta manera con el factor de impacto, para poder hacerse visibles en su ámbito y demostrar su aporte a la comunidad (Gonzalez-Pardo et al., [2020](#)). Gruber critica esta dependencia haciendo notar los esfuerzos que los autores hacen por publicar en estos *journals* por el mero hecho de aparecer en esos rankings, así como también relaciona a los *journals* con empresas de marcas reconocidas. Estas “empresas” tendrían un contenido académico limitado adrede, con poco stock para generar atracción y deseo, y mantendrían alta exclusividad por participar dentro de su producción. Otro problema de los rankings de *journals* es que no poseen formas de medir valores en concordancia unos con otros. Los rankings pueden mostrar valores diferentes para el mismo recurso con alta variabilidad. Y lo que es más, los *papers* son independientes de ellos. Un *paper* muy citado y potencialmente influyente podría estar publicado en un *journal* de bajo nivel o viceversa. Así tendríamos *papers* que no recibirían su merecido reconocimiento o *papers* que dan la impresión de ser mejores de lo que son por estar publicados en *journals* con gran renombre académico y valorados de manera alta por los indicadores de calidad e impacto.

El sistema de rankings de *journals* y factor de impacto no son defectuosos para evaluar un *journal*, ya que de manera general sirven para echar luz a las primeras nociones de calidad de los mismos. Pero no se debe abusar de él para evaluar también a los *papers* que contienen, y mucho menos para hacer evaluación individual de un investigador ya que este número se utiliza para describir un trabajo en equipo y no da una visión precisa de la contribución individual (Moed, [2020](#)). Incluso se fomenta la evaluación en este sentido, en el cual se observa más el puntaje del *journal* que del *paper* en sí. Siendo que este factor representa simplemente un promedio, y que los valores de los que se compone pueden ser más un reflejo de atraktividad que de calidad, representa poco del valor real del investigador como individuo

(Schekman, [2013](#)). Los científicos caen víctimas de este sistema en el cual luchan por ser reconocidos mediante los números de citas y aparecer en *journals* de altos factores de impacto. Adicionalmente, se sienten obligados a publicar en áreas y *journals* ya reconocidos, manipulando en ocasiones la forma en que publican sus contenidos para incrementar sus chances de ser más citados.

Sin embargo, cada día se hace más evidente el hecho de que publicar en *journals* de acceso abierto genera ventajas en el número de citas. Un *journal* o *paper* es de acceso abierto cuando su contenido es distribuido en línea de forma gratuita y sin limitaciones de ninguna índole, como tener que suscribirse al sitio que lo publica. Esto no significa que cualquier recurso incrementará su cantidad de citas por el mero hecho de ser de acceso abierto, sino que aquellos que son de buena calidad tendrán aún más citas siendo publicados en medios de acceso abierto, que si limitan su acceso con políticas de suscripción (Gargouri Y. et al., [2010](#)). Al maximizar la accesibilidad de un recurso, también lo hace su citabilidad. Los investigadores juzgan contenido y seleccionan aquellos que consideran de mayor relevancia y calidad, y siempre tenderán a elegir los mejores recursos cuando puedan acceder a ellos. Si bien los autores suelen tener que enfrentarse a costos de procesamiento de *papers* para publicar en esta modalidad, el consenso general es a favor del acceso abierto, siendo la accesibilidad y mayor diseminación las mayores causas (Enago Academy, [2020](#)). El acceso abierto, entonces, tiende a incrementar el impacto al permitir mayor nivel de accesibilidad, y fomentar el progreso de la investigación exponiendo el contenido a todo el mundo.

En este contexto, ya podemos observar que existe un nuevo bias que se introduce: el del evaluador (Meyer et al., [2009](#)). Este proceso depende de la disponibilidad de los mismos, y los más competentes no son precisamente los más disponibles. Sumado a esto, si un *journal* o base de datos no tiene amplia cobertura de todos los temas, no se puede garantizar que el contenido será evaluado por personas especializadas. Más aún, se encuentra el hecho de que

en muchos casos tampoco están publicados de manera visible los criterios mediante los cuales se seleccionan los recursos evaluados. Por ello, es que lo recomendado es complementar la información de las métricas y filtrarlas bajo interpretación humana, y no dejarse llevar por el llamativo aspecto que puede tener un valor numérico elevado.

No obstante de lo mencionado, creemos que el factor de impacto es un valor que, filtrado de sus sesgos lo más posible, puede dar una mejor medida de calidad de la que otorga en su estado actual y natural. Por ello nos parece importante su visualización en la herramienta. Debemos informar correctamente a los investigadores, para que no se sientan tentados a considerar un valor numérico aislado, y lo comparen con las debidas características que le den un sentido más objetivo. Junto con el factor de impacto y las citas, no puede faltar que se muestren la mayor cantidad de datos posibles que ayuden al investigador a realizar una mejor interpretación de los mismos. Por ejemplo, poder comparar datos de citas de un *paper* con datos de impacto de un *journal*, o información del país del que proviene y del idioma en que se encuentra para localizar posibles imparcialidades (Freyne et al. [2010](#)). También consideraremos incorporar información sobre bases de datos académicas en la web, las cuales tienen sus propios procesos de *peer review* en numerosos casos, y la información posible sobre áreas de estudio. Con todos estos datos a su disposición, un investigador tendrá a su cargo la tarea de analizarlos debidamente y podrá hacer su interpretación personal de los valores numéricos contrastados con aquellos que no lo poseen. A partir de lo logrado, esperamos que pueda hacer una elección de recursos a utilizar de manera más satisfactoria, de acuerdo a los criterios de búsqueda y selección preferidos por el investigador que la utilice.

2.2 Bases de datos académicas

Para la elección de bases de datos, hemos acotado el espectro de análisis. Existen cientos de bases de datos bibliográficas, muchas con gran prestigio, o con más uso dependiendo del área de estudio que se trate. Sin embargo, optamos por considerar las de mayor popularidad entre las multidisciplinarias y regionales. De esta forma, se decidió prescindir de bases como PubMed⁹, que se limitan al ámbito biológico y médico, o el Núcleo Básico de Revistas Científicas de CONICET, que tiene cobertura nacional.

2.2.1 Scopus

En la actualidad, las dos bases de datos de *journals* más reconocidas en muchos ámbitos científicos son Scopus¹⁰ y Web of Science¹¹ (WoS), cuyos dueños y gestores son grandes empresas. En ambas, la indexación e inclusión de *papers* y *journals* es de carácter selectivo. Existen juntas directivas de científicos (Elsevier, [2020](#)) o editores de campos específicos (Clarivate, [WoS](#)) que se aseguran que el material a incluir mantenga ciertos niveles de calidad que ellos consideran adecuados.

En el caso de Scopus, una Junta Asesora de Selección de Contenido (Content Selection & Advisory Board) compuesta de 17 profesionales de diferentes ramas (investigadores, científicos internacionales y bibliotecarios, entre otros) trabajan para asegurarse que el contenido que se indexa en Scopus sea relevante y de calidad. Para ello siguen políticas de selección

⁹ PubMed - <https://pubmed.ncbi.nlm.nih.gov/> es un motor de búsqueda gratuito utilizado para facilitar la búsqueda y obtención de literatura en ciencias de la vida y biomédicas, con el propósito de mejorar la salud tanto a nivel individual como global.

¹⁰ Scopus - <https://www.scopus.com/home.uri>

¹¹ Web of Science - <https://www.webofknowledge.com/>

robustas, disponibles para su conocimiento públicamente. Los criterios mínimos para la aceptación en Scopus son¹²:

- El contenido debe haber pasado por un proceso previo de *peer review* cuyo proceso puede conocerse de manera pública.
- Publicar regularmente y poseer un ISSN¹³ registrado con el ISSN¹⁴ International Centre.
- Tener contenido relevante y que pueda ser leído por una audiencia internacional (tener resumen y títulos en inglés, etc).
- Tener de manera públicamente disponible la declaración de ética de publicación y de malas prácticas.

Más allá de este proceso, Scopus realiza re-evaluaciones cada año para asegurarse de que el contenido indexado mantiene los estándares de calidad originales. Este proceso se realiza mediante el análisis de métricas específicas que, de llegar a ciertos límites preestablecidos, podrían derivar en la discontinuación de la indexación de un *journal*. Se hace revisión de las siguientes métricas (Elsevier, [2020](#)):

- Auto-citas. El *journal* no puede tener más del doble de auto-citas cuando se lo compara con otros de la misma área temática.
- Tasa total de citas. El *journal* no puede recibir menos de la mitad de citas que otros de la misma área temática.
- CiteScore. El *journal* no puede tener la mitad o menos del CiteScore promedio comparado con otros de la misma área temática.
- Número de *papers*. El *journal* no puede producir la mitad o menos de *papers* que aquellos de las mismas áreas temáticas.

¹² [Content Policy and Selection - Content - Scopus - Solutions - https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection](https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection)

¹³ El número ISSN (International Standard Serial Number) es un código de 8 dígitos que sirve para identificar publicaciones periódicas y recursos continuos de toda clase y editada en cualquier soporte, ya sean impresos en papel o en formato digital.

¹⁴ [ISSN International Centre](https://www.issn.org/) - <https://www.issn.org/>

- Número de hipervínculos de texto completo en Scopus.com. Los hipervínculos del *journal* son utilizados la mitad o menos comparado a los de otros de la misma área temática.
- Uso de *abstract* en Scopus.com. Los *abstracts* del *journal* son usados la mitad o menos en relación con otros de la misma área temática.

Si bien la mayor concentración de contenido que posee Scopus está publicada en *journals*, desde el 2014 han expandido su cobertura en cuanto a Conferencias. Estas Conferencias están enfocadas principalmente en las áreas que más publican por este medio, como Ingenierías, Ciencias de la Computación y Física.

Entendemos que el primer interés del investigador estará enfocado en conocer si un recurso se encuentra en Scopus o no, debido a su prestigio y larga trayectoria en el ámbito académico, y se aprecian sus métodos de evaluación y exposición de métricas de citas y valores relativos, como se observa en la Figura 1. El análisis de estas métricas y otras visibles en la página, como SJR o SNIP, así como la decisión de ser incorporadas en la herramienta, será desarrollado más adelante en el capítulo 3.

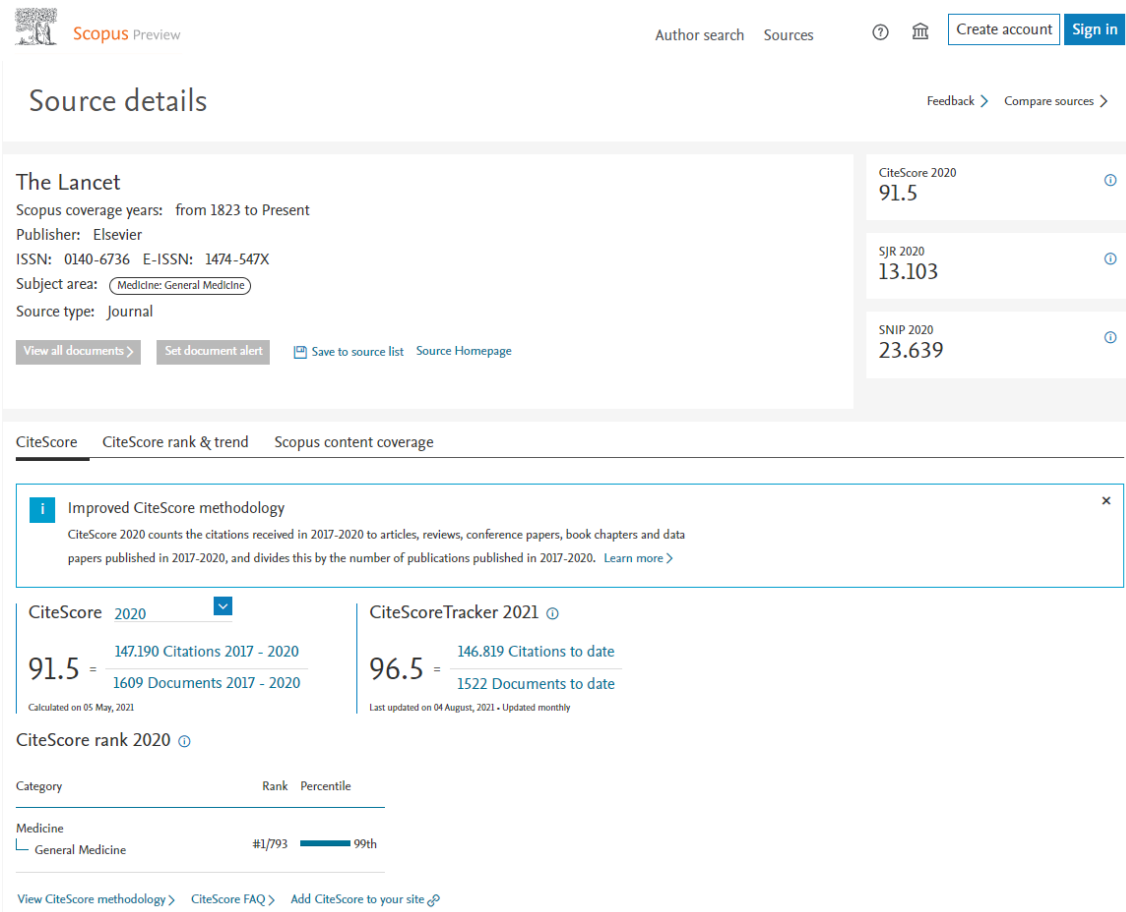


Figura 1. Captura de revista “The Lancet” en Scopus

2.2.2 Web of Science

WoS posee un proceso de admisión similar al de Scopus, tal vez incluso más riguroso. Existe un conjunto de editores expertos y especializados cada uno en diversas áreas para poder evaluar los *journals* con un entendimiento más profundo que el habitual. Estos editores no están afiliados a institutos ni editoriales en particular, intentando así eliminar cualquier imparcialidad o conflicto de interés que pudiese ocurrir en la selección.

Para seleccionar los *journals* se establecen 24 criterios de calidad, que están diseñados para evaluar la calidad editorial y las mejores prácticas a nivel

journal y 4 criterios de impacto para verificar aquellos con mayor influencia y relevancia en sus respectivas áreas temáticas, mayormente basado en las citas. Estos 28 criterios son subdivididos en 4 categorías (Clarivate, [WoS](#)):

- Triaje Inicial: son principalmente métricas de identificación del recurso, como ISSN, título, acceso al contenido y a sus políticas de *peer review*, URL, etc.
- Triaje Editorial: se revisan cuestiones editoriales para verificar si se necesita una evaluación completa. El contenido debe ser académico principalmente, poseer *abstract* y títulos en inglés, declaraciones de principios éticos, detalles de afiliaciones con autores y editoriales, etc.
- Evaluación Editorial de calidad: se asegura que el contenido publicado en el *journal* sigue los lineamientos planteados en los aspectos evaluados en el paso de Triaje Editorial.
- Evaluación Editorial de impacto: en este paso se evalúa si el recurso es de los más influyentes en su área de estudio. Se comparan tanto cantidad de citas como las fuentes de las cuales provienen, como datos más abstractos, como ser la capacidad del contenido de ser interesante. Debe contribuir de alguna forma a expandir y enriquecer de alguna forma el contenido de Web of Science.

Este proceso limita a los autores a enviar solicitudes continuamente en el caso de ser rechazadas. En las últimas 2 etapas del proceso de evaluación se especifica que en caso de no pasar exitosamente se producirá un embargo de al menos 2 años para ser re-evaluados. Sin embargo, el proceso admite *journals* en diferentes niveles de acuerdo a cuánto del mismo ha sido satisfactorio. Aquellos que pasan la evaluación editorial de calidad, son elegibles para entrar al Emerging Sources Citation Index (ESCI). Los que, además, pasan la evaluación editorial de impacto, entran al Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) o Arts & Humanities Citation Index (AHCI) dependiendo de su área temática. La

evaluación siempre continúa, siendo posible que un *journal* suba del ESCI a alguno de los posteriores si gana impacto, o que uno baje a la categoría de ESCI desde una superior por tener un decrecimiento en el mismo. Esto se realiza mediante una continua observación de la actividad de citas de cada *journal* presente en la base de datos. También pueden ser retirados de la misma aquellos que de repente no cumpliesen con alguno de los 24 criterios de calidad. Como se observa en la Figura 2, al buscar un *journal* en WoS, aparece una vista resumida de información bibliográfica del mismo, en el caso de encontrarlo. También podemos saber rápidamente a qué colección pertenece. Si se posee un perfil en la página, al iniciar sesión se logra acceder a información adicional sobre métricas del recurso.

The screenshot shows the Web of Science (WoS) interface. At the top, there is a navigation bar with the Web of Science Group logo and links for 'Search Journals', 'Match Manuscript', 'Downloads', and 'Help Center'. On the right side of the navigation bar, there are buttons for 'Login' and 'Create Free Account'. Below the navigation bar, there is a blue banner with the text 'The power of the Web of Science™ on your mobile device, wherever inspiration strikes.' and links for 'Dismiss' and 'Learn More'. The main content area is divided into several sections. On the left, there is a sidebar with filters and a 'Find a Match' button. The central section is titled 'Refine Your Search Results' and contains a search bar with the text 'Computer science' and a 'Search' button. Below the search bar, it shows 'Search Results' with 'Found 20,570 results (Page 1)' and a 'Share These Results' link. The 'Did you mean this journal?' section displays details for 'BULLETIN OF THE EUROPEAN ASSOCIATION FOR THEORETICAL COMPUTER SCIENCE', including the publisher 'EUROPEAN ASSOC THEORETICAL COMPUTER SCIENCE, C/O IOANNIS CHATZIGIANNAKIS, SEC, COMPUTER TECHNOLOGY INST & PRESS-CTI, RIO, GREECE, 26504', the ISSN/eISSN '0252-9742', and the 'Web of Science Core Collection' 'Emerging Sources Citation Index'. At the bottom right of the journal entry, there is a 'View profile page' button and a note '* Requires free login.'.

Figura 2. Captura de revista en WoS “Master Journal List”

Ambos Scopus y WoS ofrecen métricas con *rankings* y *percentiles* por categoría, las conocidas *citas* e incluso cálculos de *performance esperada*. El objetivo de sitios como éstos es garantizar siempre que la información que se

indexa es de carácter científico y relevante. Más allá de las métricas que se exhiben de aquellos elementos académicos que se encuentran indexados, es importante recalcar que el mero hecho de ser aceptado para incluir un *journal* en estas páginas posee un valor más allá del número. Esto nos da rápidamente un primer indicador de la medida de calidad del *journal* que ha completado exitosamente los criterios de aceptación.

2.2.3 Directory of Open Access Journals

Para dar respuesta a las limitaciones de Scopus y WoS (Mongeon et al., [2016](#); Visser et al., [2021](#)), surgieron diversas bases de datos provenientes de consorcios académicos, con una estrecha relación con la promoción del movimiento de acceso abierto. Un ejemplo a nivel mundial es el Directory of Open Access Journals (DOAJ)¹⁵, que evalúa las políticas de *acceso abierto* para la inclusión de *papers* o *journals* en su base de datos.

DOAJ se lanzó en 2003 como base de datos independiente, orientada a incluir datos de acceso abierto de todas las áreas, países e idiomas. Los costos por mantener esta base de datos son financiados por individuos y organizaciones que comparten la visión de DOAJ de poseer acceso abierto al contenido académico, y depende de estas donaciones y auspicios. Es manejado por una Junta Asesora y un Consejo, ambos constituidos por académicos de numerosas universidades y partes del mundo.

Un *paper* o *journal* será indexado fácilmente si no tiene restricciones de acceso, y podrá incluso acceder a un sello¹⁶ (*DOAJ Seal*) que lo destacará, si cumple aún con requisitos más específicos que demuestren su calidad bajo los criterios del sitio. Este sello no es necesario para ser indexado por DOAJ, pero destaca a aquellos *journals* que demuestran ejercer las mejores prácticas de acceso abierto.

¹⁵ [DOAJ: Directory of Open Access Journals](https://doaj.org/) - <https://doaj.org/>

¹⁶ [The DOAJ Seal – DOAJ](https://doaj.org/apply/seal/) - <https://doaj.org/apply/seal/>

De manera general, para ser indexados por DOAJ, los *journals* deberán cumplir los siguientes criterios (DOAJ [Guide to applying](#)):

- Publicar al menos 5 *papers* por año, o tener 10 *papers* publicados junto con un historial de publicación más allá del año.
- Deben ser de acceso abierto y poseer una declaración en la que se indica que sigue las políticas de acceso de acuerdo con DOAJ.
- El contenido debe estar disponible con acceso abierto de manera inmediata, sin requerimiento de estar registrado, pero se puede exigir un pago por versiones impresas.
- El *journal* debe tener su propio sitio web, el cual debe ser fácil y claro de navegar.
- El *journal* debe adherir a los Principios de Transparencia y Mejores Prácticas en Publicaciones Académicas (*Principles of Transparency and Best Practice in Scholarly Publishing*)¹⁷
- Cada *paper* incluido en el *journal* debe poseer una URL única y estar disponible en HTML o PDF al menos.
- Debe tener un ISSN oficial, exhibido en su página y que debe coincidir el título del *journal* al buscar el ISSN en la web de ISSN International Centre.
- Debe pasar por un proceso de control de calidad donde se verifica la existencia y composición de la junta editorial, editor, el proceso de *peer review* de los *papers* (que debe estar declarado de forma explícita), etc.
- Cumplir con lineamientos de licencia y *copyright* establecidos por DOAJ, quien recomienda hacer uso de una licencia del tipo Creative Commons¹⁸ o con políticas similares.

También establece un listado de la información que debe estar disponible en la página de inicio del *journal* y una restricción con respecto al uso de factores de impacto. Sólo reconoce como oficial el factor de impacto de

¹⁷ [Transparency & best practice – DOAJ](https://doaj.org/apply/transparency/) - <https://doaj.org/apply/transparency/>

¹⁸ [Creative Commons: When we share, everyone wins](https://creativecommons.org/) - <https://creativecommons.org/>

Clarivate¹⁹ y no acepta que se muestren similares métricas o rankings en la página web del *journal*. Otros requisitos adicionales pueden ser exigidos dependiendo de la naturaleza y tipo, o del área de estudio del *journal*.

De la misma forma que Scopus y WoS, estar indexado en la base de datos de DOAJ tiene una importancia de peso similar. El hecho de que un recurso se encuentre indexado por DOAJ es una rápida prueba de cumplir con ciertas pautas de calidad generales y, sobre todo, de poseer buenas políticas de acceso abierto que son exigidas por el proceso de admisión. En la Figura 3 se pone a modo de ejemplo una captura en la cual se observan el DOAJ Seal, información de costos de publicación, políticas y licencias, etc.

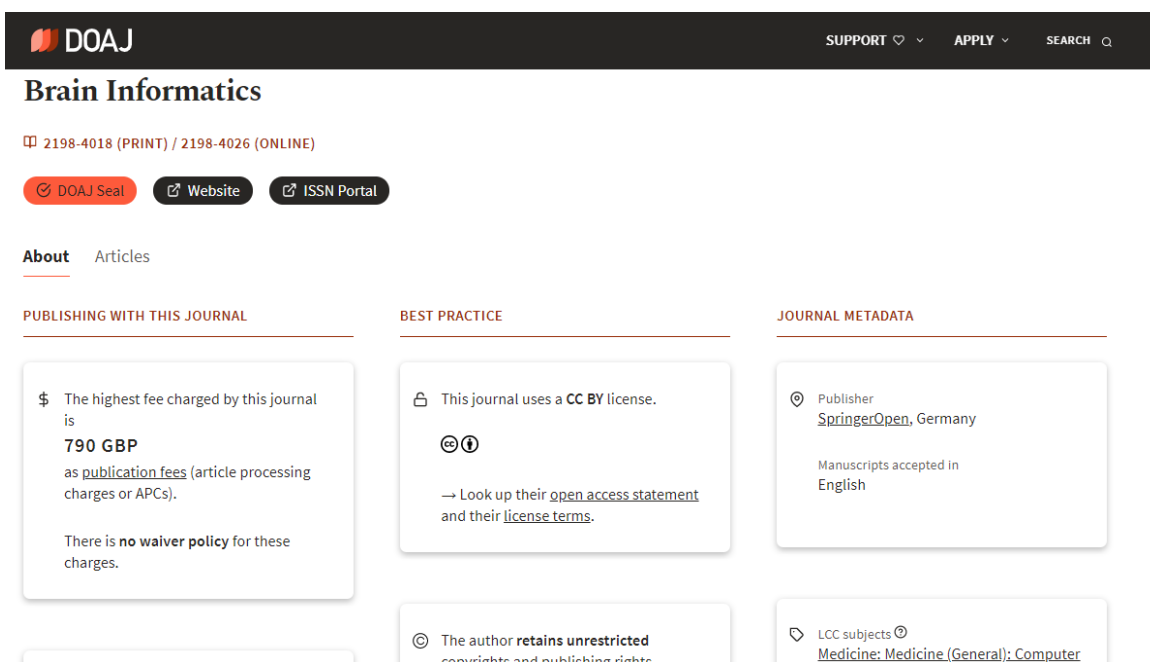


Figura 3. Captura de revista “Brain Informatics” en DOAJ

¹⁹ Clarivate es una compañía formada en 2016, adquiriendo la propiedad intelectual de Thomson Reuters (WoS, MarkMonitor y EndNote). Previamente llamado Thomson ISI, y originalmente creado en el Institute for Scientific Information (ISI), Clarivate se concentra en la analítica de datos, teniendo como objetivo el brindar información y visión de datos, información, soluciones, etc, para acelerar el paso de la innovación científica.

2.2.4 Latindex

En el ámbito latinoamericano, Latindex²⁰ se propone recopilar publicaciones científicas producidas en iberoamérica para mejorar su visibilidad en la región, integrar y expandir el producto del trabajo científico y elevar la calidad e impacto de los contenidos indexados.

Las revistas indexadas aparecerán en lo que se denomina Directorio, que sólo ofrece datos bibliográficos, como título, año de inicio, editorial, etc., y de contacto de los *journals*. En éste hay *journals* editados por países de la región iberoamericana y también algunos editados por organismos internacionales pero de los cuales participan países de América Latina, así como *journals* cuyo contenido trata temas de la región pero son editados en otros países fuera de la misma.

Además existe una sección llamada Catálogo donde sólo aparecen aquellos que se consideran de la más alta calidad, y que a partir del 2018 se restringe a aquellos *journals* que se encuentran en línea. (Latindex, [Metodología](#)). Para poder postular un *journal* para ser incluido en el Catálogo es necesario cumplir con los siguientes requisitos:

- El *journal* debe tener 2 años de antigüedad en su versión en línea.
- Debe tener un sitio web institucional u oficial dedicado.
- El sitio web debe ser de acceso libre a todos los contenidos.
- Cada *paper* publicado debe tener su propio archivo PDF y no estar todos embebidos en uno solo.
- Aquellos *journals* editados por instituciones privadas deben estar indexadas en DOAJ o SciELO para poder postularse.

Luego de la postulación, se evalúan una serie de características y se otorga un puntaje de 1 punto a aquellas que se cumplen. Para ser aceptados se debe haber cumplido con las 7 características básicas obligatorias y conseguir al

²⁰ [Latindex](https://www.latindex.org/latindex/descripcion) - <https://www.latindex.org/latindex/descripcion>

menos 23 puntos extras de las características restantes para alcanzar los 30 puntos. Las 7 características obligatorias son:

- Mencionar nombres de todos responsables editoriales en un listado
- Demostrar la generación ininterrumpida de nuevos contenidos durante los últimos dos años consecutivos, conforme la periodicidad declarada
- Tener todos los documentos firmados por las personas autoras o algo que identifique la autoría
- Aportar en lugar visible información de contacto y nombre de la entidad o institución editora, la cual deberá ser de toda solvencia académica
- Instrucciones para publicar en el sitio web del *journal*
- Detallar el procedimiento empleado para la selección de *papers*, el cual deberá ser externo al *journal*, incluyendo la instancia responsable de la decisión final
- Contar con un ISSN digital

Además de las básicas obligatorias, las características se dividen en presentación, gestión y política editorial, contenido, y revistas en línea.

Entre estas últimas características podemos encontrar algunas como:

10 - Mencionar periodicidad, con número de fascículos que edita al año, fechas que cubre, etc.

12 - Proporcionar nombres de todas las afiliaciones institucionales de los diferentes cuerpos editoriales

20 - Informar con claridad las políticas de derechos de autor respecto al acceso y uso de sus contenidos

21 - Exhibir los códigos de ética que utiliza

23 - Poseer al menos un 40% de contenido original

26 - Exigir que los *papers* tengan un resumen en el idioma original

27 - Exigir que los *papers* tengan un resumen en otro idioma

36 - Usar identificadores uniformes de recursos (Handle, DOI, ARK)

Como se observa en la Figura 4, en la web de Latindex tenemos una sección típica de búsqueda en la cual podemos buscar recursos mediante título, ISSN o término. Al concluir, se muestran los datos de:

- Título
- País
- Editorial
- Situación (Vigente/Cesada/Inactiva)
- ISSN
- Año de inicio
- Subtemas
- Si está incluido en el Catálogo y si es una revista en línea (lo cual puede filtrarse)

latindex
Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal

FAQ Ayuda Facebook Contacto Disminuir Aumentar

¿QUÉ ES? ORGANIZACIÓN SOCIOS EDITORES BIBLIOTECA DEL EDITOR DOCUMENTOS NOTICIAS IDIOMA ISSN: 2310-2799

Búsqueda básica

Directorio (seleccionado) Catálogo 2.0 Revistas en línea

Título, ISSN o término **Barra de búsqueda de journals**

Búsqueda exacta

Búsqueda avanzada

Resultado : **68 Revistas**
63 Títulos únicos

Resultado por: Informática

Directorio (seleccionado) Catálogo 2.0 Revistas en línea

Título	País	Editorial	Situación	ISSN	Año de inicio	Subtemas
Actualidad informática Aranzadi	España	Editorial Aranzadi	Vigente	1131-9623	1991	Computación, Derecho y jurisprudencia
Avances en sistemas e informática	Colombia	Universidad Nacional de Colombia	Vigente	1657-7663	2004	Tecnología
Boletim - Associação Portuguesa de Informática Médica	Portugal	Associação Portuguesa de Informática Médica	Vigente	0871-3073	1986	Medicina, Ciencias de la información

Índices

Gráficas

Resultados de la búsqueda

Figura 4. Captura de búsqueda en Latindex

Nuevamente, si bien toda la información es de utilidad, lo que más nos interesa para hacer una medida de calidad es conocer el hecho de si un recurso se encuentra en el Directorio de Latindex y, en caso afirmativo, si además está incluido en el Catálogo. Debido al riguroso proceso de postulación y aceptación, podemos estar seguros de que si un recurso se encuentra en el Catálogo, cumple con características que reflejan en buena forma la calidad del *journal*.

2.2.5 Red Iberoamericana de Innovación y Conocimiento Científico

Una alternativa existente a Latindex es la Red Iberoamericana de Innovación y Conocimiento Científico (REDIB)²¹, una plataforma que recopila contenidos científicos y académicos en formato electrónico, producidos en el ámbito iberoamericano. Está dirigida tanto a la comunidad editora y científico-académica como a la sociedad en general y al sector empresarial e industrial²².

La página web de REDIB no se limita a la indexación de *journals* o libros, sino que ofrece servicios particulares tanto para usuarios generales como para editores. De manera genérica, se pueden hacer búsquedas especializadas de contenido y/o de autores, viendo algunas métricas limitadas de datos bibliométricos, impacto, etc. de los mismos, y consultar informes avanzados de calidad editorial sobre los recursos. Incluye, además, algunos servicios pagos que se deben presupuestar, como el alojamiento e instalación de Open Journal System (OJS), la impresión de Certificados de Indexación y Certificados de posición en el Ranking de REDIB, y el otorgamiento de banners para portadas de sitios web emitidos por REDIB de manera oficial. Se puede, también, suscribirse individualmente o suscribir una institución, dando beneficios premium a los que las posean. En esta modalidad se pueden visualizar métricas de calidad bibliométrica extras y algunos de los

²¹ [REDIB - El proyecto](#)

²² [REDIB - A quién nos dirigimos - https://redib.org/quienes-somos-a-quien-dirigimos](https://redib.org/quienes-somos-a-quien-dirigimos)

presupuestos para los servicios previamente mencionados tienen costos reducidos.

Un autor, para que su *journal* sea indexado, deberá ofrecer libre acceso al contenido completo del mismo y satisfacer los Criterios de Calidad REDIB (Red Iberoamericana, [Criterios](#)) formales y de contenido. Adicionalmente tendrá que completar un formulario de inscripción de *journal* en el cual se detallan datos de contenido, idiomas, editores, URLs para cosechado de metadatos, derechos y licencias que posee, etc²³. Luego de superar la evaluación, los metadatos deberán ser incorporados a REDIB mediante OAI-PMH de ser posible, o sino de manera manual a través de un formulario web. Los criterios que REDIB utiliza para revisar la calidad de los *journals* están basados en los Criterios de Latindex (Red Iberoamericana, [Criterios](#); Latindex, [Metodología](#)) y, por ende, son muy similares. Están divididos en 5 categorías:

- Política Editorial: aspectos que aseguran el carácter científico
- Gestión Editorial: requisitos editoriales y técnicos para ejecutar la política editorial
- Presentación de la revista: aspectos formales y técnicos de acceso al recurso electrónico
- Presentación de los contenidos: aspectos que identifican contenidos y sus autores, y facilitan y aseguran su difusión e interoperabilidad
- Participación: criterios necesarios para pertenecer a REDIB

Estos criterios, a su vez, se agrupan dentro de cada categoría en 2 subgrupos:

- Obligatorios, que son esenciales para ser admitidos
- Recomendados, que son sugerencias para la mejoría futura

A diferencia de Latindex, que utilizaba un sistema de puntaje, en el caso de REDIB se deben cumplir los criterios de carácter obligatorio. Si bien esto da la impresión de ser más riguroso, teniendo en cuenta que Latindex exige un

²³ [REDIB - Inicio Solicitud revista - https://redib.org/Form/SolicitudRevista](https://redib.org/Form/SolicitudRevista)

80% del puntaje otorgado, 30/38, y que los criterios son casi idénticos, la dificultad de ser admitidos se encuentra en un nivel similar. De la misma forma, la presencia de un recurso en REDIB ya nos proporciona una medida de la calidad que éste posee al haber pasado por el proceso de aprobación.

Navegando el sitio web de REDIB, podemos buscar recursos mediante un buscador que lo localiza de manera individual, o buscarlo en el Ranking REDIB²⁴. En este último, podemos ver numerosas métricas respecto de cada *journal* que son provistas por Clarivate, como visualizamos en la Figura 5. Entre ellas contamos con información de citas, promedio de calificación global, porcentaje de *papers* citados, cuartil al que está asignada, etc.

Cuando se accede a la página individual de cada *journal*, en algunos casos se puede visualizar una miniatura del ranking, para verificar su posición y métricas, así como un widget exportable que representa una vista resumida del mismo. Adicionalmente encontramos información bibliográfica relevante, observado en la Figura 6, como título, ISSN, idioma, periodicidad, licencias, etc., y también algunos indicadores de calidad editorial, estadística con gráficos de citas y crecimiento del journal, acreditaciones nacionales e internacionales, etc.

Inicio / Ranking

Ranking de Revistas REDIB

Metodología del Ranking REDIB

Powered by Clarivate Analytics


Exportar

Nº orden selección	Revista	Clasificación global	Calificador global	Citas Recibidas	País	Perc. F. Imp. Normal	% arts. citados	% ajust. citas materia	% arts. más citados	Perc. Medio	Cuartil
1/1199	The European Journal of Psychology Applied to Legal Context	1/1199	64,722	440	España	99,625	82,258	4,379	61,290	76,056	Q1
2/1199	Comunicar	2/1199	58,364	1285	España	98,540	79,167	8,756	40,000	65,358	Q1
3/1199	Revista de Dialectología y Tradiciones Populares*	3/1199	47,765	0	España	96,289	19,424	99,992	5,755	17,365	Q0
4/1199	Psychosocial Intervention	4/1199	47,344	372	España	95,872	77,174	1,286	13,043	49,343	Q1
5/1199	Journal of New Approaches in Educational Research	5/1199	46,964	366	España	97,456	63,710	1,182	23,387	49,084	Q1
6/1199	Maderas. Ciencia y Tecnología	6/1199	43,668	594	Chile	91,868	63,712	20,425	4,986	37,351	Q0
7/1199	Psicothema	7/1199	43,526	1609	España	94,704	68,056	5,672	7,176	42,023	Q1
8/1199	Revista Iberoamericana de Psicología y Salud	8/1199	42,334	202	España	92,452	68,852	0,714	6,557	43,096	Q1
9/1199	Psicología Educativa	9/1199	42,114	241	España	92,619	63,636	1,861	8,081	44,373	Q1
10/1199	RIED. Revista Iberoamericana de Educación a Distancia	10/1199	41,991	385	España	96,205	56,345	1,346	14,721	41,336	Q1
11/1199	European Journal of Education and Psychology	11/1199	41,051	194	España	91,034	67,647	2,951	5,882	37,741	Q2

Figura 5. Captura de ranking de REDIB

²⁴ [Ranking de Revistas REDIB - https://redib.org/Ranking/Revistas](https://redib.org/Ranking/Revistas)

🔍 Citar
🖨 Imprimir
✉ Enviar por correo
📄 Exportar registro
👤 Agregar a favoritos



Journal of Computer Science & Technology

Argentina

Es una revista de acceso abierto revisada por pares que promueve la difusión de investigaciones originales y experiencias de implementación tecnológica en las áreas de las Ciencias de la Información,...

[+ Descripción completa](#)

Indicadores calidad editorial

PDF | DOI | BIBLIO | CBP

HAP

Acreditaciones internacionales

ESCI

Acreditaciones nacionales

NB CONICET

Widget exportable

<a href="https://redib.org/Record/

Números publicados

2020

+ Volumen 20

- + 2019
- + 2018
- + 2017
- + 2016
- + 2015
- + 2014
- + 2013
- + 2012
- + 2011
- + 2010
- + 2009
- + 2008
- + 2007

Figura 6. Captura de página de *journal* en REDIB

2.2.6 Crossref

Crossref surge en el año 2000²⁵, impulsado por los grandes cambios tecnológicos y la introducción de Internet a la vida cotidiana, que aceleraron el movimiento online de la investigación y comunicación en el ámbito académico. Unos años anteriores a ello, había nacido la necesidad de establecer un medio estandarizado para identificar contenidos digitales, lo que culminó en la creación del DOI²⁶ (Digital Object identifier). Desde ese momento, varias organizaciones se unieron y elaboraron un prototipo con las bases tecnológicas necesarias para crear un medio de referencia basado en este identificador y en metadatos centralizados, y de allí que varios editores se unieron y formaron Crossref para dar vida a este sistema.

²⁵ [Crossref - History](https://www.crossref.org/about/history/) - <https://www.crossref.org/about/history/>

²⁶ El DOI es una cadena alfanumérica única para identificar contenido y proveer un vínculo persistente a su localización en la internet.

En la base de datos de Crossref, se registra contenido académico en forma de metadatos. Se provee el derecho de depositar estos metadatos a aquellos productores de contenido académico que son miembros de la plataforma. Esta membresía es reservada para aquellos representantes de organizaciones que producen contenido profesional y académico, y viene acompañada de una tasa anual, existiendo costos adicionales a la misma por registrar contenido, los cuales dependen del tipo de recurso que se vaya a depositar. Los metadatos se pueden depositar mediante OJS, XML o llenando un formulario web. Adicionalmente se ofrecen otros servicios, como conteo de citas, herramientas de detección de plagio, Crossmark, etc²⁷.

Como sólo se suben metadatos, y el objetivo principal del investigador que los sube es hacer que su contenido sea más fácil de descubrir, es obligación de este mismo subir los metadatos y mantenerlos a lo largo del tiempo. Crossref ofrece completa documentación²⁸ con guías y recomendaciones para llevar a cabo esta tarea correctamente. Otra de las tareas es la de mantener un formato para la muestra del DOI del recurso de acuerdo a cómo se exige en la guía de Crossref²⁹ (Crossref, [2018](#)).

Con respecto a *papers*, entre los metadatos encontramos la métrica de la cantidad de citas que tiene el recurso, y también información útil como el título, el *abstract*, la lista de autores que lo componen, el *journal* (ISSN) al que el *paper* pertenece y la URL original del mismo. En cuanto a *journals*, podemos hallar información sobre cuántos *papers* posee y cuántos se han ido incorporando cada año. Además posee algunos metadatos de cobertura, de los cuales destaca una métrica sobre qué porcentaje de los autores que publican en la misma poseen un ORCID, lo cual es importante para el monitoreo de impacto académico de un investigador (Orduna-Malea et al., [2017](#)).

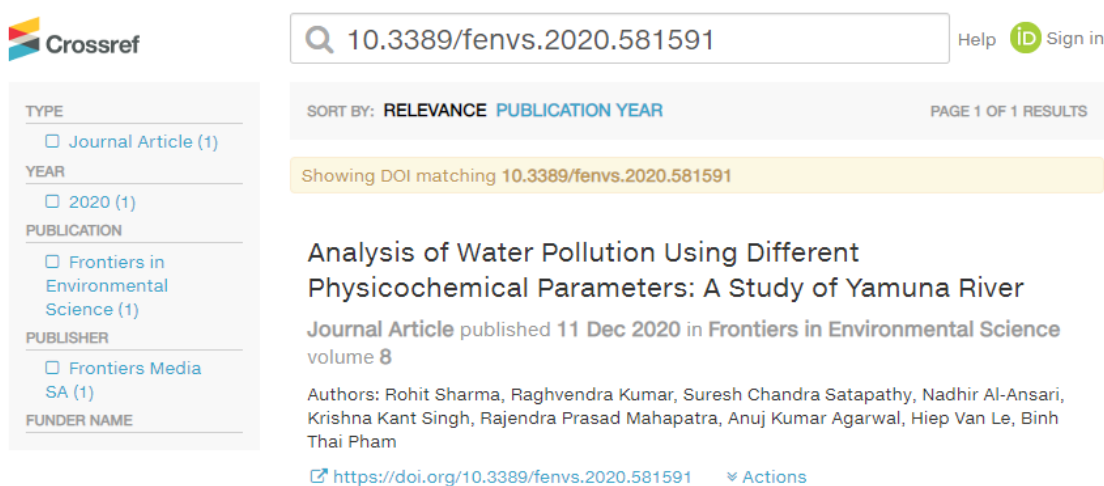
En la Figura 7 mostramos el resultado de realizar una búsqueda de un artículo en base a un DOI en Crossref. La información visible es escasa, ya que

²⁷ [Crossref - Find a service](https://www.crossref.org/services/) - <https://www.crossref.org/services/>

²⁸ [Metadata stewardship](https://www.crossref.org/documentation/metadata-stewardship/) - <https://www.crossref.org/documentation/metadata-stewardship/>

²⁹ [Crossref DOI display guidelines](https://www.crossref.org/display-guidelines/) - <https://www.crossref.org/display-guidelines/>

el propósito de Crossref es el de alojar metadatos, y no está pensada para usuarios finales. En el capítulo 3, se detallará el medio por el cual podemos acceder a estos metadatos y cómo extraerlos para analizarlos.



The screenshot shows the Crossref search interface. At the top left is the Crossref logo. A search bar contains the DOI '10.3389/fenvs.2020.581591'. To the right of the search bar are links for 'Help' and 'Sign in'. Below the search bar, there are sorting options: 'RELEVANCE' and 'PUBLICATION YEAR', and a page indicator 'PAGE 1 OF 1 RESULTS'. A yellow banner indicates 'Showing DOI matching 10.3389/fenvs.2020.581591'. The main content area displays the article title 'Analysis of Water Pollution Using Different Physicochemical Parameters: A Study of Yamuna River', its publication details 'Journal Article published 11 Dec 2020 in Frontiers in Environmental Science volume 8', and the authors: Rohit Sharma, Raghvendra Kumar, Suresh Chandra Satapathy, Nadhir Al-Ansari, Krishna Kant Singh, Rajendra Prasad Mahapatra, Anuj Kumar Agarwal, Hiep Van Le, Binh Thai Pham. At the bottom, there is a link to the DOI and an 'Actions' menu.

Figura 7. Captura de búsqueda de artículo en Crossref

2.2.7 Scimago Journal & Country Rank

SCImago Journal & Country Rank ([SCImago](https://www.scimago.org/)) es un sitio web que elabora métricas e indicadores de calidad de *journals* y países mediante el procesamiento de la información contenida en Scopus. La especialización del grupo SCImago³⁰ el análisis de información y la representación de la misma mediante técnicas de visualización.

En el ranking de *journals*, visto en la Figura 8, que es el que nos interesa, podemos ver la siguiente información de sus métricas:

- Tipo del recurso
- Cuartil
- H-index
- *Papers* totales del año anterior y de los últimos 3 años

³⁰ [Scimago Lab: Research and Web Analytics](https://www.scimago.org/) - <https://www.scimago.org/>

- Cantidad de referencias
- Citas totales en los últimos 3 años
- Métricas de citas por *paper* y referencias por *paper*

Como observamos en la Figura 9, además se puede acceder de forma individual a un *journal* y acceder a una vista del sitio web en la cual vemos información bibliográfica adicional, además de una breve descripción, y gráficos que nos muestran la evolución de la producción científica a lo largo del tiempo del *journal*.

	Title	Type	↓ SJR	H index	Total Docs. (2020)	Total Docs. (3years)	Total Refs. (2020)	Total Cites (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc. (2020)	
1	Ca-A Cancer Journal for Clinicians	journal	62,937 Q1	168	47	119	3452	15499	80	126.34	73.45	
2	MMWR Recommendations and Reports	journal	40,949 Q1	143	10	9	1292	492	9	50.00	129.20	
3	Nature Reviews Molecular Cell Biology	journal	37,461 Q1	431	115	338	8439	10844	167	32.83	73.38	

Figura 8. Captura de ranking de SJR

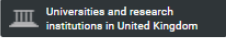
SJR Scimago Journal & Country Rank		
		Enter Journal Title, ISSN or Publisher Name
Nature Reviews Materials		
COUNTRY United Kingdom 	SUBJECT AREA AND CATEGORY Energy └ Energy (miscellaneous) Materials Science └ Biomaterials └ Electronic, Optical and Magnetic Materials └ Materials Chemistry └ Surfaces, Coatings and Films	PUBLISHER Nature Publishing Group
H-INDEX 108	PUBLICATION TYPE Journals	ISSN 20588437
COVERAGE 2016-2020	INFORMATION Homepage How to publish in this journal nrmaterials@nature.com	

Figura 9. Captura de revista “Nature Reviews Materials” en SJR

2.2.8 Google Scholar

El avance tecnológico posibilitó la indexación automática de contenido en Internet por parte de motores de búsquedas como Google o Bing!. Así es como surgen bases de datos como las de Google Scholar (GS)³¹ o Microsoft Academic (MA)³², que acumulan la información de lo que ellos, sus algoritmos de búsqueda, consideran contenido académico, siendo notorio que se puede encontrar casi cualquier documento que cumpla su forma. Los algoritmos de *machine learning* de éstos son muy potentes, e incluso capaces de indexar hasta videoconferencias grabadas, que no son *papers* per sé. Como ya mencionamos son el elemento preferido de un amplio grupo de investigadores para mostrar sus hallazgos (Godoy et al., 2015).

³¹ [Google Scholar - https://scholar.google.com/](https://scholar.google.com/)

³² [Microsoft Academic: Home - https://academic.microsoft.com/home](https://academic.microsoft.com/home)

Al ser más pragmáticos en cuanto a criterios de inclusión, se revolucionó la forma en que los académicos accedían al contenido, que hasta entonces era con una cobertura más específica y selectiva en bases como las de Scopus y WoS. Un estudio realizado, que compara las citas encontradas por GS, Scopus y WoS en diversas áreas, reveló que GS no sólo tiene una mayor cobertura, sino que encuentra una mayor cantidad de las citas en general, en relación a las que encuentran sus contrincantes (Martín-Martín et al., [2018](#)). Además, el mismo estudio refleja que el porcentaje de citas que aparecen en Scopus y WoS pero no en GS es bastante bajo. GS es, también, mucho más eficiente para encontrar material sobre Tesis, Disertaciones, Conferencias y Libros, y se demuestra que GS encuentra en general muchas más citas independientemente del área de estudio o el tipo de recurso. Sin embargo, cabe destacar que alrededor de la mitad de las citas que sólo GS encuentra, son de este tipo de material, sumándole material no publicado como *preprints*, reportes técnicos y otros documentos que tengan forma similar a un *paper*. Se sugiere cautela en el uso de GS, ya que estas citas únicas de GS son en promedio de menor impacto que las que se encuentran WoS y Scopus.

En cuanto al uso para la herramienta, si bien GS pronostica ser preciso a la hora de la búsqueda de recursos, no posee una cantidad de metadatos de gran utilidad para nuestro fin, teniendo sólo datos de citas a nivel artículo, e información de autores. El número de citas que aparece al realizar una búsqueda no está presente cuando el recurso no posee ninguna, y las demás métricas que GS ofrece están más relacionadas con la evaluación del impacto académico de un autor, no del contenido, lo cual es difícil de acceder y obtener de manera automatizada.

2.2.9 Microsoft Academic

De forma alternativa a GS, existe Microsoft Academic (MA), un motor de búsqueda de contenido académico que combina la fuerza de Bing y algoritmos de *machine learning* y técnicas de análisis de lenguaje natural y búsqueda

semántica. Así, no sólo es capaz de descubrir recursos académicos relevantes, sino que es una potente herramienta de búsqueda que ayuda a investigadores en el arduo trabajo de recopilar información³³.

Para que una publicación sea encontrada e indexada en MA debe primero poder ser indexada por Bing, y además será más fácil de descubrir si, como se recomienda en el mismo sitio, sigue los estándares web de tags HTML para contenido académico, que son los mismos que utiliza GS (Google Scholar, [Guías de Inclusión](#)).

MA posee un sistema de ranking de autores el cual refleja valores que tienen ver con la publicación continua, y el impacto que tendrá en investigaciones futuras un *paper* producido por ellos mismos. Un autor que aparece más alto en el ranking no sólo es más citado, sino que lo es en el campo de estudio particular que él trabaja. Estos campos son mayormente generados por la herramienta de machine learning de Microsoft. Entre los datos más relevantes que se pueden encontrar para evaluación de calidad de contenido académico, podemos ver información de número citas, o predicción de número de citas dependiendo si es un valor absoluto o estimado (indicado con un “*”) y áreas o campos de estudio a los que el recurso se vincula, como se puede ver en la Figura 10.

³³ [Microsoft Academic - Frequently Asked Questions](https://academic.microsoft.com/faq) - <https://academic.microsoft.com/faq>

Microsoft Academic The Google scholar experiment: How to index false papers ar X Q

The Google scholar experiment: How to index false papers and manipulate bibliometric indicators

2014 Journal of the Association for Information Science and Technology | Volume: 65, Issue: 3, pp 446-454 | DOI: 10.1002/ASI.23056

Emilio Delgado López-Cózar ¹, Nicolás Robinson-García ¹, Daniel Torres-Salinas ²
¹ University of Granada, ² University of Navarra

34 References 225 Citations*

Abstract

Paper accepted for publication at the Journal of the American Society for Information Science and Technology.
<http://www.asis.org/jasist.html>

Other Links

View PDF: arxiv.org

Website(s): ui.adsabs.harvard.edu | arxiv.org | asistdl.onlinelibrary.wiley.com | core.ac.uk | dblp.uni-trier.de | View More

Additional link(s): dx.doi.org | digibug.ugr.es | onlinelibrary.wiley.com

CORE

Other Versions

The Google Scholar Experiment: how to index false papers and manipulate bibliometric indicators
 2013 arXiv: Digital Libraries | DOI: 10.1002/ASI.23056

Related Topics

Citation analysis Bibliometrics Information science View More (5+)

Figura 10. Captura de artículo en Microsoft Academic

2.2.10 Semantic Scholar

Semantic Scholar es lanzada al público en Noviembre del 2015 (Ariana Eunjug Cha, [2015](#)) como motor de búsqueda de contenido académico desarrollado con Inteligencia Artificial (AI, por sus siglas en inglés) en el Allen Institute for AI³⁴. Surge con el objetivo de dar respuesta a la dificultad que tienen los investigadores de nuestra época de analizar y seleccionar la información adecuada para sus investigaciones proveniente de un conjunto masivo y sobrecargado de datos. En este sentido, su AI se propone acelerar

³⁴ [Allen Institute for AI](https://allenai.org/) - <https://allenai.org/>

este proceso ayudando a los investigadores a descubrir el contenido que es más relevante. Con técnicas de *machine learning* y de análisis de lenguaje natural, ayuda a hacer conexiones entre *papers* de relevancia similar, así como también es capaz de generar un resumen de los *papers* llamado TLDR³⁵³⁶ para resaltar aspectos claves del mismo.

Semantic Scholar toma su contenido de bases de datos y *journals* con los que se afilia³⁷, y también es posible que un autor individual suba su contenido al sitio web de Semantic Scholar siempre y cuando se cumplan las siguientes condiciones:

- Tener una cuenta en Semantic Scholar, en la cual se debe crear un perfil de autor para subir el *paper*
- El *paper* debe estar disponible en formato PDF, debe ser de acceso abierto y debe ser accesible sin ningún tipo de suscripción o login
- La fuente de donde el *paper* proviene originalmente debe ser de las aprobadas por Semantic Scholar o debe estar pertenecer a un *journal* indexado en DOAJ

El proceso de indexación es automático, realizado por la AI que procura mantener una calidad alta en los *papers* y relacionar los contenidos de manera relevante. Una desventaja es que al día de la fecha, el algoritmo sólo indexa contenidos que se encuentren en inglés.

Al navegar la web de Semantic Scholar y acceder a la página individual de un *paper*, como mostramos en la Figura 11, podemos visualizar información bibliográfica detallada del mismo, así como ver citas que recibe, referencias que realiza, y enlaces a cada uno de los recursos relevantes. En particular, ofrece filtros a las citas, para identificar aquellas que citan al *paper* como trasfondo, que citan sus métodos, o que citan sus resultados. Algunas de las citas son clasificadas como “altamente influyentes”, que son aquellas que el algoritmo de Semantic Scholar detecta que tienen un gran impacto en el *paper*

³⁵ [Semantic Scholar FAQ](https://www.semanticscholar.org/faq#tldr) - <https://www.semanticscholar.org/faq#tldr>

³⁶ [TLDR Feature](https://www.semanticscholar.org/product/tldr) - <https://www.semanticscholar.org/product/tldr>

³⁷ [Publishers](https://www.semanticscholar.org/about/publishers) - <https://www.semanticscholar.org/about/publishers>

que lo cita, ya sea porque tratan de un tema y área muy similar o porque a su vez se basan en los mismos contenidos de referencia.

The screenshot shows the Semantic Scholar interface for the article "The Google scholar experiment: How to index false papers and manipulate bibliometric indicators". The page includes the Semantic Scholar logo, a search bar, and navigation options like "Sign In" and "Create Free Account". The article title is prominently displayed, along with the authors' names: Emilio Delgado López-Cózar, N. Robinson-García, and D. Torres-Salinas. It is published in 2014 in the "Journal of the Association for Information Science and Technology". The article's abstract is partially visible, discussing Google Scholar's reception and the introduction of new bibliometric products. A citation statistics box on the right shows 174 total citations, with 7 highly influential, 39 background, and 8 methods citations. Navigation buttons for "View on Wiley" and "arxiv.org" are present, along with options to "Save to Library", "Create Alert", "Cite", and "Launch Research Feed". At the bottom, a horizontal menu offers links to "Abstract", "Topics", "Paper Mentions", "174 Citations", "50 References", and "Related Papers".

Figura 11. Captura de artículo en Semantic Scholar

2.2.11 Servicios de métricas alternativas

Por último, encontramos servicios como Altmetric³⁸ o Dimensions³⁹, que generan aplicaciones que intentan enriquecer el proceso de evaluación de *journals y/o papers* proveyendo de una visión más integradora de valores nuevos que surgen con la época para evaluarlos. Los *altmetrics* nacen bajo la necesidad de rastrear los recursos, que con el pasar del tiempo han sido almacenados de manera online, y de proporcionar filtros adicionales a los tradicionales. Gracias a la era digital, muchos recursos antes inaccesibles, pueden ser localizados y rastreados, y pueden constituirse medidas de impacto que incluyen nuevos tipos de recursos, como diseños experimentales, blogs, comentarios en internet, etc (Priem et al., 2010).

³⁸ [What are altmetrics? – Altmetric](https://www.altmetric.com/about-altmetrics/what-are-altmetrics/) - <https://www.altmetric.com/about-altmetrics/what-are-altmetrics/>

³⁹ [Dimensions - Going beyond traditional databases](https://www.dimensions.ai/dimensions-data/) - <https://www.dimensions.ai/dimensions-data/>

La importancia que se le da a los valores de las altmetrics debe servir de manera complementaria y no definitiva. Las altmetrics miden popularidad, que no necesariamente es una medida de calidad académica en investigación (Gruber, [2014](#)). Las menciones en redes sociales están más vinculadas a una audiencia no académica, y tanto éstas como datos de uso como los números de descargas son valores altamente manipulables, además de venir usualmente incompletos y no estandarizados entre proveedores de los mismos (Moed, [2020](#); Gruber, [2014](#)). Aunque es recalable su uso como indicador de la conexión que existe entre el ámbito científico, y el interés público y social que se le da al tema. Consideramos que las altmetrics son un buen indicador de exposición del contenido académico al mundo y que es útil para entender en dónde es que el contenido es popular y por qué.

Altmetric es uno de los servicios que nace de este movimiento, desarrollado en base a un blog agregador de datos científicos llamado Postgenomic. En sus inicios se creó como una simple aplicación que participó en la competencia de Elsevier Apps for Science en el año 2011, la cual acabó ganando⁴⁰. Con futuras inversiones, y sumándose al grupo de Digital Science⁴¹, se convierte en lo que es hoy día. Altmetric se dispone a rastrear y analizar la actividad online que ocurre en la periferia de la producción científica, mostrando la atención que recibe, quiénes la produjeron, si hay preguntas al respecto, y en general ayudando a entender cómo es que ésta es diseminada y contribuye al mundo. El servicio de Altmetric toma sus datos recorriendo la web, incluyendo información sobre *peer reviews*, citas en Wikipedia y documentos públicos, discusiones en blogs de investigación, medios de comunicación mainstream, citas guardadas en manejadores de referencias como Mendeley, y menciones en redes sociales como Twitter o Reddit, entre otros.

Para brindar este servicio, y poder mostrar su información en cualquier repositorio institucional, página de *journal* o perfil personal, ofrece un *badge* importable. En este *badge* se puede visualizar un “puntaje de atención” que es

⁴⁰ [Altmetric | Apps for Science](https://appsforscience.devpost.com/submissions/4138-altmetric/) - <https://appsforscience.devpost.com/submissions/4138-altmetric/>

⁴¹ [Digital Science: Home](http://www.digital-science.com/) - <http://www.digital-science.com/>

una suma de las menciones encontradas en diferentes redes sociales, periódicos, blogs, etc, como se observa en la Figura 12. Esta suma no se calcula de manera directa, sino que se le asigna un peso al valor obtenido de acuerdo al sitio en el cual se ha mencionado cada recurso. No tiene el mismo valor ser mencionado en una red social de uso general, como por ejemplo Twitter o Facebook, que en un blog científico o un reconocido sitio de noticias. Además el algoritmo que calcula la suma intenta identificar *posts* duplicados con el objetivo de limitar la manipulación de los valores. Como se puede apreciar en la Figura 13, al acceder a la página de detalles de las métricas también puede visualizarse información geográfica y demográfica con respecto a la fuente de las citas.

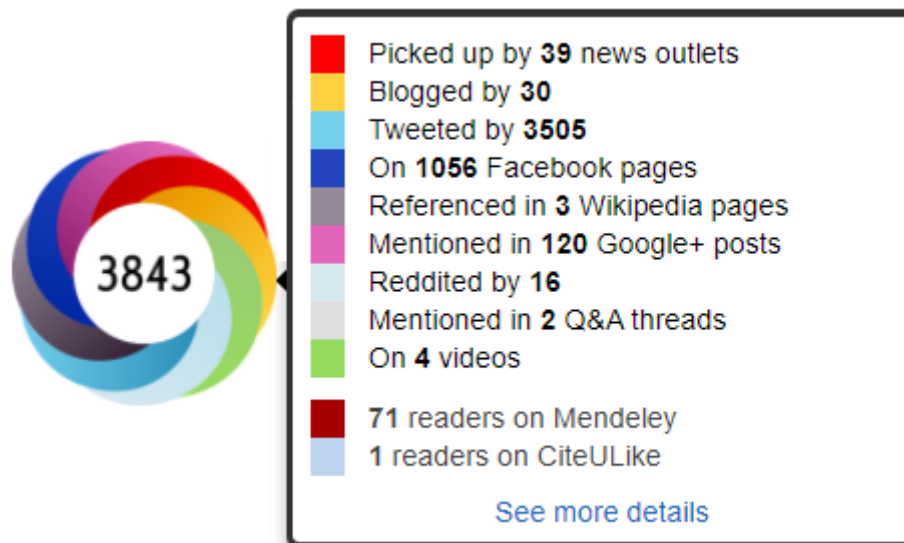


Figura 12. Captura de *badge* de Altmetric al colocar el mouse encima

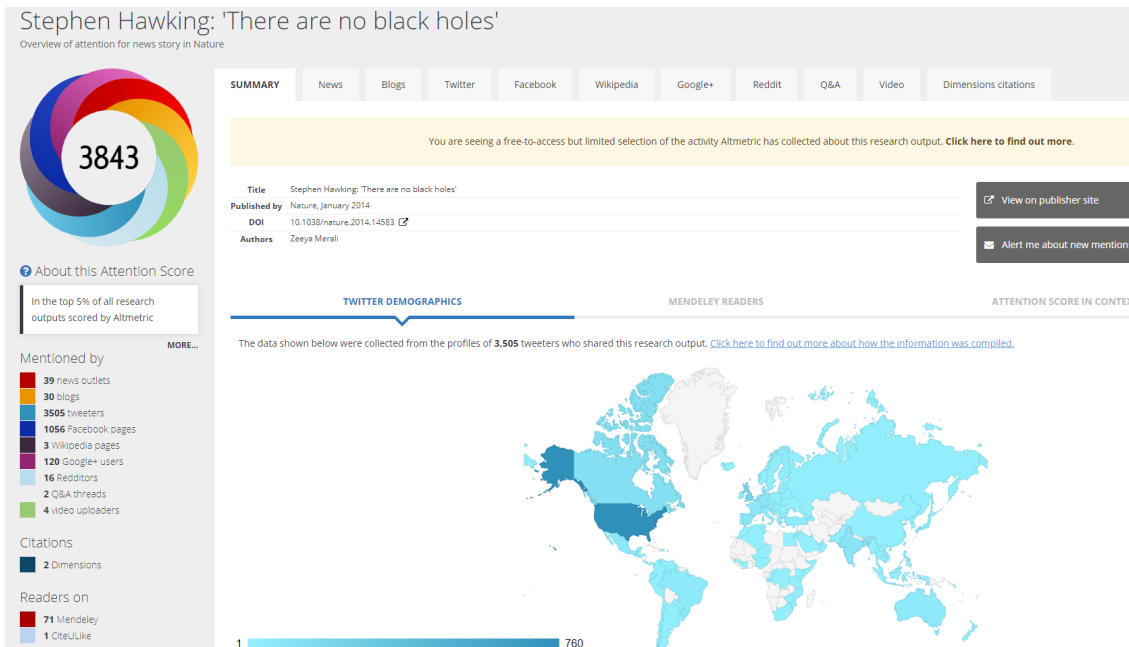


Figura 13. Captura de detalle al clicar *badge* de Altmetric

Dimensions, desarrollado por la misma compañía que Altmetric, Digital Science, se propone alojar en una misma base de datos contenido de diversos tipos y vinculado de maneras comprensibles. Esto lo logra siendo capaz de enlazar publicaciones y citas con subsidios, patentes, políticas de artículos, etc. para dar una visión holística del proceso que acarrea la investigación (Dimensions, [Why Dimensions](#)). El objetivo principal de Dimensions no es el de elaborar métricas, sino el de proveer un abanico más abarcativo de datos de forma gratuita para que los investigadores puedan extraerlos y analizarlos para así crear la nueva generación de indicadores. No obstante, Dimensions analiza las citas de las publicaciones y calcula algunos indicadores a nivel *paper*.

La Dimensions Badge⁴² nos presenta esta información de manera gráfica y fácilmente importable a cualquier página web. En esta *badge* podemos ver información de número de citas a las publicaciones, y además realiza cálculos sobre las mismas para determinar otras métricas:

- Citas Recientes, en los últimos 2 años

⁴² [Dimensions Badge Builder](https://badge.dimensions.ai/) - <https://badge.dimensions.ai/>

- Proporción de Citas por Campo, que indica el desempeño relativo de un *paper* comparado con *papers* de antigüedad similar en el mismo tema de estudio
- Proporción de Citas Relativa, que indica el desempeño relativo de un *paper* comparado con *papers* del mismo área de estudio

Al hacer clic en la *badge* (Figura 14) accedemos al sitio web de Dimensions, que ofrece más detalles de la obra o recurso en particular; allí se cuenta con información sobre, por ejemplo, qué porcentaje de las citas totales son recientes. También posee información sobre las áreas de estudio en las cuales se cita más al elemento *paper* en cuestión.

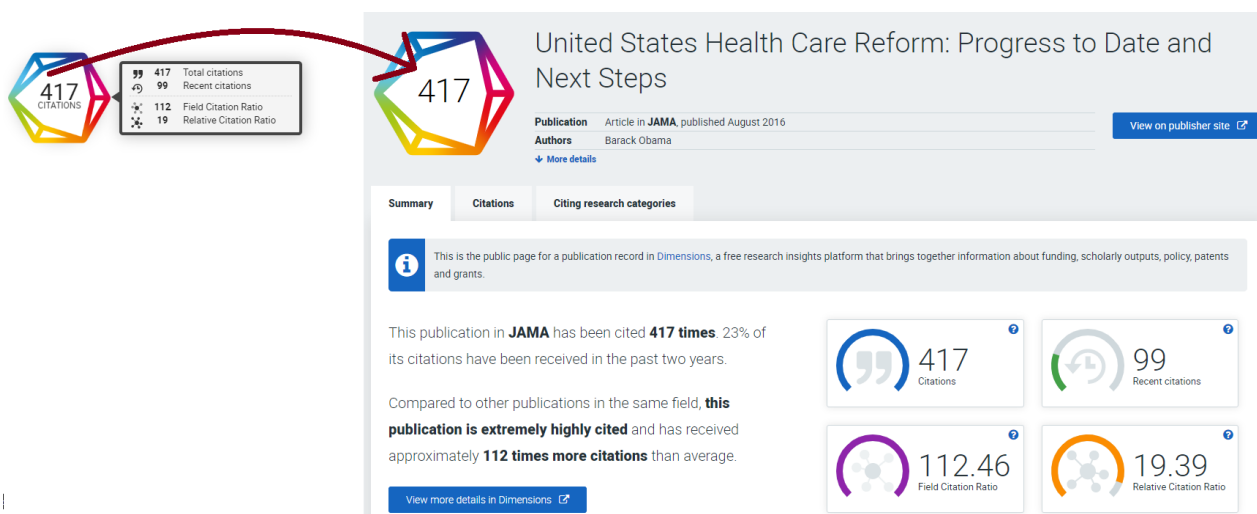


Figura 14. Captura de *badge* de Dimensions y acceso a la página de detalle cuando se hace clic en la misma

2.3 Métricas bibliográficas y alternativas

En este apartado mencionaremos brevemente, y en conjunto, todas las métricas con las que nos topamos en búsqueda de material para el desarrollo

de la herramienta. Si bien algunas ya fueron explicadas en detalle, esta sección servirá como referencia rápida para recordar de manera concisa su significado.

2.3.1 Número de citas

Es el número de veces que un recurso, generalmente un *paper*, es mencionado por otros autores en sus trabajos. Como ya mencionamos en [2.1.1](#), este valor es muy utilizado, pero a pesar de aparentar ser objetivo, no lo es en realidad, y por ello es que debe complementarse con otras métricas para cobrar sentido. De los sitios accedidos, encontramos número de citas en Crossref, SCImago (de manera gráfica en su web), Scopus, Microsoft Academic y Dimensions ([2.4](#)). También se desprenden otros valores de este número, como el *tracker* de Scopus mediante CiteScore o las *citas estimadas* de Microsoft Academic, el *citation velocity* de Semantic Scholar, y las proporciones que Dimensions calcula en base a temas y áreas de estudio.

2.3.2 Factor de impacto

El factor de impacto, o *impact factor*, es una medida que nos dice qué tanto el *paper* promedio de un *journal* fue citado en un año específico. Sirve para medir la importancia que un *journal* tiene en base a cuánto se citan los *papers* en él. Se calcula realizando la siguiente ecuación

$$FI_y = \frac{Citas_y}{Publicaciones_{y-1} + Publicaciones_{y-2}}$$

Donde “*FI*” es el factor de impacto, “*y*” es el año para el cual se quiere averiguar, “*Citas*” son las citas de los *papers* sumadas del mismo año, y “*Publicaciones*” es la cantidad de ítems citables publicados el año. De esta manera, el cálculo representa la relación entre el número de citas recibido en

un año para publicaciones en ese *journal* que fueron introducidas en los últimos dos años, y la suma total de ítems citables publicados en esos dos años.

Es importante recordar lo mencionado en 2.1.2 respecto de las ventajas y desventajas que acarrea el uso de éste valor, como que cada base de datos elige qué revistas tiene en cuenta para el factor de impacto, o el hecho de que es un valor definido por Clarivate. No tenemos toda la información necesaria para verificar su cálculo ya que los criterios por los cuales se seleccionan los documentos y las citas no están publicados (Meyer et al., [2009](#)).

2.3.3 Cuartil

El cuartil es una medida que se extrae del factor de impacto, siendo el cociente entre el ranking de un *journal* en una categoría específica, y el número total de *journals* en esa categoría. La utilidad de este valor está en poder apreciar el valor de impacto más eficientemente, al compararlo sólo con el de *journals* del mismo tiempo y no con aquellos que pertenecen a áreas de estudio menos populares. Al igual que el factor de impacto, este valor es proporcionado por SCImago.

2.3.4 Índice H

El índice H, o *h-index*, fue creado por J.E. Hirsch en el año [2005](#) a modo de poder evaluar la producción científica de un investigador. Se define como “el mayor número *h* de *papers* tal que estos poseen un número de citas $\geq h$ ”.

Si bien pensamos que es una buena medida de impacto de un autor, debemos tener en cuenta al utilizarlo que el *h-index* no provee un contexto para las citas, no tiene en cuenta el tiempo que el autor lleva produciendo contenido y tampoco es una medida universal, entre otras cosas (Becker, [What is the h index?](#)). También su valor puede ser afectado por problemas de nombres similares, o cuando no se controlan las auto-citas, lo que ya es un problema para otras métricas de impacto (Freyne et al., [2010](#)). Una alternativa ofrecida

por Hirsch para comparar científicos de diferente antigüedad es el “valor m ”, el cual divide el *h-index* por la cantidad de años desde que el autor produjo su primer *paper*, y puede ser considerado una medida del éxito del investigador.

De los sitios analizados, el único que nos provee un *h-index* de forma directa es SCImago ([2.2.7](#)). En el caso de esta página, el *h-index* no es referente a los *papers* de un autor, sino a los de un *journal*, lo cual nos da una idea de la productividad y actividad que un *journal* posee. Google Scholar, provee también el *h-index*, pero éste es del autor de los *papers* indexados.

2.3.5 Índice G

Es un valor propuesto como una mejora del H-index por Leo Egghe, y se calcula tomando un grupo de *papers* enumerados en orden decreciente en base a sus números de citas, del cual el G-index se traduce como el mayor y único número tal que los g *papers* superiores recibieron juntos al menos g^2 citas (Egghe, [2006](#)). Si bien supone ayudar a evaluar la performance de los mejores *papers* de un autor y a hacer más evidente la diferencia entre el impacto relativo de varios autores, todavía no es muy utilizado. No es calculado por ninguna de las bases de datos analizadas para el desarrollo de la herramienta.

2.3.6 Índice i10

Este índice fue creado por Google Scholar para calcular el número de publicaciones que al menos poseen 10 citas. Es simple y bastante obvio y evidente, por lo que los sesgos e imparcialidades que se presentan al analizar números de citas podrían verse reducidos, ya que es fácil de interpretar qué hechos pueden afectar al valor y no es una medida muy profunda de impacto académico.

2.3.7 Menciones y cálculos alternativos

La idea detrás de los sistemas alternativos es la de contar con un sistema con métricas diversas, ya que el contenido es cada vez más variado. Es imprescindible tener una herramienta de métricas con un fuerte valor semántico, más veloz y de conocimiento más abierto que los sistemas de evaluación tradicionales.

Esto cobra vida en lo que son las altmetrics. Los servicios de altmetrics presentan usualmente versiones ampliadas de las citas típicas, integrando menciones en redes sociales, videos, y conferencias, números de descargas, y en cálculos de diversas índoles. Por ejemplo, se basan en estadísticas de acceso a los recursos, comparaciones con publicaciones de los mismos campos, y observaciones de citas en intervalos de tiempo específicos, y así otorgan de manera sintética indicadores de tendencias y predicciones que permiten evaluar rápidamente el impacto y contribución que un recurso logra en el mundo académico.

Al calcular valores de números de citas, generalmente se les aplica un significado o un peso a las mismas, de manera de dar importancia a las citas de fuentes de renombre, y también se hacen rigurosos controles contra la manipulación de sus valores.

2.3.8 Resumen de información de las bases analizadas

En la siguiente tabla se visualiza a modo de resumen lo sonsacado de cada base de datos respecto de métricas obtenibles, modo de acceso, y el tipo de evaluación que realiza sobre el recurso:

Tabla 1. Resumen de información provista por las bases de datos

Nombre de Base de Datos	Métodos de selección de contenido	(Principales) Métricas e indicadores provistos	Recurso ⁴³ considerado
Scopus	Proceso de evaluación por Junta Asesora	Número de citas Predicción de citas Factores de impacto	Journal
Web of Science	Proceso de evaluación por editores expertos	Índice de WoS al que pertenece Información bibliográfica	Journal
DOAJ	Proceso de evaluación de acceso abierto y calidad	Licencias Sello de DOAJ	Paper & Journal
Latindex	Proceso de evaluación de calidad (sólo iberoamérica o contenido de la región)	Información bibliográfica Presencia en el Catálogo	Journal
Red Iberoamericana	Proceso de evaluación de calidad (sólo iberoamérica o contenido de la región)	Posición en ranking Información bibliográfica Indicadores de calidad editorial y Acreditaciones	Journal
Crossref	Metadatos estandarizados por miembros de la plataforma	Número de citas Estadísticas de <i>papers</i> por año y autores con ORCID Información bibliográfica	Paper & Journal

⁴³ Si bien se usó el término *Journal*, también pueden estar considerados Conferencias y Libros

Scimago Journal & Country Rank	Recursos contenidos en Scopus	Posición en ranking global Cuartil H-index del <i>journal</i> Información bibliográfica	Journal
Google Scholar	Indexación por Google mediante algoritmos de machine learning	Número de citas H-index de autores	Paper
Microsoft Academic	Indexación por Bing mediante algoritmos de machine learning	Número de citas Predicción de citas Información bibliográfica	Paper
Semantic Scholar	Indexación mediante algoritmos de machine learning ejecutados sobre Bases de Datos Afiliadas	Número de citas Frecuencia de citas Número de citas influyentes	Paper
Altmetric	Realiza cálculos reconociendo menciones en redes sociales, foros, etc. mediante IA	Puntaje de atención basado en menciones con peso dependiendo el sitio Presencia de menciones en redes sociales	Paper
Dimensions	Recopilación de contenidos de bases abiertas (Crossref y PubMed)	Número de citas Número de citas recientes (2 años) Proporción de citas por campo y relativa	Paper

2.4 Recolección de datos en la web

Entre los diversos métodos para obtener datos de las bases de datos mencionadas, analizamos aquellos que eran provistos por las mismas de manera más frecuente. Estos son el uso de APIs y el protocolo OAI-PMH. Como no todos poseen mecanismos de extracción programática, en algunos casos se recurrió al web scraping. En este apartado, explicaremos en qué consisten estos tres métodos más detalladamente.

2.4.1 OAI-PMH

Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)⁴⁴ es un protocolo para la cosecha de metadatos creado por la Open Archive Initiative, que tenía como objetivo el desarrollo de métodos de interoperabilidad para difundir de manera más eficiente los contenidos en Internet (Barrueco et al., 2003). De esta manera, y mediante el uso de este protocolo, diversos sistemas pueden comunicarse e intercambiar información representada en la forma de metadatos en el formato establecido por Dublin Core⁴⁵, y el mecanismo es independiente de la aplicación.

Para este intercambio, los repositorios juegan el rol de *Data Providers* y exponen metadatos utilizando OAI-PMH. Luego los *Service Providers* son aquellos que, utilizando los seis verbos de OAI-PMH, realizan peticiones HTTP para cosechar los metadatos. Éstos le son otorgados en formato XML cuando la petición resulta de forma exitosa.

Entre las peticiones posibles, las más generales son la de *ListIdentifiers* y la de *ListRecords*. *ListIdentifiers* es una forma abreviada del segundo, devolviendo sólo encabezados, los cuales pueden ser filtrados mediante argumentos en la petición. Éste se usa cuando se quiere encontrar información

⁴⁴ [Open Archives Initiative Protocol for Metadata Harvesting \(OAI-PMH\)](https://www.openarchives.org/pmh/) - <https://www.openarchives.org/pmh/>

⁴⁵ [The Dublin Core Metadata Initiative \(DCMI\)](https://dublincore.org/specifications/dublin-core/) - <https://dublincore.org/specifications/dublin-core/>

sobre la existencia de registros que cumplan ciertos criterios, pero no se necesita la información completa del recurso, por lo que se solicitan sus identificadores⁴⁶. *ListRecords* nos trae el registro completo. Un registro contiene todos los metadatos de un recurso, expresado en un formato único, y se identifica por su identificador único combinado con el prefijo de metadatos utilizado en el formato del mismo, y la fecha.

2.4.2 Application Programming Interface

Una Application Programming Interface (API)⁴⁷ es un conjunto de definiciones y protocolos que constituyen un software intermediario que se utiliza para lograr conectar y comunicar dos aplicaciones. No es necesario saber cómo una API está implementada para poder utilizarla y comunicarse con un servidor, sino sólo conocer las rutas a los puntos de entrada de la aplicación y el verbo y formato de datos en el que la API espera recibir peticiones para poder otorgarlas correctamente.

Generalmente, las peticiones a una API se hacen por HTTP, y en su mayoría adhieren a las restricciones REST (*representational state transfer*)⁴⁸. El formato más común para representar la información transferida en una petición a una API es JSON por ser independiente del lenguaje, pero no se limita a éste y puede ser también HTML, XLT, PHP, etc.

Realizar peticiones a APIs es de simple implementación. Es sencillo realizar pruebas sobre las mismas y modificarlas en caso de necesitar mantenimiento. Normalmente contamos con documentación sobre cada una de las rutas a la cual se debe realizar cada petición, bajo qué verbo, y qué formato de datos y encabezados son requeridos. No obstante, cuando esta documentación es escasa, de todas formas es fácil probar su utilización mediante herramientas como Postman⁴⁹. También podemos, en limitados

⁴⁶ Un identificador único identifica de forma no ambigua un ítem dentro de un repositorio

⁴⁷ [What is an API?](https://www.redhat.com/en/topics/api/what-are-application-programming-interfaces) - <https://www.redhat.com/en/topics/api/what-are-application-programming-interfaces>

⁴⁸ [Topics Integration What is a REST API?](https://www.redhat.com/en/topics/api/what-is-a-rest-api) - <https://www.redhat.com/en/topics/api/what-is-a-rest-api>

⁴⁹ <https://www.postman.com/company/about-postman/>

casos, sonsacar los métodos necesarios para su correcto uso mediante inspecciones del tráfico en la red donde se visualizan los llamados a estas API (cuando son de carácter público), pero además existen casos en los que las mismas API responden mensajes de error con información sobre cómo hacer una petición correctamente.

2.4.3 Web Scraping

Web scraping es un mecanismo de extracción de datos en la web que provee una solución alternativa cuando requerimos de acceso a estructuras de datos en forma automática pero el sitio del cual queremos obtenerlos no posee una API pública o cuenta con una pero el acceso a los datos que la misma otorga está limitado⁵⁰.

Para hacer uso de este método, es necesario emular de alguna forma el acto humano de navegar por internet, y así obtener los vínculos de las páginas en las que se encuentran los recursos y contenidos requeridos. El proceso automatizado de navegar la web se llama *web crawling* y trabaja en conjunto con *web scraping* para obtener datos específicos. En navegadores y aplicaciones sofisticadas, como Google Scholar, este proceso de *web crawling* está programado con la inteligencia artificial capaz de acabar encontrando contenidos académicos para su indexación.

La obtención de datos simples no requiere implementar algoritmos complejos, ya que por medio de peticiones HTTP GET con los parámetros adecuados, es posible recuperar el documento HTML referente a la búsqueda de un recurso académico. Esto es, la página web resultado que un usuario humano visualizaría al realizar una búsqueda normal en el sitio web en cuestión. Luego es posible navegarlo mediante el uso de selectores CSS y la HTML DOM API para extraer el enlace al sitio donde el recurso se encuentra de manera individual. Una vez recuperado el documento HTML de este último,

⁵⁰ [What Is Web Scraping And How Does It Work? - https://www.zyte.com/learn/what-is-web-scraping/](https://www.zyte.com/learn/what-is-web-scraping/)

se realiza *web scraping* para obtener los datos preferidos que se encuentren en dicho documento.

2.5 Herramientas existentes para la evaluación de recursos académicos

Como ya mencionamos en [1.1.2](#), existen ya algunas aplicaciones que intentan brindar apoyo a los investigadores en la evaluación de contenido académico, mediante recopilación de datos selectiva, muestra de métricas, y otros métodos. En esta sección haremos un breve análisis de las fortalezas que cada una posee pero también de las debilidades, que representan parte de los catalizadores que nos inspiraron al desarrollo de nuestra herramienta.

2.5.1 ¿Dónde lo publico?

Como describimos anteriormente, ¿Dónde lo publico? (DLP)⁵¹ recopila información de *journals* de ciencias sociales y humanidades, que estén en español o portugués. Surge con el objetivo de ayudar a los investigadores de estas áreas a contar con información confiable y actualizada sobre revistas científicas de sus áreas de trabajo⁵². Al acceder a una revista encontrada en este sitio, podemos visualizar de forma rápida y sencilla que el recurso ya se encuentre indexado por otras bases de datos. Algunas de las bases de datos que el sitio contempla son DOAJ, WoS, REDALyC, Latindex, Núcleo Básico Revistas Argentinas, etc. Otra información provista en la página de cada *journal* de DLP son datos bibliográficos, como ISSN, Título, País, etc., e información con respecto a si el *journal* cuenta con alguna evaluación llevada a cabo por organismos nacionales de ciencia y técnica.

Como puntos positivos de DLP mencionamos cómo considera de valor el hecho de mostrar en qué bases de datos se incluye un recurso, punto que

⁵¹ <http://www.dondelopublico.com/>

⁵² <http://www.dondelopublico.com/public/faq.php>

comparte con la siguiente aplicación que describiremos. Otra característica destacable, es que si bien sólo muestra *journals* de ciencias sociales y humanidades en español y portugués, minimiza un poco algún favoritismo regional que pudiera darse si se incluyeran contenidos en inglés (Moed, [2020](#); Lee et al., [2012](#)).

Desgraciadamente, la actualización de DLP ha sido discontinuada desde 2018, aunque aún se puede consultar la información que poseía hasta ese momento.

2.5.2 Matriz de Información para el Análisis de Revistas

La segunda opción ya mencionada es la Matriz de Información para el Análisis de Revistas (MIAR)⁵³, que da una vista ampliada e integral de métricas de un *journal* en particular que hayamos ingresado en su barra de búsqueda. Para ello, reúne información para su análisis, y las agrupa en grandes áreas científicas, subdivididas en campos académicos específicos. Luego se genera una correspondencia mediante una matriz, y las identifica por ISSN y las bases de datos donde se encuentran indexadas. Incorpora datos de bases de datos como Scopus, WoS, Latindex, REDIB, Crossref, DOAJ, etc.

Al visualizar un *journal* se puede apreciar información con respecto a algunas bases de datos en las que la revista se encuentra y métricas como por ejemplo:

- Artículos por año (Crossref)
- Difusión/ICDS, una medida de cuánta difusión y representación internacional posee el *journal*⁵⁴
- Información bibliográfica
- Métricas extraídas de los sitios mencionados u otros

⁵³ <https://miar.ub.edu/>

⁵⁴ <https://miar.ub.edu/about-icds>

MIAR se acerca mucho más a la idea que resulta del desarrollo de nuestra herramienta, ofreciendo información muy detallada de manera matricial para *journals*. Sin embargo, notamos que no provee información a nivel de *paper*.

Capítulo 3 - Propuesta

En este capítulo presentaremos el propósito de nuestra herramienta, así como las decisiones de diseño tomadas en forma previa al desarrollo de la solución, el análisis de las bases de datos vistas, y las especificaciones técnicas.

3.1 Propósito

La herramienta desarrollada apunta a simplificar, agilizar y dar apoyo al proceso de evaluación de recursos académicos. Para este propósito, recopila métricas e indicadores de calidad e impacto de un recurso determinado, tomadas de diversas bases de datos, para luego ser exhibidas de forma integrada al usuario.

La agilidad, simplificación y apoyo de recopilación de métricas estarán dados por la ayuda que brindará a los investigadores en los aspectos de búsqueda. Al tener a su alcance todo en un mismo sitio, los eximirá de la ardua tarea de realizar búsquedas en sitios externos de forma individual para extraer la información de métricas deseada. Bastará con que un investigador coloque el identificador de un recurso en un campo de búsqueda y aparecerán en pantalla todos los resultados recopilados de métricas de calidad. Se presentarán los datos recopilados en un portal web, dando mayor visibilidad a una gran variedad de métricas, aportando cada una su porción de valor a la evaluación de calidad. Rápidamente se podrán conocer valores de números de citas o factores de impacto, información sobre licencias y acceso abierto, valores de citas relativas y predicciones, e inclusión en bases de datos reconocidas.

Esto presenta una gran ventaja para realizar tareas de análisis de confianza y procesos de evaluación gracias a poseer una mayor disposición de métricas, permitiendo visualizar en conjunto diferentes aspectos de la visión de

calidad de los recursos que no sólo se basan en valores numéricos sino también en políticas de licencia o pertenencia a bases de datos reconocidas.

Sin esta herramienta, los investigadores estarían condicionados a tener que consultar varias bases académicas en caso de querer realizar el análisis más objetivo posible de un recurso, lo cual puede resultar tedioso por el tiempo que implica. Este escenario generalmente derivaría en realizar valoraciones más subjetivas al sólo consultar aquellas bases preferidas por el investigador. HERA no sólo brinda su apoyo en el aspecto de la velocidad de recopilación, sino que de alguna forma pretende poner en foco toda la información que se ha considerado digna de no ser ignorada a la hora de querer interpretar el valor de calidad de un recurso académico.

3.2 Análisis y decisiones de diseño

En esta sección describiremos el análisis realizado que explica las decisiones de diseño tomadas en forma previa al desarrollo de la herramienta. Estas decisiones abarcan el modo de identificar un recurso, la modalidad y medios de acceso a la información de las diferentes bases de datos, y la selección de aquellas que finalmente serán utilizadas para realizar peticiones de datos de métricas.

3.2.1 Identificadores de los recursos

Para recopilar los datos acerca de un recurso académico, primero es necesario buscarlo en diversas bases de datos, de las cuales es posible extraer los metadatos que describen cada recurso.

En una primera instancia fue considerada la alternativa de buscar un recurso en base a su título. Esta opción probó ser de poca utilidad, ya que no sólo pocas bases de datos ofrecen la posibilidad de buscar un recurso por este medio, sino que las que sí otorgan esta posibilidad devuelven información en

base a similitud. El utilizar un título puede, además, generar problemas de ambigüedad cuando los títulos fueran iguales o muy similares. Entonces, cuando un recurso no se encontraba en éstas bases de datos que permitían trabajar de esta forma, se obtenía aquel elemento con el título más parecido, agregando a la herramienta la carga de tener que verificar que la información recibida fuera relevante.

De manera alternativa, y dando solución a estos inconvenientes, se tomó la decisión de utilizar identificadores capaces de localizar de manera unívoca a los recursos académicos: DOI para *papers* e ISSN para *journals*. Todas las bases de datos seleccionadas permiten realizar búsquedas utilizando estos identificadores (según corresponda), y además nos evita tener que verificar que los datos devueltos sean los correctos ya que los identificadores son únicos para cada recurso. Colateralmente, la herramienta es incapaz de encontrar un recurso que no esté identificado por alguno de estos valores, y se introduce la leve complejidad de buscar por estos identificadores, siendo que resulta más sencillo utilizar algo natural como lo es un título.

3.2.2 Modalidad de acceso a la información de las bases de datos

Evaluamos la posibilidad de almacenar información en una base de datos, con el objetivo de enriquecer aún más la evaluación de calidad de los recursos. Esta información se recopilaría con mecanismos de *web crawling* o similares, y los datos cosechados se replicarían en nuestra base de datos. Sin embargo, esta técnica implica almacenar un volumen de datos muy grande para *papers*, aunque es viable para *journals*. Además se deberían realizar actualizaciones periódicas para que la información mostrada fuese coherente con lo que se muestra en las bases de datos originales que la poseen. La ventaja de este medio sería que no se depende del estado de los servicios de las bases de datos, así como de su velocidad de respuesta. Pero analizando el espacio físico necesario y la necesidad de contar con un mecanismo de

sincronización y actualización constante, se optó por la alternativa de trabajo en tiempo real.

El trabajo en tiempo real implica la consulta de datos sobre un recurso en el momento en que se desean conocer. Esto introduce una pequeña carga de trabajo, ya que se debe esperar que las diversas bases de datos respondan con su información y luego se rendericen en la aplicación. No obstante, los tiempos de espera no son extensos y esta alternativa es válida tanto para *papers* como para *journals*. También es destacable que la información mostrada sería siempre la más actualizada, aunque en el caso de que uno de los servicios estuviera caído no veríamos datos del recurso para ese mismo. Sin embargo la herramienta no sería afectada por la falla de un servicio externo, funcionando de todas formas aunque todas las bases estuvieran sin servicio. En estos casos el error de las bases de datos podría ser procesado como un *timeout* para recibir respuesta de los mismos o simplemente un mensaje error, dependiendo de cómo cada una maneje estos casos. Para ambos, en HERA sólo se mostraría que el recurso no fue encontrado.

Pensamos, entonces, que mantener una base de datos propia y actualizada replicando las externas es un trabajo dificultoso. El mantenimiento de una base propia implica tanto tareas de actualización constante como de verificación de consistencia. Además, la actualización tampoco podría realizarse si los sitios externos estuvieran fuera de servicio. En caso de que éstos decidieran cambiar la forma de sus métricas, tendría incidencias en el esquema para el modelado de los datos. Por ello finalmente se decidió por la alternativa de trabajo en tiempo real.

3.2.3 Medios de acceso a las bases de datos

Como mecanismo para obtener datos de las bases de datos seleccionadas, analizamos los tres ya mencionados en la sección [2.4](#).

La primera opción evaluada fue la de utilizar el protocolo OAI-PMH, por su gran uso en el ámbito académico y científico^{55 56 57}. En cuanto al uso de este mismo, encontramos algunos inconvenientes que nos llevaron a revisar las otras dos alternativas:

- Imposibilidad de extracción simple de un registro: para extraer un registro de metadatos, o el ítem completo, necesitamos conocer el identificador único de dicho elemento. Estos identificadores no se pueden conocer de antemano, se obtienen al ejecutar una petición de *ListIdentifiers* o *ListRecords*, y generalmente son propios de cada base de datos
- Alta carga en procesamiento: la imposibilidad de OAI-PMH de trabajar a nivel registro obliga a procesar un conjunto de todos ellos hasta encontrar aquel solicitado. Este procesamiento introduce un *overhead* que puede resultar significativo para la escalabilidad de la herramienta.
- Inconsistencia de datos entre la interfaz OAI-PMH y la interfaz web del repositorio: en algunos casos, utilizando mecanismos manuales improvisados para realizar pruebas con este protocolo, se encontró que algunos recursos estaban indexados en los repositorios y eran visibles en su página web, pero las peticiones por OAI-PMH indicaban que el recurso no se encontraba.

Afortunadamente, en la mayoría de los casos que se contaba con la posibilidad de utilizar OAI-PMH, también se encontraba una API disponible. Por este motivo, nos inclinamos a la opción de utilizar las API, que entendemos son más apropiadas para realizar búsquedas simples. La gran mayoría de los sitios analizados exponen una API pública, e incluso los que poseen servicios

⁵⁵ [Repositorios DSpace con múltiples contextos OAI-PMH - http://sedici.unlp.edu.ar/handle/10915/44572](http://sedici.unlp.edu.ar/handle/10915/44572)

⁵⁶ [Interoperabilidad entre el Repositorio Institucional y servicios en línea en la Universidad Nacional de La Plata - http://sedici.unlp.edu.ar/handle/10915/27406](http://sedici.unlp.edu.ar/handle/10915/27406)

⁵⁷ [Harvesting Full text and Metadata of OpenDOAR through DSpace OAI-PMH : A Framework for Institutional Digital Repositories - https://www.ijlis.org/articles/harvesting-full-text-and-metadata-of-opendoar-through-dspace-oaipmh--a-framework-for-institutional-digital-repositories.pdf](https://www.ijlis.org/articles/harvesting-full-text-and-metadata-of-opendoar-through-dspace-oaipmh--a-framework-for-institutional-digital-repositories)

basados en API con suscripción paga suelen proveer una alternativa gratuita con información reducida, son de fácil uso.

En aquellos sitios en los cuales no se proveía de una API pero resultaban de gran interés para extraer datos, optamos por utilizar la técnica de *web scraping*. Sin embargo, debido a lo intrincado de recolectar datos de un documento HTML, y a las amplias diferencias que tienen sitio a sitio que obliga a desarrollar un algoritmo específico para cada uno, nos hemos limitado a sólo dos sitios en los cuáles aplicar este método. Además, se optó por extraer información limitada, pero relevante, a modo de aligerar la carga de este procesamiento. En caso de que interese ver la información completa de algunas de las bases de datos, la herramienta incluye el enlace correspondiente.

3.2.4 Selección de bases de datos y métricas

3.2.4.1 Análisis sobre Scopus

Scopus ofrece numerosas API para acceder a diferentes métricas o metadatos de los recursos que posee⁵⁸, lo cual resulta muy conveniente para hacer análisis de las mismas en la herramienta. En el análisis de las APIs provistas encontramos que algunas son de acceso bajo suscripción paga, otras están limitadas bajo políticas de uso, propósitos específicos o cuota limitada. La que llamó interés para nuestros propósitos es la API de Journal Metrics⁵⁹, que contiene métricas de CiteScore, Source Normalized Impact per Paper (SNIP) y SCImago Journal Rank (SJR), accesible por métodos HTTP GET mientras se cuente con una APIKey. Ésta se otorga de manera gratuita a aquellos que se registran en el sitio web de Elsevier, aunque se debe respetar una cuota de uso. Además, la API otorga algunos metadatos de otras índoles, pero de un carácter diferente al que buscamos para la herramienta. CiteScore

⁵⁸ [Elsevier Developer Portal - https://dev.elsevier.com/use_cases.html](https://dev.elsevier.com/use_cases.html)

⁵⁹ Journal Metrics - [Elsevier Developer Portal - https://dev.elsevier.com/journal_metrics.html](https://dev.elsevier.com/journal_metrics.html)

revisa la cantidad de citas y documentos recibidos en una ventana de 4 años previa. En base a estos datos es capaz de calcular un coeficiente que sirve para estimar una posición en el ranking futuro respecto de sus citas actuales, denominado CiteScoreTracker.

Estas métricas nos resultaron útiles y fueron incorporadas para ser visualizadas en la herramienta a modo de ser comparadas, en caso que se considere necesario, con las de otros sitios en los cuáles el recurso también puede estar indexado y valorado de manera diferente.

3.2.4.2 Análisis sobre WoS

En WoS también encontramos que para saber si un *journal* se encuentra indexado, u obtener métricas de los mismos, existen numerosas API accesibles desde la web de Clarivate⁶⁰. Lamentablemente, las que refieren a WoS no proveen de manera pública una *API key* para su uso. Sin embargo, encontramos de utilidad utilizar la API pública del sitio web Master Journal List (Web of Science Group)⁶¹, que nos permite recuperar un número acotado, pero suficiente, de métricas de WoS mediante solicitudes HTTP POST a la API. Cabe destacar que existen casos los cuales un recurso puede ser obtenido a través de la API, pero haciendo una búsqueda manual del mismo en la página de Master Journal List el recurso no aparece entre los resultados. Entre las métricas que la API otorga mostramos algunas de las que consideramos más se destacan:

- Frecuencia de publicación
- Ediciones por año
- País de origen y año de inicio de publicación
- Idiomas del recurso

⁶⁰ [Clarivate APIs](https://developer.clarivate.com/apis) - <https://developer.clarivate.com/apis>

⁶¹ [Web of Science Master Journal List - WoS MJL by Clarivate](https://mjl.clarivate.com/home) - <https://mjl.clarivate.com/home>

- Categorías (entre las que se encuentra a cuál de los niveles de índices pertenece el recurso en WoS)

3.2.4.3 Análisis sobre DOAJ

Existen algunos métodos por los cuales podemos extraer información de DOAJ. Por un lado, posee acceso a grandes volúmenes de metadatos mediante el uso del protocolo OAI-PMH⁶². También existe la posibilidad de utilizar una API⁶³ pública para extraer información de recursos de forma individual o grupal mediante búsquedas bajo identificadores como DOI o ISSN. Mediante consultas a la API conseguimos extraer numerosos datos del recurso, de los cuales nos enfocamos en:

- Tipo de licencia del recurso.
- Costos de publicación del journal
- Existencia de un DOAJ Seal asociado al recurso

3.2.4.4 Análisis sobre Latindex

Latindex no posee una API pública para la obtención de datos de los recursos en su base de datos. En los intentos de realizar *web scraping*, hallamos que además utiliza métodos para generar las vistas en su página web que nos imposibilitan la recopilación de datos a partir del documento HTML. Por esto hemos decidido prescindir de la inclusión de datos provenientes de Latindex para el desarrollo de la herramienta.

⁶² OAI-PMH – DOAJ - <https://doaj.org/docs/oai-pmh/>

⁶³ DOAJAPI - Documentación - <https://doaj.org/api/v2/docs>

3.2.4.5 Análisis sobre REDIB

En el caso de REDIB tampoco se cuenta con una API para acceder a sus datos de forma externa, pero gracias a la sencilla organización del documento, y dado que es la única alternativa disponible, hemos realizado *web scraping* del sitio. Al hacerlo, cabe notar que si bien hacer un scraping directo del Ranking REDIB y tomar todas las métricas exhibidas puede parecer la opción ideal, en realidad puede conducir a una pérdida de información. Al realizar el scraping, notamos que varios de los *journals* indexados por REDIB, no necesariamente aparecían en el ranking. Esto se debe a que el ranking se arma a partir de métricas provistas por Clarivate, lo que implica que para visualizarse, la revista debe haber sido indexada previamente por Clarivate.

El scraping se logró, entonces, por medio de una búsqueda del recurso y de su página individual con solicitudes HTTP GET a la URL de búsqueda utilizando los parámetros adecuados. De los resultados obtenidos, extrajimos la página del recurso de la cual se muestra la información sobre los indicadores de calidad editorial y aquella respecto a las acreditaciones que el recurso posea. Cuando fue posible, también obtuvimos el widget exportable con la información resumida del Ranking REDIB, recordando que éste puede no aparecer por lo mencionado anteriormente respecto del ranking de Clarivate.

En rigor de verdad, podríamos haber extraído algunas métricas o información adicional que se visualiza en la página de cada recurso, como los gráficos estadísticos de citas y crecimiento del *journal*, pero elegimos no hacerlo ya que consideramos que replicar la funcionalidad de la web de REDIB no es nuestro objetivo, sino el de mostrar el hecho de que un recurso esté indexado en el sitio como una “métrica” adicional para el investigador. Las pocas métricas extraídas son a modo de ejemplo y simbólico, como para intentar captar la atención del investigador que se encontrará analizando la calidad de un *journal*, y tendrá disponible el enlace en caso de querer acceder a la información completa del recurso en el sitio original.

3.2.4.6 Análisis sobre Crossref

En cuanto a medidas de calidad, pudimos extraer algunas métricas o información interesante mediante el uso de la API REST⁶⁴ pública de Crossref, la cual está disponible sin necesidad de una membresía y de forma gratuita.

Encontramos a Crossref cumpliendo su objetivo natural incluso en nuestra herramienta, siendo de gran utilidad para ser el punto principal de referencia del cual obtendremos datos generales del recurso buscado, y otros que lo vinculan con otros recursos. También lo utilizamos como punto de partida en los llamados a las numerosas API REST, ya que luego de numerosas pruebas percibimos que en el mayor de los casos la búsqueda en Crossref daba resultados positivos, mientras que los recursos a veces no eran hallados exitosamente en las bases de datos restantes.

3.2.4.7 Análisis sobre SCImago

SCImago no posee una API REST ni otra forma conveniente de extraer su información, salvo un widget exportable que se puede extraer de la página individual de cada *journal*. Dado que encontramos que el documento estaba organizado de manera sencilla, fue posible realizar scraping del sitio para extraer algunas de las métricas, como el H-index, la cobertura, el país y el widget exportable, las cuales se incorporaron en la herramienta junto con lo extraído de Scopus para su fácil análisis. Decidimos no tomar la información de los gráficos de citas y documentos que posee, ya que sentimos que el mérito de los mismos pertenece a SCImago, y nos pareció más apropiado mostrar lo suficiente como para captar la atención de un investigador, y que luego éste pueda decidir ir al portal original del que la información proviene y así analizar los datos más detenidamente.

⁶⁴ [API CROSSREF - Swagger UI - https://api.crossref.org/swagger-ui/index.html](https://api.crossref.org/swagger-ui/index.html)

3.2.4.8 Análisis sobre Google Scholar

GS no cuenta con una API REST o un método simple para la extracción de los datos en grandes contenidos, por lo que tendríamos que haber recurrido al scraping, como en los casos anteriores. En el caso de GS esto revisamos que el proceso podía resultar tedioso y poco escalable, hecho evidente al analizar cómo algunos autores tampoco han recurrido a estos métodos (Freyne et al., [2010](#); Martín-Martín et al., [2018](#)). Más allá de la labor extra que constituye el *web scraping*, Google, y por consiguiente GS, utilizan herramientas que generan los tags de su documento HTML de forma automática, los cuales suelen cambiar frecuentemente. En forma similar, la estructura general de un mismo documento sufre cambios de manera continua. Por la complejidad y dificultad de sostener un mecanismo de extracción de datos de la web de GS, no lo consideramos para el desarrollo de la herramienta.

3.2.4.9 Análisis sobre Microsoft Academic

Microsoft Academic cuenta con poderosas herramientas para compilar y dar un sentido mayor a sus datos, como lo son MAG⁶⁵ y MAKES⁶⁶. Sin embargo, particularmente para hacer una aplicación como la que nos proponemos, MAKES pareció ser la opción más adecuada.

MAKES posee una API REST para acceder a los datos con sus propios métodos específicos. Los datos son extraídos con un algoritmo de machine learning desde los resultados proporcionados por Bing. Desgraciadamente, el uso de MAKES requiere una suscripción en Azure, por lo que recurrimos a otra alternativa, que es la Project Academic Knowledge API. Ésta cuenta con una

⁶⁵ [Microsoft Academic Graph \(MAG\) Analytics](#) -

<https://www.microsoft.com/en-us/research/project/microsoft-academic-graph-mag-analytics/>

⁶⁶ [About Microsoft Academic Knowledge Exploration Service - Microsoft Academic Services](#) -

<https://docs.microsoft.com/en-us/academic-services/knowledge-exploration-service/introduction?view=makes-3.0>

suscripción gratuita pero con usos limitados en transacciones y demoras específicas para cada distinto método utilizado de la API REST, que permite acceder a información académica. Dado que la herramienta desarrollada no realiza peticiones de manera excesiva, la encontramos conveniente, y de ella es que pudimos conseguir los datos de métricas mencionados en [2.2.9](#).

Un hecho a tener en cuenta de esta API REST es que no soporta búsqueda de *journals* por ISSN, que es el identificador por el cual recopilamos las métricas de los mismos. Es por ello que limitamos el uso de MA al de la búsqueda de métricas de *papers*.

3.2.4.10 Análisis sobre Semantic Scholar

Para extraer datos de Semantic Scholar, el sitio ofrece una API REST⁶⁷ de libre acceso de la cual pudimos obtener las siguientes métricas:

- Número de citas
- Número de citas en los últimos 3 años
- Número de citas que son influyentes
- Información bibliográfica (título, *abstract*, autores, referencias)
- URL del recurso en semantic scholar

El uso de esta API está limitado a 100 peticiones cada 5 minutos, lo cual es suficiente para el uso dado a nuestra herramienta. En caso de ser necesaria una cuota mayor, se puede completar un formulario para solicitar un permiso de acceso privilegiado. Encontramos de gran utilidad poseer información de número de citas de varios sitios diferentes, con políticas de inclusión diferentes, para combatir las imparcialidades que pueden existir en este valor, como mencionamos en el capítulo [2.1.1](#).

⁶⁷ <https://www.semanticscholar.org/product/api#Fetch-Paper>

3.2.4.11 Análisis sobre servicios de métricas alternativas

Los proveedores de altmetrics utilizados para la herramienta son Altmetric y Dimensions. Esta decisión fue tomada en base a que por un lado ofrecen información diferente una de la otra, y por otro el hecho de que son atractivas y fáciles de importar. Ambas ofrecen tanto una API REST para extraer datos y un *badge* exportable que nos muestra la misma información pero de manera sintética y atractiva. Luego un usuario puede hacer clic en estos *badges* y ser llevado a una página donde se hace un análisis más detallado del significado de las métricas que mostraba la misma.

Los datos de Dimensions, en particular, son accesibles por dos medios: la Dimensions Metrics API⁶⁸, que por medio de una API REST devuelve las métricas en formato JSON, y la Dimensions Badge, que presenta dichas métricas de manera gráfica y que pueden ser integradas fácilmente a cualquier página web. Existen numerosas API provistas por Dimensions además de la mencionada, pero para utilizarlas es necesario una suscripción.

Esperamos que la incorporación de los *badges* de Altmetric y Dimensions sean de utilidad para que los investigadores posean una vista más amplia de más métricas, y que éstas proporcionen una opinión más cercana a la que un investigador común podría emitir que a la que entidades evaluadoras ya mencionadas podrían definir.

3.4.2.12 Decisiones sobre las métricas de citas y factor de impacto

De los datos disponibles respecto del número de citas, no extrajimos el gráfico de citas SCImago por lo mencionado en [2.2.7](#), y destacamos que si bien Scopus posee esta información en su sitio, su API sólo nos devolvía el

⁶⁸ [Dimensions Metrics API Documentation - https://figshare.com/articles/journal_contribution/Dimensions_Metrics_API_Documentation/5783694](https://figshare.com/articles/journal_contribution/Dimensions_Metrics_API_Documentation/5783694)

CiteScore ya calculado. De todas formas este valor nos resultó de gran fortaleza por ser una medida relativa al total de documentos. Los demás fueron incluidos de forma directa para poder ser comparados entre ellos.

En cuanto al factor de impacto, en nuestro caso la medida utilizada fue aquella provista por la API de Scopus denominada *SNIP* (Source Normalized Impact per paper), complementada por la información de SCImago Journal Rank. El *SNIP* corrige las desviaciones de impacto cuando se trata con áreas temáticas que tienden a tener menor cantidad de citas, ya que asigna un peso mayor al valor de una cita en estas áreas poco citadas.

3.3 Diseño y desarrollo

Para la implementación de la herramienta se desarrolló una aplicación web, que está dividida en una aplicación front-end, desarrollada con ReactJS, y en una aplicación back-end, desarrollada con NodeJS, como se muestra en la Figura 15. La aplicación de back-end tiene como objetivo servir de puente entre la aplicación front-end y los diversos servidores que proveen las métricas de los recursos académicos, concentrando de esta forma la mayor carga de trabajo en sí misma. La aplicación de front-end realiza peticiones a la de back-end mediante una API REST, recopilando los resultados, y es la encargada de generar las visualizaciones de los mismos en un portal web.

En esta sección detallaremos el diseño de las mismas, de la tecnología utilizada y cuestiones relevantes a la hora del desarrollo y prueba de la herramienta.

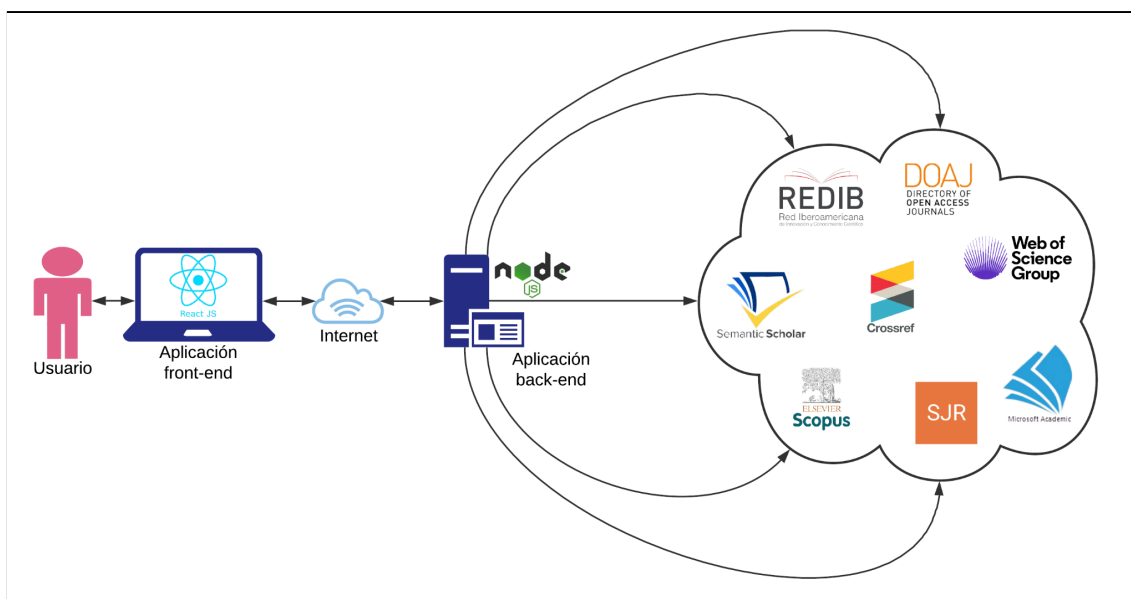


Figura 15. Esquema de comunicación entre las aplicaciones

3.3.1 Diseño de aplicación back-end

Esta aplicación fue desarrollada con NodeJS y utilizando el framework *express*⁶⁹ para definir los endpoints de la API. También incorpora el módulo *node-fetch* para ser capaz de realizar peticiones por HTTP a las diferentes bases de datos y procesar las respuestas. Consta de dos funcionalidades principales:

- Almacenamiento de constantes de acceso, requeridas para acceder a las APIs de Scopus y Microsoft Academic
- Exposición de una API que brinde la funcionalidad necesaria para que la aplicación de front-end pueda delegar las solicitudes de datos a las diferentes bases de datos

⁶⁹ <https://expressjs.com/>

Tabla 2. Endpoints de la API

Método HTTP	Endpoint	Parámetros	Descripción
POST	/api/getContentsDefault	- req: objeto con requerimientos. Se extrae de forma general el campo "url" del body del mismo para indicar a qué API se desea hacer una petición de contenido	Retorna una representación en JSON de la información que responde el sitio web indicado por el campo "req.body.url"
POST	/api/getContentsMicrosoft		Realiza lo mismo que /api/getContentsDefault, agregando la clave de suscripción de Microsoft que está almacenada en esta aplicación
POST	/api/getContentsScopus		Realiza lo mismo que /api/getContentsDefault, agregando la clave de suscripción de Scopus que está almacenada en esta aplicación
POST	/api/getContentsHtml		Retorna un objeto cuyo campo "html" es la representación en formato string del HTML del sitio web "req.body.url". Utilizado para tareas de <i>web scraping</i> .

3.3.2 Diseño de aplicación front-end

La aplicación de front-end fue realizada con ReactJS, con el objetivo de elaborar una aplicación web en la que se visualicen las métricas recopiladas. Esta es la encargada de interpretar el pedido de un usuario y comunicarse con la aplicación de back-end para obtener datos de métricas relevantes.

ReactJS es una librería de JavaScript, que permite desarrollar interfaces de usuario interactivas simples. El modelo en el cual se presentan los datos en

aplicaciones que utilizan ReactJS está basado en componentes, los cuales encapsulan su propio estado y forma de renderizarse, además de funciones propias que pudieran tener cada uno⁷⁰. En una página desarrollada con ReactJS, todo lo que se visualiza son componentes, y la filosofía principal es la de no recargar toda la página cuando sólo una porción de la información en ella cambia, sino sólo las partes competentes. Para ello, ReactJS detecta cambios en los estados de sus componentes de manera automática y eficiente, lo que desata re-renderizaciones sólo de aquellos en los cuáles el estado es afectado.

Por ser una librería de JavaScript, la parte lógica y funcional de la aplicación puede ser programada con JavaScript puro o incorporar cualquier otra librería o framework que se desee. Para el caso de esta aplicación, en la parte funcional el único framework utilizado es la Web API Fetch⁷¹, similar al uso de XMLHttpRequest.

3.3.3 Componentes de la aplicación

Los componentes utilizados en el desarrollo de la aplicación son los que se visualizan en la Figura 16. Además existen pequeñas librerías de código *helpers* que ayudan a organizar la funcionalidad de la misma, que si bien no son visualizadas en el diagrama al no hacer a la arquitectura, serán explicadas en las secciones de los respectivos componentes.

⁷⁰ [ReactJS: An Open Source JavaScript Library for Front-end Development - https://www.theseus.fi/handle/10024/130495](https://www.theseus.fi/handle/10024/130495)

⁷¹ [Fetch API - Web APIs | MDN - https://developer.mozilla.org/en-US/docs/Web/API/Fetch_API](https://developer.mozilla.org/en-US/docs/Web/API/Fetch_API)

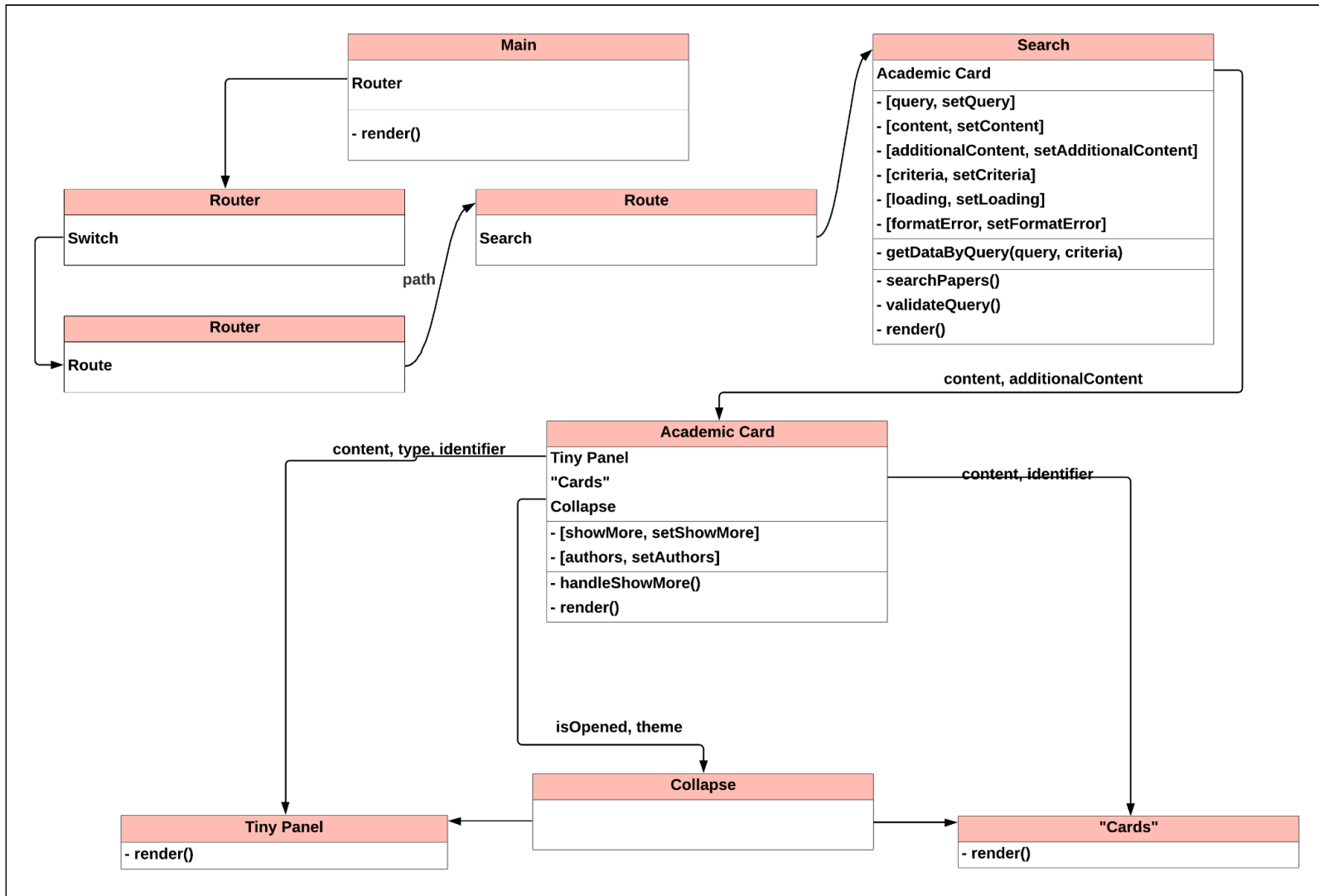


Figura 16. Diagrama de componentes de la aplicación front-end

Main: componente principal de la aplicación, donde se configuran las diferentes rutas que la misma pudiera poseer y cómo se manejan los accesos a éstas.

Router: componente que capta los cambios en las rutas del navegador y se comunica con la aplicación para actualizar el estado del *path* actual.

Switch: está contenido dentro del Router. Cuando se actualiza la URL actual, el Switch se renderiza, y busca en sus hijos “Route” el que tenga un *path* coincidente con la URL.

Route: contenido en el Switch, este componente define qué se renderiza de acuerdo a la coincidencia entre la URL actual y su propiedad *path*. En nuestra aplicación sólo posee la ruta “/” para renderizar el resto de la aplicación, ya que no se ofrece funcionalidad adicional a la ya mencionada. También se incluye un mecanismo de redirección para que las rutas diferentes lleven siempre al *path* “/”.

Search: es el principal componente funcional para el propósito de búsqueda de métricas. Visualmente se renderiza como una barra de búsqueda con un botón para buscar y dos elementos para seleccionar el tipo de búsqueda que se desea realizar (se proveerán capturas en la siguiente sección).

Este componente es el que valida formatos de búsqueda y luego comienza la tarea de recopilar las métricas de un recurso, construyendo las peticiones a realizar valiéndose de scripts importados y *helpers*. Estos scripts construyen las URL de búsqueda apropiadas, se comunican con los correspondientes endpoints del back-end para que éste obtenga los datos y los retorne, y arman los datos en forma entregable al componente Search. Un *helper* provee de expresiones regulares para validar los formatos de DOI e ISSN ingresados en el search, y un mensaje de error es renderizado cuando se detecta un mal formato. Cuando la información es recibida, la organiza en un formato apropiado y se la envía al siguiente componente en la cadena, *AcademicCard*, para que la renderice apropiadamente.

AcademicCard: es un componente hijo de Search. Además de mostrar información general bibliográfica de un recurso, es el contenedor principal de todos los componentes donde se visualiza la información de las métricas obtenidas de cada sitio.

AcademicCard contiene a los componentes *TinyPanel*, *Collapse* y las “Cards”. Mediante ellos, y decide qué información debe renderizarse de acuerdo a las propiedades que recibe del *Search* y en base a su estado. También provee la funcionalidad de poder elegir entre una vista simplificada de las métricas o una detallada mediante un botón. La información resumida se renderiza en el componente *TinyPanel* y la detallada en sus respectivas “Cards”.

Collapse: este componente es importado de react-collapse⁷² con el objetivo de proveer animaciones de renderizado dinámico. “Encierra” a los componentes *TinyPanel* y “Cards”, mostrándolos u ocultándolos de acuerdo a su propiedad *isOpen*. También es posible configurar sus estilos mediante su propiedad *theme*.

“**Cards**”: nos referimos de esta forma genérica a esta familia de componentes ya que en realidad no existe como tal, sino que existe un componente “Card” diferente para cada base de datos consultada, y éste contiene la información de las métricas de la misma. Aun así, el propósito de ellos es el mismo en todos los casos: dar una visualización general e integral de todas las métricas recopiladas de una base de datos.

Se han desarrollado de manera distinta e individual debido a las grandes diferencias entre métricas que muchas poseen, lo que demanda diferentes formas de evaluación y renderizado. Para visualizar estas “Cards” se debe activar la vista detallada mediante el botón mencionado de la *AcademicCard*.

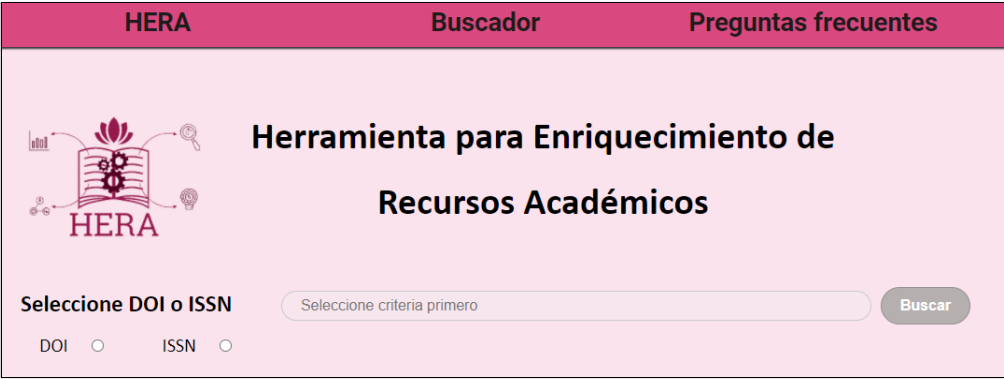
TinyPanel: diseñado con el propósito de dar una vista resumida de las métricas más importantes de todas las bases de datos. En este componente

⁷² [react-collapse](https://www.npmjs.com/package/react-collapse) - <https://www.npmjs.com/package/react-collapse>

encontramos algunas métricas de cada base de datos consultada, organizada en formas de pequeñas “medallas” para cada base de datos. Su objetivo es el de poder hacer un rápido y simple análisis de métricas al poseer información en forma compacta pero relevante, cautivando tal vez un interés inicial del investigador. El *TinyPanel* se visualiza por defecto en la *AcademicCard* luego de realizar una búsqueda, y representa la vista simplificada ya mencionada.

3.4 Interfaz gráfica de usuario

La aplicación cuenta con una única interfaz de usuario en la cual se destaca la barra de búsqueda donde los usuarios deben introducir el identificador del recurso que se quiere buscar, como se muestra en la Figura 17.



The image shows a web interface for the HERA tool. At the top, there is a navigation bar with three items: 'HERA', 'Buscador', and 'Preguntas frecuentes'. Below this, the main content area has a light pink background. On the left, there is a logo for HERA, which consists of a stylized book with a flower-like shape above it and the word 'HERA' below. To the right of the logo, the text 'Herramienta para Enriquecimiento de Recursos Académicos' is displayed in a bold, black font. Below the title, there is a search form. The form starts with the text 'Seleccione DOI o ISSN' followed by two radio buttons labeled 'DOI' and 'ISSN'. To the right of these is a search input field with the placeholder text 'Seleccione criterio primero'. At the end of the input field is a button labeled 'Buscar'.

Figura 17. Interfaz de búsqueda de la herramienta con input de búsqueda y selectores de tipo de identificador (DOI e ISSN)

En ella, el usuario debe seleccionar el tipo de búsqueda, que puede ser por DOI o por ISSN, y luego colocar en la barra el identificador correspondiente. En las siguientes secciones mostraremos el flujo de búsqueda de cada tipo de recurso.

3.4.1 Búsqueda de *journal*

Al introducir un ISSN en la barra de búsqueda, se realiza la búsqueda de un *journal*. De este obtenemos primero la información sobre la editorial y link al recurso, así como la vista del resumen de métricas recopiladas de los diferentes sitios, visualizado en la Figura 18. Deslizándose hacia abajo se visualiza la vista resumida de las métricas del *journal* que fueron recopiladas en las bases de datos. Contamos con un botón para expandir la información y así apreciar la totalidad de su contenido, como mostramos en las Figuras 19, 20 y 21.

The screenshot shows the HERA search interface. At the top, there are navigation links for 'HERA', 'Buscador', and 'Preguntas frecuentes'. The main heading is 'Herramienta para Enriquecimiento de Recursos Académicos'. Below this is a search bar with the text 'Ingrese ISSN del contenido a analizar' and the value '1989-4007'. There are radio buttons for 'DOI' and 'ISSN', with 'ISSN' selected. A 'Buscar' button is to the right. The search results are displayed in a white box with a red border. The title is 'The European Journal of Psychology Applied to Legal Context'. Below the title, it says 'Tipo de recurso: Journal' and 'Editorial: Sociedad Española de Psicología Jurídica y Forense'. There is a link 'Click aquí para ir al recurso'. The results are presented in a grid of six cards: Crossref (25 artículos publicados), DOAJ (Directory of Open Access Journals), Ranking 2020 (Calificación Global: 84,37), Web of Science Group (Indexado por Web of Science), Scopus (CiteScore 10.7 - 2020, CiteScore Tracker 11.8 - 2021), and SJR (SJR 2020 7.02). A 'Ver más' button is at the bottom of the results box.

Figura 18. Resultados generales de la búsqueda de un *journal*

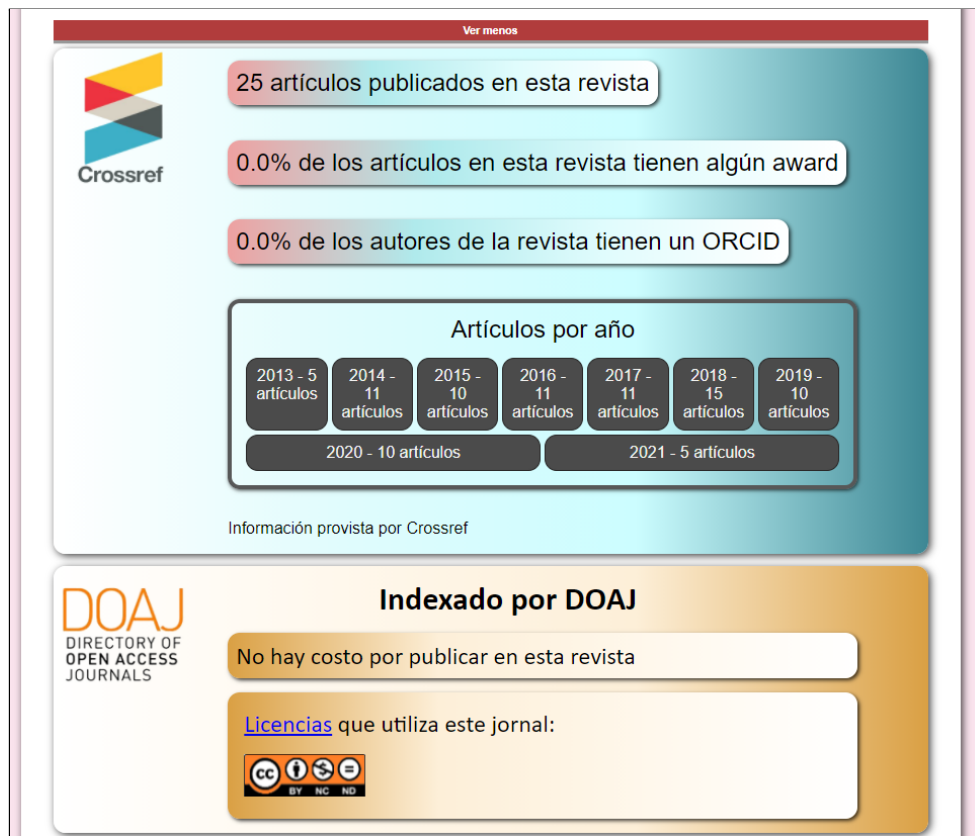


Figura 19. Vista expandida de métricas de un *journal* (1)

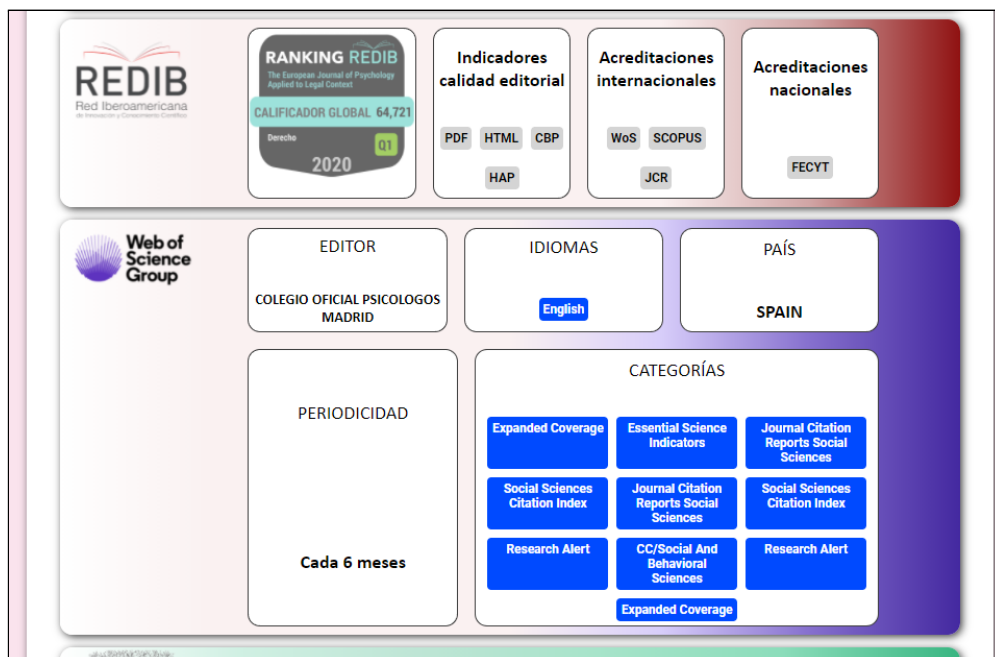


Figura 20. Vista expandida de métricas de un *journal* (2)

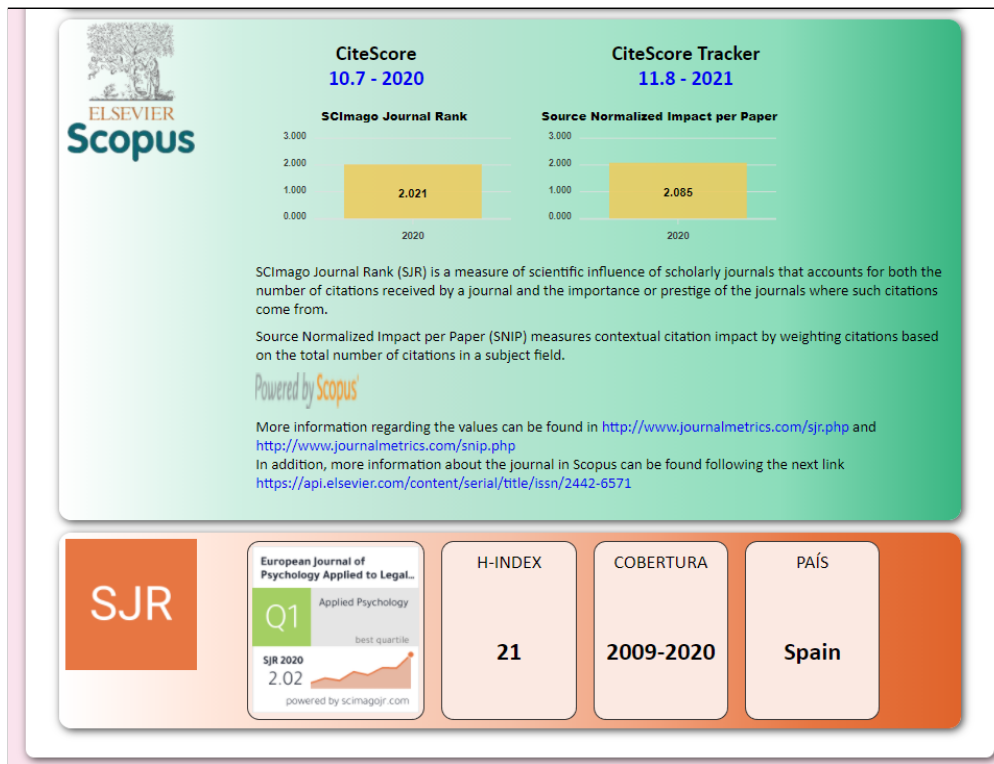


Figura 21. Vista expandida de métricas de un *journal* (3)

3.4.1 Búsqueda de *paper*

Luego de introducir un DOI en la barra de búsqueda y presionar el botón buscar, las métricas son recopiladas de las diferentes bases de datos y se renderizan en pantalla, como se muestra en la Figura 22. Entre los resultados se pueden apreciar algunos metadatos de los recursos, como ser título, tipo de recurso, abstract, entre otros, así como el nombre del *journal* donde se publicó el *paper* y un link para dirigirnos a la página original del mismo. También, de ser posible, la aplicación intentará recolectar métricas del *journal* donde se encuentra el *paper* para mostrarlas en conjunto, como se ve en la Figura 23. Para ello, se extrae de los metadatos recuperados el ISSN del *journal* y se ejecuta de manera automática y en background una búsqueda por ISSN. De esta forma se amplía el contexto de evaluación de un *paper*, al poder visualizar sus métricas de forma aislada y también tener acceso a las del *journal* y así entender qué influencia podría tener en las del primero. En general, la calidad

de un *paper* está asociada a la del *journal* en el que está publicado. Al presionar el botón “Ver más” que se ve en esta última imagen, se genera una vista expandida de las métricas en la cual vemos información en forma más detallada (en aquellos casos donde las bases de datos nos provean algún dato adicional), ilustrado en la Figura 24.

HERA Buscador Preguntas frecuentes

Herramienta para Enriquecimiento de Recursos Académicos

Ingrese DOI del contenido a analizar 10.3389/fenvs.2020.581591 **Buscar**

DOI ISSN

Analysis of Water Pollution Using Different Physicochemical Parameters: A Study of Yamuna River

Tipo de recurso: Journal article

Autores: Sharma Rishi, Kumar Raghendra, Sotopathy Suresh Chandra, Al-Arazi Nadiyah, Singh Krishna Kant, Mahapatra Rajendra Prasad, Agarwal Anuj Kumar, Le Hiep Van, Pham Binh Thoi.

Título de la revista: Frontiers in Environmental Science

Editorial: Frontiers Media S.A.

Año de publicación: 2020

ABSTRACT: The Yamuna river has become one of the most polluted rivers in India as well as in the world because of the high-density population growth and speedy industrialization. The Yamuna river is severely polluted and needs urgent revival. The Yamuna river in Dehradun is polluted due to exceptional tourist activity, poor sewage facilities, and insufficient wastewater management amenities. The measurement of the quality can be done by water quality assessment. In this study, the water quality index has been calculated for the Yamuna river at Dehradun using monthly measurements of 12 physicochemical parameters. Trend forecasting for river water pollution has been performed using different parameters for the years 2020-2024 at Dehradun. The study shows that the values of four parameters namely, Temperature, Total Calcium, TDS, and Hardness are increasing yearly, whereas the values of pH and DO are not rising heavily. The considered physicochemical parameters for the study are TDS, Chloride, Alkalinity, DO, Temperature, CO₂, BOD, pH, Magnesium, Hardness, Total Calcium, and Calcium. As per the results and trend analysis, the value of total calcium, temperature, and hardness are rising year by year, which is a matter of concern. The values of the considered physicochemical parameters have been monitored using various monitoring stations installed by the Central Pollution Control Board (CPCB), India.

[Click aquí para ir al recurso](#)

Crossref 7 citas

DOAJ DIRECTORY OF OPEN ACCESS JOURNALS 7 citas

Microsoft Academic Predicciones de citas

Semantic Scholar 7 citas

Sin información de tendencias

Se habla del tema

Figura 22. Resultados generales de la búsqueda de un *paper* (vista resumida)

Publicado en: Frontiers in Environmental Science

Crossref 965 artículos publicados

DOAJ DIRECTORY OF OPEN ACCESS JOURNALS 7 citas

REDIB Recurso no encontrado

Web of Science Group Indexado por Web of Science

ELSEVIER SCOPUS CiteScore 4.4 - 2020 CiteScore Tracker 4.7 - 2021

SJR Frontiers in Environmental Science Q1 SJR 1.23

[Ver más](#)

Figura 23. Métricas correspondientes al *journal* donde se encuentra el *paper* (vista resumida)



Figura 24. Vista expandida de métricas de un *paper*

Capítulo 4 - Pruebas y Resultados

En este capítulo expondremos el funcionamiento general de la herramienta, incluyendo capturas que visualicen la interfaz general por la que el usuario navega y su utilización. También se incluirán capturas con ejemplos de los diferentes escenarios posibles al buscar un artículo y una revista, confirmando que la información que aparece está en las bases de datos realmente, y que la que no aparece tampoco lo está.

4.1 Pruebas y resultados en búsquedas de *journals*

En esta sección mostraremos cómo se visualizan los resultados al realizar búsqueda de información de *journals*, y cómo se vería en los sitios de los cuales es extraída, así como el caso en el que no se hallara en los mismos. Con este propósito, realizamos búsquedas de métricas de diversos *journals*, los cuales pueden aparecer en Crossref, DOAJ, REDIB, Web of Science, Scopus y SCImago Journal Ranking (SJR).

Al escribir el identificador, si el campo de ISSN del *journal* a buscar se ingresa de forma incorrecta, aparece en pantalla un error que indica este evento, ilustrado en la Figura 25.

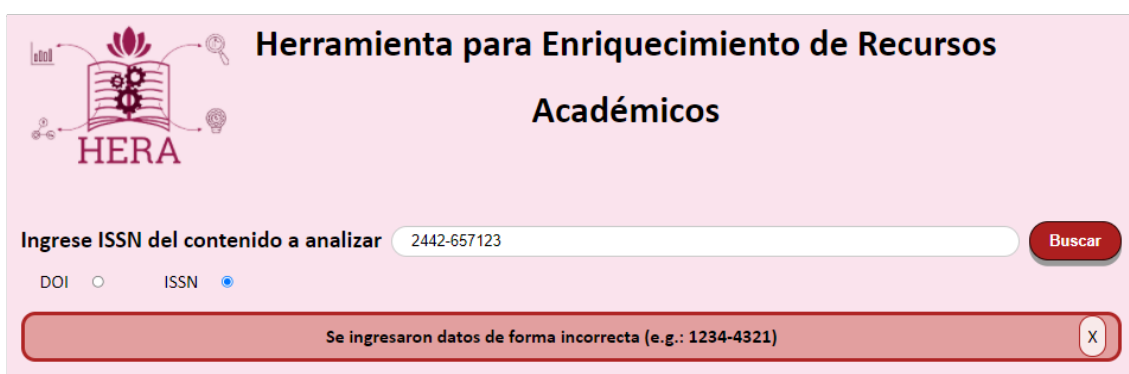


Figura 25. Error de formato de ISSN

4.1.1 Resultados de Crossref

La información recibida desde Crossref para métricas de un *journal* se muestra en la vista resumida como lo exhibe la Figura 26. A su lado observamos que el recurso se encuentra también indexado por el sitio web de Crossref. Es importante en Crossref que la opción “Journal Issue” se encuentre seleccionada en la lista de filtros para buscar puntualmente publicaciones del *journal*, y no mezclar resultados con *papers* donde es mencionado. En la Figura 27 vemos que cuando un recurso no se encuentra en Crossref, en la herramienta aparece con el ícono del logo más transparente, además de darnos el mismo mensaje que con los *papers*. En el sitio web, realizar la búsqueda de este *journal* nos devuelve “Internal Server Error”.

IET Cyber-Systems and Robotics
 Tipo de recurso: Journal
 Editorial: Wiley
 Click aquí para ir al recurso

62 artículos publicados

DOAJ
 DIRECTORY OF OPEN ACCESS JOURNALS

REDIB
 Recurso no encontrado

Web of Science Group
 Indexado por Web of Science

ELSEVIER
 Scopus
 Recurso no encontrado

SJR
 Recurso no encontrado

Crossref

2631-6315

TYPE: Journal Issue (5)

YEAR: 2021 (3), 2020 (2)

PUBLICATION: IET Cyber-Systems and Robotics (5)

PUBLISHER: Institution of Engineering and Technology (IET) (5)

FUNDER NAME

Showing results for ISSN matching 2631-6315

Journal Issue published Sep 2020 in IET Cyber-Systems and Robotics volume 2 issue 3
<https://doi.org/10.1049/csy2.v2.3>

Journal Issue published Dec 2020 in IET Cyber-Systems and Robotics volume 2 issue 4
<https://doi.org/10.1049/csy2.v2.4>

Journal Issue published Mar 2021 in IET Cyber-Systems and Robotics volume 3 issue 1

Figura 26. Métricas de Crossref para *journal* en la herramienta y en el sitio real

Revista Iberoamericana de Psicología y Salud
 Tipo de recurso: Journal
 Editorial: Elsevier Doyma
 Click aquí para ir al recurso

Recurso no encontrado

DOAJ
 DIRECTORY OF OPEN ACCESS JOURNALS

RANKING REDIB
 2020

Web of Science Group
 Indexado por Web of Science

ELSEVIER
 Scopus
 CiteScore 4.0 - 2020
 CiteScore Tracker 3.9 - 2021

SJR
 Revista Iberoamericana de Psicología y Salud
 Q2
 0.69

Internal Server Error

https://search.crossref.org/?q=1989-9246&from_ui=yes

Figura 27. Métricas de Crossref para *journal* no halladas por la herramienta ante su ausencia en esta base de datos

4.1.2 Resultados de DOAJ

Los resultados de DOAJ incluyen el hecho de poseer el DOAJ Seal y las licencias del *journal*, las cuales se ven en el panel de la vista resumida de la forma en que se observa en la Figura 28, información que verificamos que existe también al acceder a su página original de DOAJ. Para verificar que un recurso que no muestra métricas de DOAJ realmente no se encuentra allí, realizamos una búsqueda por ISSN en el sitio y corroboramos su ausencia, mostrado en la Figura 29.

The image shows two side-by-side screenshots of the DOAJ interface. The left screenshot displays a summary card for 'International Journal of Advances in Intelligent Informatics' with various metrics: Crossref (70 articles published), DOAJ Seal, REDIB (Recurso no encontrado), Web of Science (Indexado por Web of Science), Scopus (CiteScore 2.1 - 2020, CiteScore Tracker 2.6 - 2021), and SJR. The right screenshot shows the journal's main page with buttons for 'DOAJ Seal', 'Website', and 'ISSN Portal'. A purple arrow points from the DOAJ Seal icon in the summary to the DOAJ Seal button on the main page.

Figura 28. Métricas de DOAJ para *journal* en la herramienta y en el sitio real

The image shows two side-by-side screenshots of the DOAJ search interface. The left screenshot shows a search result for 'Nereis Interdisciplinary Ibero-American Journal of Methods Modelling and Simulation' with metrics: Crossref (15 articles published), DOAJ Seal, REDIB (Indicadores de calidad editorial), Web of Science (Recurso no encontrado), and SJR (Recurso no encontrado). The right screenshot shows the search results page for ISSN 2531-2545, displaying '0 indexed journals' and a message: 'You searched for '2531-2545' and we found no results. Search terms must be in English. Try removing some of the filters you have set, modifying the text in the search box, or using less specific search terms.' A red arrow points from the DOAJ Seal icon in the summary to the DOAJ logo in the search results header.

Figura 29. Métricas de DOAJ para *journal* no halladas por la herramienta ante su ausencia en esta base de datos

4.1.3 Resultados de REDIB

Como muestra la Figura 30, la vista resumida de REDIB nos trae el widget exportable desde la web. En caso de que este widget no existiera, de todas formas nos informa la existencia de un recurso en REDIB, como se observó anteriormente en la Figura 29. La vista maximizada muestra ambas métricas, de la forma en que se visualizó en la sección 4.1.2. La Figura 31 comprueba que cuando el recurso no muestra métricas de REDIB es porque realmente tampoco está indexado en el sitio original.

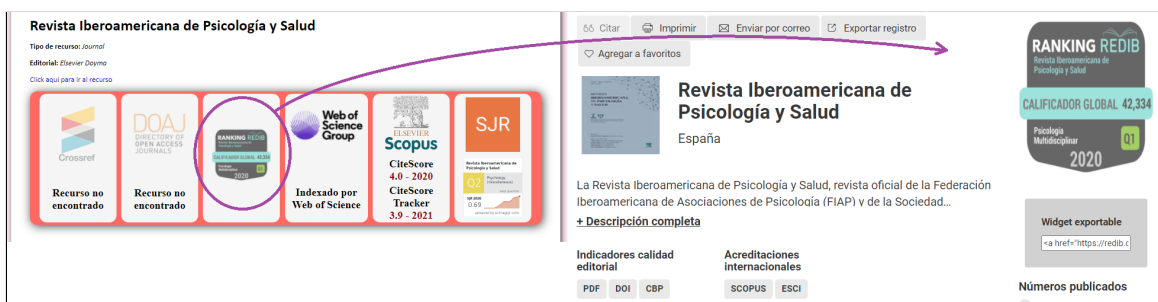


Figura 30. Métricas de REDIB para *journal* en la herramienta y en el sitio real

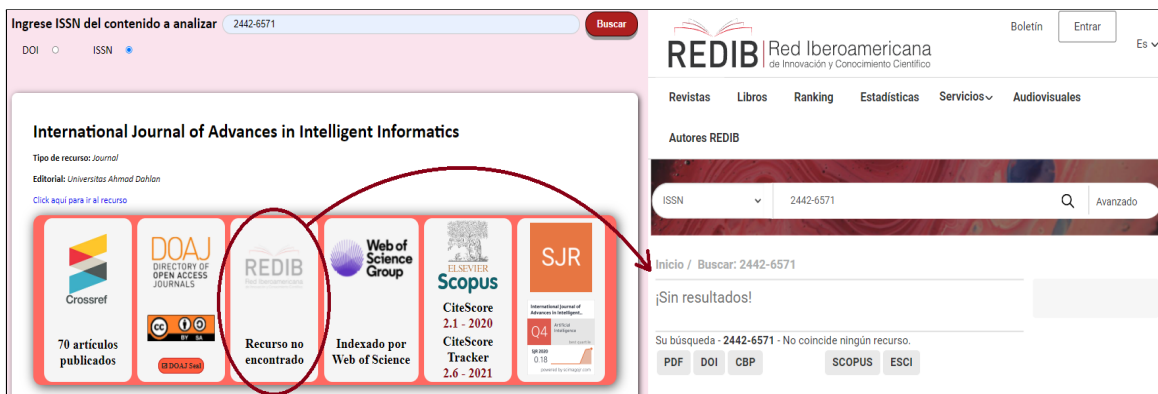


Figura 31. Métricas de REDIB para *journal* no halladas por la herramienta ante su ausencia en esta base de datos

4.1.4 Resultados de Web of Science

Cuando en la herramienta se encuentran métricas de Web of Science, lo verificamos buscando el recurso en la web de Master Journal List de Clarivate. La vista que mostramos en la Figura 32 es simplemente el resultado de búsqueda por ISSN del mismo *journal*. En la Figura 33 repetimos el ejercicio, ahora con un *journal* que la herramienta dice no poseer métricas de Web of Science, y verificado cuando no aparecen resultados en la web de la que se extrae la información.

Durante el desarrollo y pruebas de la aplicación han existido casos donde la API devolvía información, pero al ir al sitio de Master Journal List no se encontraba el recurso. Suponemos que se debía a una falta de sincronización entre lo que estaba realmente cargado en los servidores de Clarivate en forma de metadatos y lo que poseían en formato de texto para mostrar en la web.

The screenshot displays the Master Journal List interface. At the top, there is a search bar with the ISSN '1666-6038' and a 'Buscar' button. Below the search bar, the journal 'Journal of Computer Science and Technology' is listed with its ISSN and a 'Find a Match' button. A purple circle highlights the 'Web of Science' metric in the journal's profile, and a purple arrow points from this circle to the 'Web of Science' filter in the search results section. The search results section shows 'Found 1 results (Page 1)' and 'Exact Match Found' for the 'JOURNAL OF COMPUTER SCIENCE & TECHNOLOGY'.

Figura 32. Métricas de REDIB para *journal* en la herramienta y en el sitio real

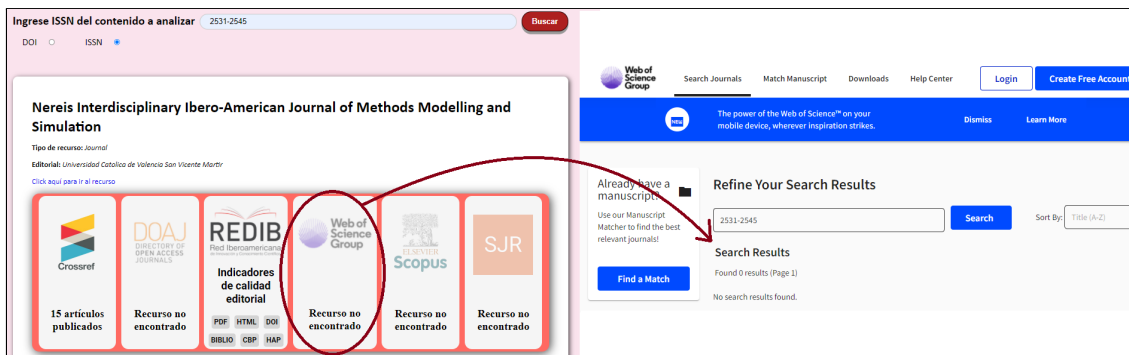


Figura 33. Métricas de WoS para *journal* no halladas por la herramienta ante su ausencia en esta base de datos

4.1.5 Resultados de Scopus y SJR

Como las métricas de SJR son elaboradas a partir de la información de Scopus, este ejemplo lo hemos realizado en conjunto. En la Figura 34 vemos que los resultados de Scopus, particularmente el CiteScore, aparecen en la página de Scopus en mayor detalle, pero verificamos que el valor es el mismo. También, en la parte inferior de la figura, vemos que el recurso se encuentra indexado en SJR, y mostramos cómo se ve el widget en la web (aunque esto es un recorte para simplificar el tamaño de la imagen).

Los recursos que no posean métricas de ninguno de los sitios muestran su logo transparente, como en todos los casos previos, y la búsqueda de los mismos en las respectivas páginas de internet otorga resultados nulos, como se observa en la Figura 35. Debido al vínculo entre estos dos sitios, lo más probable es que ambos tengan métricas de un *journal* o ambos no las posean.

Ingrese ISSN del contenido a analizar 1538-3598 **Buscar** Scopus Preview Author search Sources

DOI ISSN

JAMA

Tipo de recurso: Journal
Editorial: American Medical Association
[Click aquí para ir al recurso](#)

5026 artículos publicados

Recurso no encontrado

Recurso no encontrado

Indexado por Web of Science

CiteScore 24.8 - 2020
CiteScore Tracker 35.6 - 2021

JAMA - Journal of the American Medical Association
Q1
SJR 2020 4.69

JAMA - Journal of the American Medical Association
Scopus coverage years: from 1883 to Present
Publisher: American Medical Association
ISSN: 0098-7484 E-ISSN: 1538-3598
Subject area: Medicine: General Medicine
Source type: Journal

[View all documents >](#) [Set document alert](#) [Save to source list](#) [Source Homepage](#)

CiteScore CiteScore rank & trend Scopus content coverage

Improved CiteScore methodology
CiteScore 2020 counts the citations received in 2017-2020 to articles, reviews, conference papers, book chapters and data papers published in 2017-2020, and divides this by the number of publications published in 2017-2020. [Learn more >](#)

CiteScore	2020	CiteScoreTracker 2021
24.8	63.604 Citations 2017 - 2020 2561 Documents 2017 - 2020	35.6 = 70.094 Citations to date 1968 Documents to date

SJR Scimago Journal & Country Rank

JAMA - Journal of the American Medical Association

← Show this widget in your own website

Just copy the code below and paste within your html code:

``

powered by scimagojr.com

COUNTRY

United States

Universities and research institutions in United States

SUBJECT AREA AND CATEGORY

Medicine
↳ Medicine (miscellaneous)

PUBLISHER

American Medical Association

H-INDEX

680

Figura 34. Métricas de Scopus y SJR para *journal* en la herramienta y en el sitio real

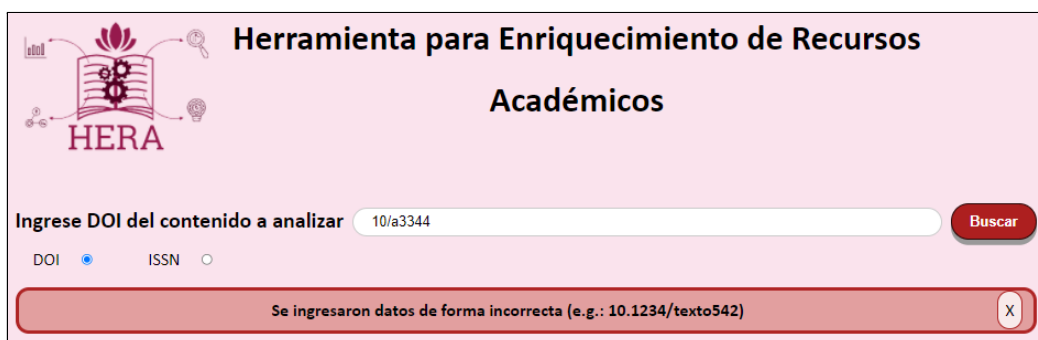
The screenshot displays a search interface for the journal "IADIS INTERNATIONAL JOURNAL ON COMPUTER SCIENCE AND INFORMATION SYSTEMS" (ISSN: 1646-3692). The interface includes a search bar with the ISSN, a "Buscar" button, and a "Filter refine list" section. The filter list shows "Scopus" and "SJR" both marked as "Recurso no encontrado". A message box states "No sources were found. Please check your search terms and filters or try a different combination of search criteria." Below the filter list, there is a search bar with the ISSN "1646 3692" and a search icon, with the message "Sorry, no results were found." below it.

Figura 35. Métricas de Scopus y SJR para *journal* no halladas por la herramienta ante su ausencia en estas bases de datos

4.2 Pruebas y resultados en búsquedas de *papers*

En esta sección mostraremos capturas en las cuales se verifica que la información exhibida en la herramienta es coherente con aquella que realmente existe en las bases de datos utilizadas. Para ello, realizamos búsquedas de *papers* en las cuales pudiéramos visualizar casos en los que se cuenta con información de un sitio, y casos en los que no. Las bases de datos involucradas en este caso son Crossref, DOAJ, Semantic Scholar, Microsoft Academic, Dimensions y Altmetric. Cabe mencionar que para encontrar un artículo que no se encuentre en Crossref hemos tenido que forzar el caso mediante un DOI falso.

En el caso de introducir mal el formato de un DOI para buscar, no se realizará una búsqueda y se indicará al usuario del error cometido para su corrección, como se indica en la Figura 36 a continuación.



The image shows a screenshot of the HERA tool interface. At the top, it says "Herramienta para Enriquecimiento de Recursos Académicos". Below that, there is a search input field with the text "Ingrese DOI del contenido a analizar" and the value "10/a3344". To the right of the input field is a "Buscar" button. Below the input field, there are radio buttons for "DOI" (selected) and "ISSN". At the bottom of the interface, there is a red error message box that says "Se ingresaron datos de forma incorrecta (e.g.: 10.1234/texto542)".

Figura 36. Error de formato de DOI

4.2.1 Resultados de Crossref

Como se muestra en la Figura 37, si buscamos un recurso en la herramienta y vemos que posee información provista por Crossref, luego podemos encontrarlo realizando la misma búsqueda en la página de Crossref. En esta última no se visualizan las métricas, ya que Crossref no es un sitio

donde el recurso se encuentre almacenado y sólo provee metadatos que son buscados a través de su API. En la Figura 38 podemos observar el resultado de una búsqueda que no posee resultados en Crossref, y ya que el DOI utilizado fue inventado, tampoco se exhibirán métricas de ningún otro.

Figura 37. Métricas de Crossref en la herramienta y en el sitio real

Figura 38. Métricas de Crossref no halladas por la herramienta ante su ausencia en esta base de datos

4.2.2 Resultados de DOAJ

En la Figura 39 podemos apreciar las métricas de DOAJ con resultados positivos. El recurso se encuentra en DOAJ y nos muestra que posee el DOAJ Seal. Este sello es otorgado, en rigor de verdad, a *journals*. Por ello es que el hecho de visualizarlo indica que el *journal* donde está publicado este *paper* es quien lo posee, lo cual se comprueba al ver las métricas del *journal* debajo de las del *paper*. A la derecha podemos comprobar que ambos recursos se encuentran en el sitio web de DOAJ y la existencia del sello. La Figura 40 muestra un ejemplo de un artículo que no se encuentra indexado por DOAJ, lo cual evidenciamos a un costado, donde vemos que el recurso no aparece en la lista de resultados del sitio.

The figure consists of two panels. The top panel displays metrics for a specific article: 'Analysis of Water Pollution Using Different Physicochemical Parameters: A Study of Yamuna River'. It shows 7 citations, a DOAJ Seal, a Scopus logo with 7 citations, 'Sin información de tendencias', and 'Se habla del tema'. The article is published in 'Frontiers in Environmental Science'. The bottom panel displays metrics for the journal 'Frontiers in Environmental Science', showing 957 articles published, a DOAJ Seal, a Scopus logo with a CiteScore of 4.4 (2020) and CiteScore Tracker of 4.7 (2021), and an SJR logo.

Figura 39. Métricas de DOAJ en la herramienta y en el sitio real

The figure consists of two panels. The top panel displays metrics for an article: 'Solar superstorms: planning for an internet apocalypse'. It shows 0 citations, a DOAJ Seal with the text 'Recurso no encontrado', a Scopus logo with 'No hay información de citas', 'Sin información de tendencias', and 'Hot topic en redes sociales!'. The article is published in 'Proceedings article'. The bottom panel shows a search result for '0 indexed articles' with the text 'You searched for '10.1145/3452296.3472916' and we found no results. Search terms must be in English.'.

Figura 40. Métricas de DOAJ no halladas por la herramienta ante su ausencia en esta base de datos

4.2.3 Resultados de Microsoft Academic

En la prueba de búsquedas de métricas de Microsoft encontramos información de citas, predicciones de citas y etiquetas con las que se relaciona al recurso. En la vista resumida de la herramienta, visualizada en la Figura 41, sólo mostramos el dato de predicción de citas por simplicidad. Podemos ver en la misma imagen que el recurso se encuentra en Microsoft Academic, aunque esta información de predicciones se visualiza de otra forma. El número “18” que se observa, es en realidad perteneciente a este cálculo que es estimado por el algoritmo de Microsoft. La Figura 42 muestra un ejemplo del caso en que no se proporcionan métricas de un *paper* por parte de Microsoft Academic. Del lado derecho de la imagen se observa la búsqueda del mismo recurso, que sólo podemos hacer mediante título al utilizar el sitio web (a diferencia de la API que busca por DOI) y observamos que no se encuentra indexado.

The image shows a side-by-side comparison of a search result for a scientific paper. On the left is a search result card from a tool, and on the right is the actual Microsoft Academic search result page. The card on the left includes a 'Predicciones de citas' (Citation Predictions) section with a red circle around the number '18'. The Microsoft Academic page on the right shows the same paper title and abstract, with 25 references and 18 citations listed.

Figura 41. Métricas de Microsoft Academic en la herramienta y en el sitio real

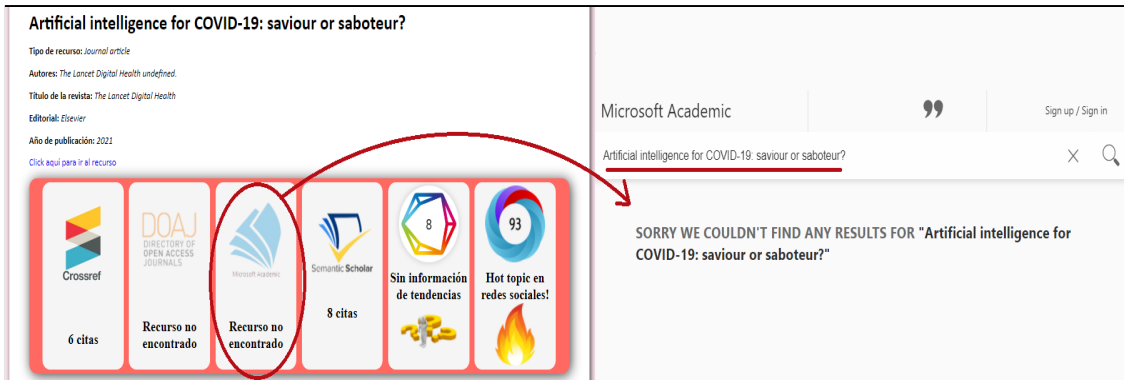


Figura 42. Métricas de Microsoft no halladas por la herramienta ante su ausencia en esta base de datos

4.2.4 Resultados de Semantic Scholar

Cuando recopilamos métricas de Semantic Scholar, veremos en la vista resumida la información con respecto a la cantidad de citas que el *paper* posee, y, de ser posible, la cantidad de citas que fueron “influyentes”, como se observa en la Figura 43. La misma información puede hallarse buscando el recurso en la página de Semantic Scholar, visualizada en forma de recuadro a la derecha. La Figura 44 muestra el ejemplo cuando un recurso no posee métricas en Semantic Scholar, que debe buscarse en su sitio de la misma forma que en Microsoft; por el título del mismo.



Figura 43. Métricas de Semantic Scholar en la herramienta y en el sitio real

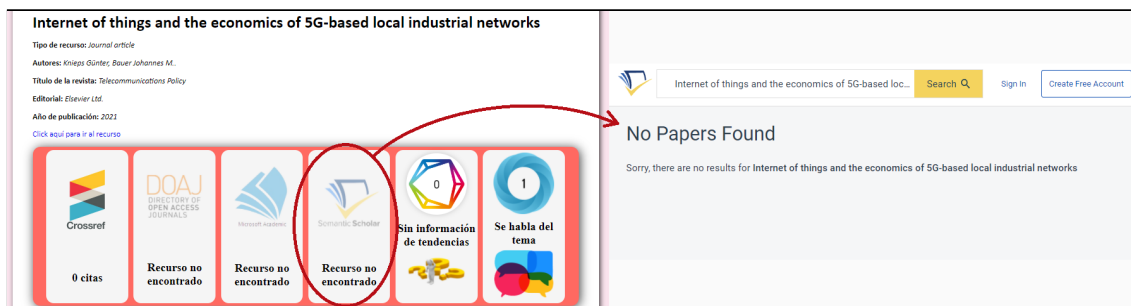


Figura 44. Métricas de Semantic Scholar no halladas por la herramienta ante su ausencia en esta base de datos

4.2.5 Resultados de Dimensions

Como vemos en la Figura 45, los resultados de las métricas provistas por la API de Dimensions vienen en la forma de un *badge* con un valor numérico de citas, y que al posicionar el mouse encima añade información.

Si se cliquea el *badge*, el navegador nos dirige a un sitio que detalla esta misma información, y que luego provee otro link al recurso almacenado en la base de datos. Cuando la API no nos devuelve información, aparecerá un signo de pregunta (“?”) en lugar de un número de citas en el *badge*, como se observa en la Figura 46. Además, al hacer clic nos lleva a una página de Dimensions que indica que el recurso no se encuentra.

Cloud Computing, Big Data and the Industry 4.0 Reference Architectures

Tipo de recurso: *Journal article*
 Autores: Velásquez Nancy, Estevez Elsa, Pesado Patricia.
 Título de la revista: *Journal of Computer Science and Technology*
 Editorial: Postgraduate Office, School of Computer Science, Universidad Nacional de La Plata
 Año de publicación: 2018

ABSTRACT: The Industry 4.0 promotes the use of Information and Communication Technologies (ICT) in manufacturing processes to obtain customized products satisfying demanding needs of new consumers. The Industry 4.0 approach transforms the traditional pyramid model of automation to a network model of interconnected services, combining operational technology (OT) with Information Technology (IT). This new model allows the creation of ecosystems enabling more flexible production processes through connecting systems and sharing data. In this context, cloud computing and big data are critical technologies for leveraging the approach. Thus, this paper analyzes cloud computing and big data under the lenses of two leading reference architectures for implementing Industry 4.0: 1) the Industrial Internet Reference Architecture (IIIRA), and 2) the Reference Architecture Model Industrie 4.0 (RAMI 4.0). A main contribution of this paper is to present a comparative analysis of IIIRA and RAMI 4.0, discussing needs, benefits, and challenges of applying cloud computing and big data in the Industry 4.0.

Dimensions Log in Register

Publication - Article
Cloud Computing, Big Data and the Industry 4.0 Reference Architectures
Journal of Computer Science and Technology, 18(03), e29-e29 - December 2018
<https://doi.org/10.24215/16666038.18.e29>

Authors
 Nancy Velásquez - National University of La Plata
 Elsa Estevez - Universidad Nacional del Sur
 Patricia Pesado - National University of La Plata

Abstract
 The Industry 4.0 promotes the use of Information and Communication Technologies (ICT) in manufacturing processes to obtain customized products satisfying demanding needs of new consumers. The Industry 4.0 approach transforms the traditional pyramid model of automation to a network model of interconnected services, combining operational technology (OT) with Information Technology (IT). This new model allows the creation of ecosystems enabling more flexible production processes through connecting systems and sharing data. In this context, cloud computing and big data are critical technologies for leveraging the

More

Summary Citations Citing research categories

This is the public page for a publication record in Dimensions, a free research insights platform that brings together information about funding, scholarly outputs, policy, patents and grants.

This publication in **Journal of Computer Science and Technology** has been cited **14 times**. 92% of its citations have been received in the past two years, which is **higher than you might expect**, suggesting that it is currently receiving a lot of interest.

Compared to other publications in the same field, **this publication is extremely highly cited** and has received approximately **7.28 times more citations** than average.

View more details in Dimensions

14 Total citations
 13 Recent citations
 7.28 Field Citation Ratio
 n/a Relative Citation Ratio

14 Citations
 13 Recent citations
 7.28 Field Citation Ratio
 n/a Relative Citation Ratio

Figura 45. Métricas de Dimensions en la herramienta y en el sitio real

6G Cognitive Information Theory: A Mailbox Perspective

Tipo de recurso: *Journal article*
 Autores: Hao Yixue, Miao Yiming, Chen Min, Gharavi Hamid, Leung Victor C. M.
 Título de la revista: *Big Data and Cognitive Computing*
 Editorial: MDPH AG
 Año de publicación: 2021

ABSTRACT: With the rapid development of 5G communications, enhanced mobile broadband, massive machine type communications and ultra-reliable low latency communications are widely supported. However, a 5G communication system is still based on Shannon's information theory, while the meaning and value of information itself are not taken into account in the process of transmission. Therefore, it is difficult to meet the requirements of intelligence, customization, and value transmission of 6G networks. In order to solve the above challenges, we propose a 6G mailbox theory, namely a cognitive information barrier to enable distributed algorithm embedding for intelligence networking. Based on Mailbox, a 6G network will form an intelligent agent with self-organization, self-learning, self-adaptation, and continuous evolution capabilities. With the intelligent agent, redundant transmission of data can be reduced while the value transmission of information can be improved. Then, the features of mailbox principle are introduced, including polarity, traceability, dynamics, convergence, figurability, and dependence. Furthermore, key technologies with which value transmission of information can be realized are introduced, including knowledge graph, distributed learning, and blockchain. Finally, we establish a cognitive communication system assisted by deep learning. The experimental results show that, compared with a traditional communication system, our communication system performs less data transmission quantity and error.

Click aquí para ir al recurso

Crossref 0 citas
 DOAJ DIRECTORY OF OPEN ACCESS JOURNALS Recurso no encontrado
 Microsoft Academic Recurso no encontrado
 Semantic Scholar Recurso no encontrado
 Recurso no encontrado
 Recurso no encontrado

Dimensions Embed badge What is this page?

404 NOT FOUND

Sorry, we can't find a record for that publication!
 It could be that it doesn't exist, or just that there's something odd happening with how we're calling the Dimensions API.
 Try visiting Dimensions and searching for it there.

Figura 46. Métricas de Dimensions no halladas por la herramienta ante su ausencia en esta base de datos

4.2.6 Resultados de Altmetric

La información provista por la API de Altmetric se visualiza de forma similar a la de Dimensions, por medio de un *badge*, como podemos ver en la Figura 47. Al posicionar el mouse encima, vemos una breve descripción de los sitios donde el recurso es mencionado, y podemos presionar el botón para ver más detalles y dirigirnos a una página web. En ésta, podemos analizar información demográfica y conocer los comentarios originales realizados al *paper*. Cuando la búsqueda en Altmetric no es exitosa, mostrará un signo de pregunta (“?”) al igual que Dimensions, y clicar el *badge* nos lleva a un sitio que indica la falta de datos del recurso en Altmetric, visualizado en la Figura 48. Es necesario mencionar que existen casos en que el *badge* de altmetric es renderizado con el signo de pregunta y el mensaje de “Recurso no encontrado”, como si no tuviese métricas disponibles, pero aún así al hacer clic nos lleva a una página con detalles de información. Esto se debe a que el *badge* se nutre de la misma información que nuestra herramienta, que es la API de Altmetric. Esta API puede devolver un código de error 404 (no encontrado), pero aún así al clicar en el *badge* encontrar información, ya que Altmetric mismo puede consultar por metadatos de los recursos interactuando con otros servicios. Nuestra herramienta no tiene mecanismo por el cual saber que habrá datos al clicar el *badge* ya que la API no devuelve señales de este evento, sólo un simple “404 Not Found”.

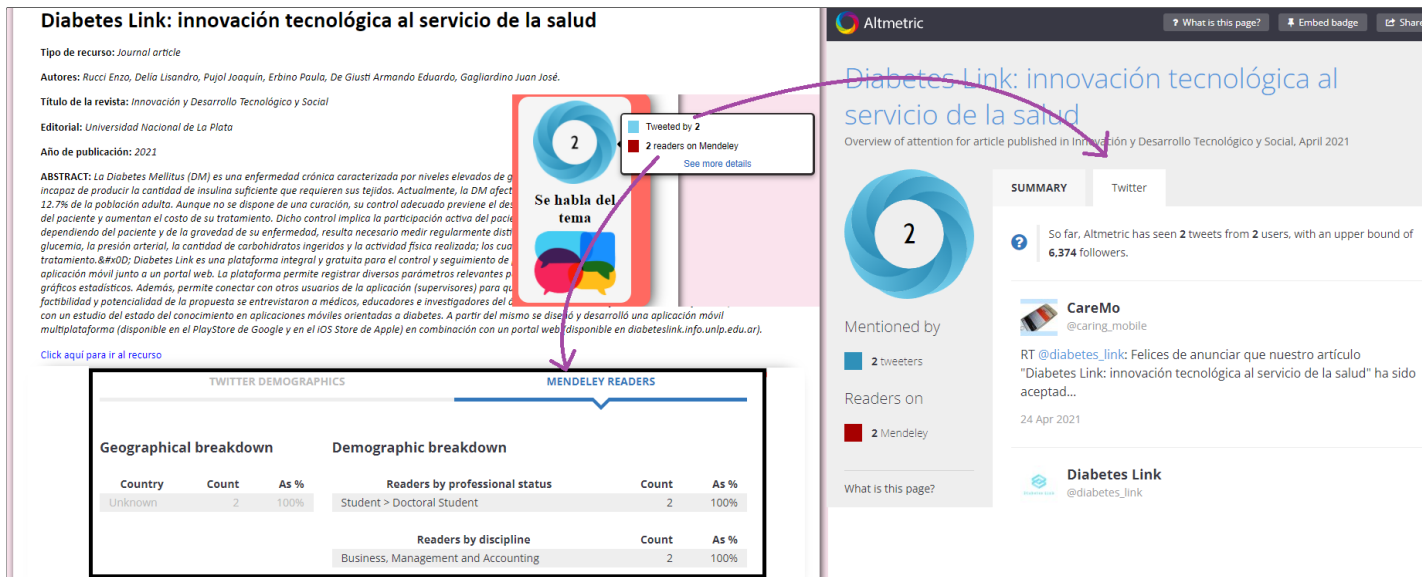


Figura 47. Métricas de Altmetric en la herramienta y en el sitio real

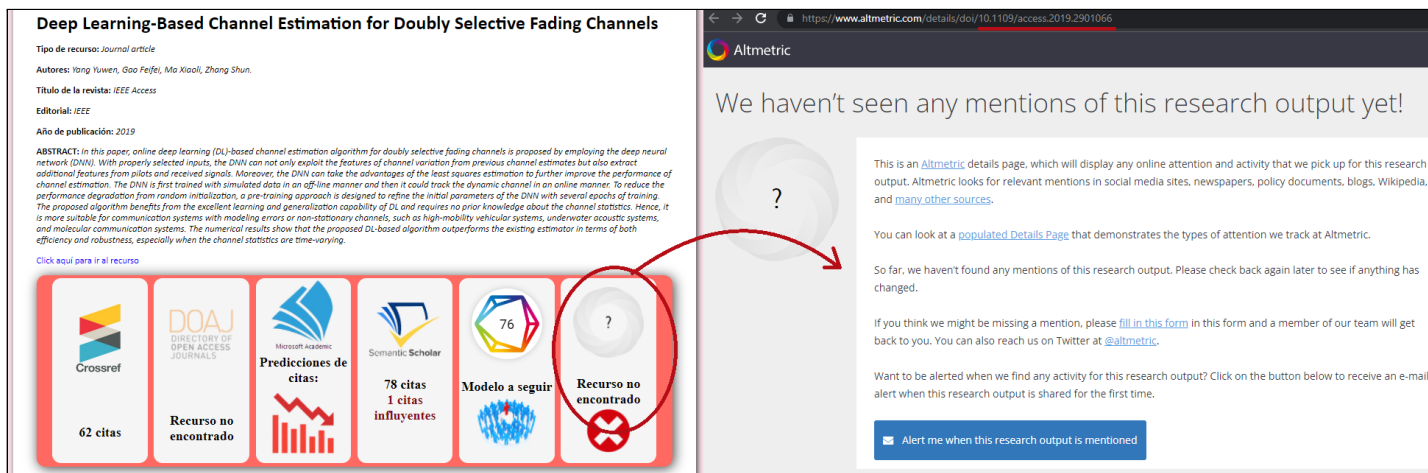


Figura 48. Métricas de Altmetric no halladas por la herramienta ante su ausencia en esta base de datos

Capítulo 5 - Conclusiones y trabajos futuros

En este último capítulo, en primer lugar, resumimos el trabajo realizado y presentamos las conclusiones de esta tesina. En segundo lugar, detallamos aquellos trabajos que podrían realizarse a futuro a modo de expandir la solución propuesta o mejorarla.

5.1 Conclusiones

En el ámbito académico-científico, múltiples investigadores de diversos campos de estudio se lanzan en la misión de la búsqueda del nuevo conocimiento y saber. Tanto para apoyar, fundamentar y complementar su trabajo, como para documentar sus hallazgos finales, producen lo que denominamos “artículo científico”. En el proceso de producción científica incorporan el trabajo producido previamente por sus colegas, lo que los enfrenta al desafío de seleccionar aquellos recursos que le resulten más relevantes y de mayor calidad. Este desafío se torna cada vez más dificultoso debido al crecimiento exponencial en la cantidad de recursos para analizar gracias al desarrollo tecnológico y a la existencia de múltiples indicadores y métricas de diversa índole para cada uno de ellos.

Del estudio realizado sobre las diversas métricas y métodos de evaluación de calidad e impacto de *papers* y *journals* y de las tendencias en sus usos, identificamos que existe una necesidad de proporcionar asistencia a los investigadores en su tarea de recopilar métricas de contenido para analizarlos y/o evaluarlos de una forma más elaborada. En este contexto, propusimos el desarrollo de una aplicación que intenta dar respuesta a esta necesidad, permitiendo de forma ágil buscar y recopilar datos de numerosas y diversas bases de datos académicas, y visualizarlos de manera integral en un sólo sitio web.

Previo al inicio del desarrollo de la herramienta, debimos realizar un análisis de diversas bases de datos para estudiar el contenido disponible, métodos de evaluación utilizados por las mismas, métricas exhibidas, y medios de acceso a sus datos. Realizamos ésto para incluir aquellas que proveían de la información más relevante y abarcativa, así como también para analizar cuáles medios de acceso eran viables para nuestro desarrollo. Los medios finales por los que decidimos obtener información son las APIs REST provistas por algunas bases, y *web scraping* en los casos donde la primera no fuera posible.

La herramienta consta de dos aplicaciones: una para el frontend realizada con ReactJS y otra para el backend desarrollada con Node.js. Esta última consiste de una aplicación liviana, cuyo único propósito es el de interactuar con la de frontend y, en base a sus peticiones, disparar búsquedas a servidores externos para luego devolver los datos. En la aplicación frontend utilizamos un modelo basado en componentes para que cada uno contenga las métricas de cada base de datos, lo que permite incorporar información de nuevas bases de datos para visualizar fácilmente, sin alterar significativamente el código presente en el resto de la aplicación. Cada componente de cada base de datos se renderiza en el navegador de forma única debido a la diversidad de datos que cada una puede poseer.

Las búsquedas realizadas en la herramienta se hacen a nivel de *paper*, identificado por un DOI, y de *journal*, identificado por un ISSN. En la búsqueda de *papers* se decidió incluir también una segunda búsqueda automática en segundo plano, que incorpora a las métricas del *paper* las del *journal* en el cual está publicado, para así dar mayor contexto al análisis que pudiese realizar un investigador.

Las métricas se consultan de múltiples bases de datos con diversos objetivos y características, lo que contribuye a una evaluación más amplia de los recursos:

- Scopus y WoS son dos de las bases de datos de *journals* más reconocidas en muchos ámbitos científicos, cuyos dueños y gestores son grandes empresas.
- SCImago elabora métricas e indicadores de calidad de *journals* basándose en la información recopilada por Scopus.
- REDIB aparece como una alternativa de la región iberoamericana que concentra *journals* producidos en esta parte del mundo.
- DOAJ es una de las bases de datos de acceso abierto más ejemplares a nivel mundial. Es financiada por empresas que comparten su visión de acceso abierto al contenido, incluyendo contenidos de todas las áreas, países e idiomas.
- Crossref se enfoca en la estandarización de identificadores de recursos para poder referenciarlos y encontrarlos fácilmente.
- Microsoft Academic surge como una alternativa a Google Scholar, haciendo hincapié en medios de inteligencia artificial con análisis semántico para descubrir contenido académico que se encuentra en la web y así poder indexarlo.
- Semantic Scholar va un paso más allá en la búsqueda de recursos, queriendo ayudar a acelerar el proceso de selección de información mediante algoritmos que conectan los recursos de mayor relevancia entre ellos, siendo capaz también de generar resúmenes que resaltan aspectos claves de los mismos.
- Finalmente, sitios como Dimensions y Altmetric proponen la integración de métricas de manera compacta y de fácil acceso. Utilizando alta tecnología, recopilan información de toda la web para elaborar demográficas, menciones en redes sociales y impacto relativo a recursos similares.

Las consultas de datos a estas bases se realizan en tiempo real, con el propósito de contar con la información en su estado más actualizado y de prescindir de una base de datos para almacenarlas en caso contrario. Como

los tiempos de espera para obtener una respuesta de las bases de datos no son extensos, la carga adicional introducida resulta despreciable.

A partir de los resultados obtenidos, consideramos que el objetivo planteado inicialmente se ha cumplido. HERA representa una herramienta que de manera ágil permite integrar y visualizar métricas de bases de datos relevantes y reconocidas en un único sitio. Esto es de vital importancia a la hora de evaluar contenido, no sólo por eximir a los investigadores de la tediosa tarea de buscar métricas en muchas páginas web, sino también por facilitar la visualización de métricas para un análisis integral de estas. Considerando las características de HERA, esperamos que los miembros de la comunidad académico-científica la encuentren útil para evaluar la calidad y el impacto de los recursos académicos y que contribuya a facilitar y acelerar dicha tarea.

5.2 Trabajos futuros

En esta sección mencionaremos ciertos aspectos de la herramienta que pueden mejorarse o extenderse en el futuro:

5.2.1 Expandir el banco de bases de datos académicas

En este momento, la herramienta recopila métricas e indicadores de bases de datos académicas cuidadosamente seleccionadas. Sin embargo, en el análisis también forzó una limitación a la cantidad de fuentes para no sobrecargar el nivel de complejidad inicial del desarrollo. De esta forma no se descarta que, siempre que se realizara un análisis previo adecuado, se incorporaran nuevos índices y fuentes de datos en el futuro para ser añadidos en la herramienta. Opciones atractivas incluyen la incorporación de índices regionales, o fuentes de datos adicionales como OJS, y también bases reconocidas pero de disciplinas más específicas como el caso de PubMed. Dada esta posibilidad de contar con numerosas bases, sería conveniente en

ese caso agregar la funcionalidad de poder seleccionar de qué bases de datos interesa obtener información.

5.2.2 Implementación de una API REST

La herramienta podría exponer parte de su funcionalidad mediante el desarrollo de una API REST para así desarrollar nuevas herramientas que se nutran de los datos brindados por HERA. Por ejemplo, de tener una API se podrían implementar soluciones como:

- Extensiones web para navegadores que muestren métricas de un recurso en tiempo real. De esta forma no sería necesario tener que ingresar identificadores a mano en la herramienta, haciéndola aún más automática.
- Generación de archivos de datos masivos que recopilen métricas de varios recursos académicos a la vez. Esta funcionalidad puede resultar ideal para evaluar recursos dentro de un contexto específico, como ser un proyecto particular, un área temática o una institución.
- Integración con OJS para visualizar métricas desde el mismo y obtener datos a través de terceros de forma más transparente.

5.2.3 Optimización de tiempo de búsquedas de métricas

En el caso de realizar la misma búsqueda varias veces, la herramienta demora tanto como los sitios externos demoren en devolver esa información, con un límite para el tiempo de espera preestablecido. Sería de utilidad poseer mecanismos de caché para optimizar los tiempos en caso de búsquedas recurrentes. Además, algunas de las APIs de sitios externos utilizadas tienen cuotas máximas específicas de uso que se podría evitar alcanzar al tener la información ya almacenada.

Referencias

- Becker Bernard. [Tools for Authors: What is the h index?](https://beckerguides.wustl.edu/authors/hindex#_edn1). *Becker Medical Library*. Recuperado de https://beckerguides.wustl.edu/authors/hindex#_edn1 (último acceso 2/12/2021)
- Clarivate - [History of citation indexing - Web of Science Group](https://clarivate.com/webofsciencegroup/essays/history-of-citation-indexing/). Recuperado de <https://clarivate.com/webofsciencegroup/essays/history-of-citation-indexing/> (último acceso 2/12/2021)
- Clarivate. [Web of Science Journal Evaluation Process and Selection Criteria - Web of Science Group](https://clarivate.com/webofsciencegroup/journal-evaluation-process-and-selection-criteria/) . Recuperado de <https://clarivate.com/webofsciencegroup/journal-evaluation-process-and-selection-criteria/> (último acceso 2/12/2021)
- Crossref (2018). [Membership terms](https://www.crossref.org/membership/terms/). Recuperado de <https://www.crossref.org/membership/terms/> (último acceso 2/12/2021)
- Delgado López-Cózar, Emilio; Robinson-García, Nicolás; Torres Salinas, Daniel (2012). Manipulating Google Scholar Citations and Google Scholar Metrics: simple, easy and tempting. *EC3 Working Papers* 6: 29 May, 2012
- Design Education Forum of Southern Africa - [What is an Academic Paper?](https://www.defsa.org.za/what-academic-paper/) . Recuperado de https://www.defsa.org.za/what-academic-paper (último acceso 2/12/2021)
- Directory of Open Access Journals. [Guide to applying – DOAJ](https://doaj.org/apply/guide/) . Recuperado de <https://doaj.org/apply/guide/> (último acceso 2/12/2021)
- Egghe, L. Theory and practise of the *g*-index. *Scientometrics* 69, 131–152 (2006). <https://doi.org/10.1007/s11192-006-0144-7>
- Enago Academy . Recuperado de <https://www.enago.com/academy/open-access-publications-get-cited-more-often/> (último acceso 2/12/2021)
- Elsevier (2020). [Scopus Content Coverage Guide](https://www.elsevier.com/?a=69451) . Recuperado de <https://www.elsevier.com/?a=69451> (Último acceso 2/12/2021)
- Elsevier - [What is peer review? - the peer review process](https://www.elsevier.com/reviewers/what-is-peer-review) . Recuperado de <https://www.elsevier.com/reviewers/what-is-peer-review> (Último acceso 2/12/2021)

Eunjung Cha, Ariana (2015). [Paul Allen's AI research group unveils program that aims to shake up how we search scientific knowledge. Give it a try.](https://www.washingtonpost.com/news/to-your-health/wp/2015/11/02/paul-allen-s-ai-research-group-unveils-program-that-aims-to-shake-up-how-we-search-scientific-knowledge-give-it-a-try/) *The Washington Post*. Recuperado de <https://www.washingtonpost.com/news/to-your-health/wp/2015/11/02/paul-allen-s-ai-research-group-unveils-program-that-aims-to-shake-up-how-we-search-scientific-knowledge-give-it-a-try/> (Último acceso 2/12/2021)

Freyne, J., Coyle, L., Smyth, B., & Cunningham, P. (2010). Relative status of journal and conference publications in computer science. *Communications of the ACM*, 53(11), 124–132.

Gargouri Y, Hajjem C, Larivie`re V, Gingras Y, Carr L, et al. (2010) Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *PLoS ONE* 5(10): e13636.
<https://doi.org/110.1371/journal.pone.0013636>

Godoy, D., Zunino, A. & Mateos, C. Publication practices in the Argentinian Computer Science community: a bibliometric perspective. *Scientometrics* **102**, 1795–1814 (2015). <https://doi.org/10.1007/s11192-014-1450-0>

Gonzalez-Pardo, R.; Repiso, R.; Arroyave-Cabrera, J. (2020). Revistas iberoamericanas de comunicación a través de las bases de datos Latindex, Dialnet, DOAJ, Scopus, AHCI, SSCI, REDIB, MIAR, ESCI y Google Scholar Metrics. *Revista Española de Documentación Científica*, 43 (4): e276.
<https://doi.org/10.3989/redc.2020.4.1732>

Dimensions, [Why did we build Dimensions](https://www.dimensions.ai/why-dimensions/) . Recuperado de <https://www.dimensions.ai/why-dimensions/> (Último acceso 2/12/2021)

Google Scholar, [Inclusion Guidelines for Webmasters](https://scholar.google.com/scholar/inclusion.html#indexing). Recuperado de <https://scholar.google.com/scholar/inclusion.html#indexing> (Último acceso 2/12/2021)

Hirsch JE. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* (2005, November 15) 102(46): 16569–16572.
<https://doi.org/10.1073/pnas.0507655102>

Howard D. White, Sebastian K. Boell, Hairong Yu, Mari Davis, Concepción S. Wilson, Fletcher T.H. Cole (2009) - [Lbcitations: A measure for comparative assessment of book publications in the humanities and social sciences](#), *Journal*

of the American Society for Information Science and Technology, 60(6), 1083-1096. <https://doi.org/10.1002/asi.21045>

Johnson Rob, Watkinson Anthony, Mabe Michael. [The STM Report 1968-2018](#) 5 ed., 2018, *International Association of Scientific, Technical and Medical Publishers Prins Willem Alexanderhof 5, The Hague, 2595BE, The Netherlands*

Lam Díaz, Rosa María. (2016). La redacción de un artículo científico. *Revista Cubana de Hematología, Inmunología y Hemoterapia*, 32(1), 57-69.

Latindex - *Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal*. [Metodología del Catálogo 2.0 de Latindex](#) . Recuperado de <https://www.latindex.org/latindex/meto2> (Último acceso 2/12/2021)

Lee, C.J., Sugimoto, C.R., Zhang, G. and Cronin, B. (2013), Bias in peer review. *J Am Soc Inf Sci Tec*, 64: 2-17. <https://doi.org/10.1002/asi.22784>

Lindsey, D. (1989). Using citation counts as a measure of quality in science measuring what's measurable rather than what's valid. *Scientometrics*, 15(3-4), 189-203.

Martín-Martín Alberto, Orduna-Malea Enrique, Thelwall Mike, Delgado López-Cózar Emilio. *Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories*. <https://doi.org/10.1016/j.joi.2018.09.002>

Martijn Visser, Nees Jan van Eck, Ludo Waltman; Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies* 2021; 2 (1): 20-41. https://doi.org/10.1162/qss_a_00112

Meyer, B., Choppy, C., Staunstrup, J., & van Leeuwen, J. (2009). Viewpoint: Research evaluation for computer science. *Communications of the ACM*, 52(4), 31-34.

Moed, H. F. (2020). Appropriate Use of Metrics in Research Assessment of Autonomous Academic Institutions. *Scholarly Assessment Reports*, 2(1): 1. DOI: <https://doi.org/10.29024/sar.8>

Mongeon, P., Paul-Hus, A. The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics* 106, 213-228 (2016). <https://doi.org/10.1007/s11192-015-1765-5>

Orduna-Malea, Enrique, Delgado López-Cózar, Emilio. *Taller: Research Analytics: la monitorización del impacto académico de un investigador* <http://sedici.unlp.edu.ar/handle/10915/63599>

Priem J., Taraborelli D., Groth P., Neylon C. (2010), *Altmetrics: A manifesto*, 26 October 2010. <https://altmetrics.org/manifesto>

REDIB - *Red Iberoamericana de Innovación y Conocimiento Científico*. [Criterios de calidad para revistas](https://redib.org/criterios-calidad). Recuperado de <https://redib.org/criterios-calidad> (Último acceso 2/12/2021)

Repiso, Rafael Cómo identificar una revista de calidad. *Cardiocre* [en línea]. 2015, 50(2), 46-48 [fecha de Consulta 6 de Agosto de 2021]. ISSN: 1889-898X. Disponible en: <https://www.redalyc.org/articulo.oa?id=277041630002>

Rhaim, M. Measurement and determinants of academic research efficiency: a systematic review of the evidence. *Scientometrics* 110, 581–615 (2017). <https://doi.org/10.1007/s11192-016-2173-1>

Rozemblum, C. (2020). El irrefrenable impulso por ser parte del MAINSTREAM. *Revista ECONO. Facultad De Ciencias Económicas. UNLP*, (21). Recuperado a partir de <https://revistas.unlp.edu.ar/econo/article/view/11030>

Rozemblum, C.; Unzurrunzaga, C.; Banzato, G. y Pucacco, C. 2015. Calidad editorial y calidad científica en los parámetros para inclusión de revistas científicas en bases de datos en Acceso Abierto y comerciales. *Palabra Clave (La Plata)*, vol. 4, n° 2, p. 64-80. Disponible en: <https://www.palabraclave.fahce.unlp.edu.ar/article/view/PCv4n2a01>

Schekman, R. (2013, December 9). How journals like Nature, Cell and Science are damaging science. *The Guardian*. Recuperado de <https://www.theguardian.com/commentisfree/2013/dec/09/how-journals-nature-science-cell-damage-science> (Último acceso 2/12/2021)

SCImago, (n.d.). SJR — SCImago Journal & Country Rank [Portal]. Recuperado de <http://www.scimagojr.com> (Último acceso 2/12/2021)

Thorsten Gruber (2014) Academic sell-out: how an obsession with metrics and rankings is damaging academia, *Journal of Marketing for Higher Education*, 24:2, 165-177, <https://doi.org/10.1080/08841241.2014.970248>

Tim C.E Engels, Andreja Istenič Starčič, Emanuel Kulczycki , Janne Pölonen ,
Gunnar Sivertsen (2018) - [Are book publications disappearing from scholarly
communication in the social sciences and humanities?](#), *Aslib Journal of
Information Management*