

HERA - Herramienta para Enriquecimiento de Recursos Académicos

JUAN FRANCISCO PORTO

Facultad de Informática, Universidad Nacional de La Plata (UNLP) (Argentina)

ENZO RUCCI

III-LIDI, Facultad de Informática, Universidad Nacional de La Plata (UNLP) (Argentina)

Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC) (Argentina)

erucci@lidi.info.unlp.edu.ar

GONZALO LUJÁN VILLARREAL

PREBI-SEDICI, Universidad Nacional de La Plata (UNLP) (Argentina)

CESGI-Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC) (Argentina)

gonzalo@prebi.unlp.edu.ar

RESUMEN

En la actualidad, determinar la calidad y el impacto de un recurso académico (revista o artículo científico) representa un verdadero desafío para un investigador. Al problema inicial de la existencia de múltiples indicadores y métricas de diversa índole para cada uno de ellos, se ha sumado el crecimiento exponencial en la cantidad de recursos para analizar, gracias al desarrollo tecnológico. En esta ponencia se presenta el diseño y desarrollo de una herramienta web que busca dar respuesta a esta problemática. HERA es una herramienta que apunta a simplificar, agilizar y apoyar el proceso de determinar la calidad y el impacto de un recurso académico. Para ello, HERA consulta múltiples fuentes en tiempo real para luego ofrecer información de un recurso como sus metadatos, su pertenencia a índices y bases de

datos, sus métricas de citas y menciones, e información de la publicación donde figura dicho recurso en caso de que corresponda.

PALABRAS CLAVE

Artículo científico, revista científica, *paper*, *journal*, bibliometría, bases de datos académicas, bases de datos bibliográficas, cienciometría, recuperación de información, evaluación bibliográfica.

Este trabajo presenta un resumen de: PORTO, JUAN FRANCISCO (2021). HERA: Herramienta para Enriquecimiento de Recursos Académicos. Tesis de grado, Facultad de Informática, Universidad Nacional de La Plata (UNLP). <http://hdl.handle.net/10915/129874>

La evaluación de la calidad e impacto en el presente

A la hora de utilizar la información publicada en revistas, congresos, libros, tesis, etc., los lectores deben contar con criterios que les permitan evaluar el grado de veracidad del contenido de cada artículo. La gran disponibilidad de contenido académico y de espacios donde se publica este contenido hace inviable conocer de antemano los procesos que se utilizan para controlar la calidad, asegurar la transparencia, y obtener métricas en cada uno de ellos. Así es como nace la necesidad de establecer criterios genéricos 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 una revista (o *journal*) como de un artículo (o *paper*) es el impacto de estos. Mientras que la calidad de una revista se asocia a los procesos que sigue y a los requisitos que impone para que un artículo sea finalmente publicado, el impacto se asocia a la repercusión que la revista o el artículo tiene en la comunidad científica, usualmente medido en número de citas y menciones (REPISO, 2015). La necesidad de determinar la calidad y el

impacto de las publicaciones se traduce en lo que hoy vemos como sistemas de evaluación, indicadores y métricas de revistas o artículos. 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 artículo o esta revista de buena calidad? ¿es este artículo o esta revista de alto impacto? ¿se puede predecir el éxito? ¿se puede entender si es de utilidad? Cabe destacar que los valores obtenidos al medir calidad e impacto poseen un carácter subjetivo muy alto por no ser un estándar y por existir diferentes formas de calcularlos y valorarlos, de acuerdo a las distintas áreas temáticas, organismos de medición y evaluación, métodos utilizados, fuentes de datos, y estrategias y tecnologías utilizadas.

En el caso de las revistas, 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 artículos habitualmente se asocia a la revista en que fueron publicados, por la necesidad de cumplir sus requisitos de publicación.

En adición a las métricas clásicas basadas en citas desde recursos científicos o académicos, han surgido soluciones alternativas, como, por ejemplo, las altmetrics. Dicho simple, este tipo de métricas engloban e incorporan nuevos aspectos para evaluar contenido, y están muy vinculadas al mundo hiperconectado actual: descargas, menciones en redes sociales, citas en blogs científicos e incorporaciones en sistemas de gestión bibliográfica son algunos de los valores que registran a la hora de medir el impacto de una publicación. Su punto fuerte yace en la facilidad de interpretación visual, sus valores numéricos simples y su conveniente portabilidad. Mientras que su

atractivo es indiscutible, no dejan de introducir nuevos valores a ser interpretados por los investigadores.

Más allá del tipo de métricas o del mecanismo utilizado para evaluar calidad e impacto de una publicación, es evidente que los investigadores poseen hoy en día mucha información a disposición, con más que variados métodos de análisis. Surgen muchas preguntas respecto de los tipos de métricas ya mencionadas, y las problemáticas en general son el natural carácter subjetivo de ellas, 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 difícil.

De este análisis de revistas y artículos, sus métricas y los problemas que conlleva analizarlos, 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. Existen 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 ejemplo, el sitio web [¿Dónde lo Publico?](#) recopila información de revistas de ciencias sociales y humanidades en español o portugués. Una opción superadora es la [Matriz de Información para el Análisis de Revistas](#) (MIAR), que da una vista ampliada e integral de métricas de las revistas que posee en su base de datos. Cabe aquí mencionar que MIAR no provee información a nivel de artículo, y Dónde lo Publico 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; RHAJEM, 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 las métricas mencionadas, sin que los investigadores deban recorrer decenas de sitios web buscando información sobre publicaciones o trabajos. Bajo esta idea surge la Herramienta para Enriquecimiento de Recursos Académicos, o simplemente HERA, como una solución centralizada que unifica información de revistas y de artículos, automatizando su recolección y brindando un punto de acceso único y simplificado a distintas métricas de calidad e impacto.

Cómo funciona HERA

HERA es una herramienta web que se encuentra disponible en <http://hera.sedici.unlp.edu.ar>. Su funcionamiento es relativamente sencillo: el usuario cuenta con una barra de búsqueda donde deberá introducir un identificador, que puede ser el ISSN si desea analizar una publicación, o un DOI si desea analizar un artículo (ver FIGURA 1). Luego, deberá hacer clic en el botón “Buscar” que indica a HERA que debe iniciar la búsqueda de información del recurso asociado a dicho identificador.

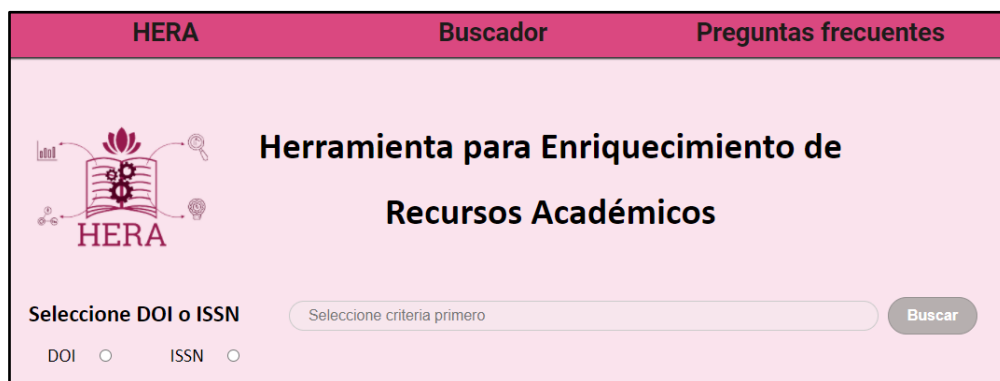


FIGURA 1. Interfaz de búsqueda de la herramienta con input de búsqueda y selectores de tipo de identificador (DOI e ISSN)

Una vez iniciada la búsqueda, HERA se conectará de manera concurrente con un conjunto de servicios en línea y sitios web (bases de datos de aquí en adelante), desde los cuales intentará obtener información del recurso de interés. Asimismo, HERA decidirá cuál es la mejor estrategia para obtener información a partir de cada base de datos (por ejemplo, una API REST o *web scrapping*), así como también qué tipo de información puede obtener a partir de cada una de ellos: pertenencia o no a la base de datos, citas, gráficos, altmetrics, etc. La siguiente tabla lista las bases de datos que consulta HERA, así como también las métricas que se obtienen en cada caso:

NOMBRE DE BASE DE DATOS	MÉTODOS DE SELECCIÓN DE CONTENIDO	(PRINCIPALES) MÉTRICAS E INDICADORES PROVISTOS	RECURSO¹ CONSIDERADO
SCOPUS	Proceso de evaluación por junta asesora	Número de citas Predicción de citas Factores de impacto	<i>Journal</i>
WEB OF SCIENCE	Proceso de evaluación por editores expertos	Índice de WoS al que pertenece Información bibliográfica	<i>Journal</i>
DOAJ	Proceso de evaluación de acceso abierto y calidad	Licencias Sello de DOAJ	<i>Paper & Journal</i>
LATINDEX	Proceso de evaluación de calidad (sólo iberoamérica o contenido de la región)	Información bibliográfica Presencia en el catálogo	<i>Journal</i>
RED IBEROAMERICANA	Proceso de evaluación de calidad (sólo Iberoamérica o contenido de la región)	Posición en ranking Información	<i>Journal</i>

¹ Si bien se usó el término *Journal*, también pueden estar considerados Conferencias y Libros

NOMBRE DE BASE DE DATOS	MÉTODOS DE SELECCIÓN DE CONTENIDO	(PRINCIPALES) MÉTRICAS E INDICADORES PROVISTOS	RECURSO ¹ CONSIDERADO
		bibliográfica Indicadores de calidad editorial y acreditaciones	
CROSSREF	Metadatos estandarizados por miembros de la plataforma	Número de citas Estadísticas de papers por año y autores con ORCID Información bibliográfica	<i>Paper & Journal</i>
SCIMAGO JOURNAL & COUNTRY RANK	Recursos contenidos en Scopus	Posición en ranking global Cuartil H-index del journal Información bibliográfica	<i>Journal</i>
GOOGLE SCHOLAR	Indexación por Google mediante algoritmos de <i>machine learning</i>	Número de citas H-index de autores	<i>Paper</i>
MICROSOFT ACADEMIC	Indexación por Bing mediante algoritmos de <i>machine learning</i>	Número de citas Predicción de citas Información bibliográfica	<i>Paper</i>
SEMANTIC SCHOLAR	Indexación mediante algoritmos de <i>machine learning</i> ejecutados sobre bases de datos afiliadas	Número de citas Frecuencia de citas Número de citas influyentes	<i>Paper</i>
ALTMETRIC	Realiza cálculos reconociendo menciones en redes sociales, foros, etc. mediante IA	Puntaje de atención basado en menciones con peso dependiendo el sitio	<i>Paper</i>

NOMBRE DE BASE DE DATOS	MÉTODOS DE SELECCIÓN DE CONTENIDO	(PRINCIPALES) MÉTRICAS E INDICADORES PROVISTOS	RECURSO ¹ CONSIDERADO
		Presencia de menciones en redes sociales	
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	<i>Paper</i>

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

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 2. 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 en un portal web.

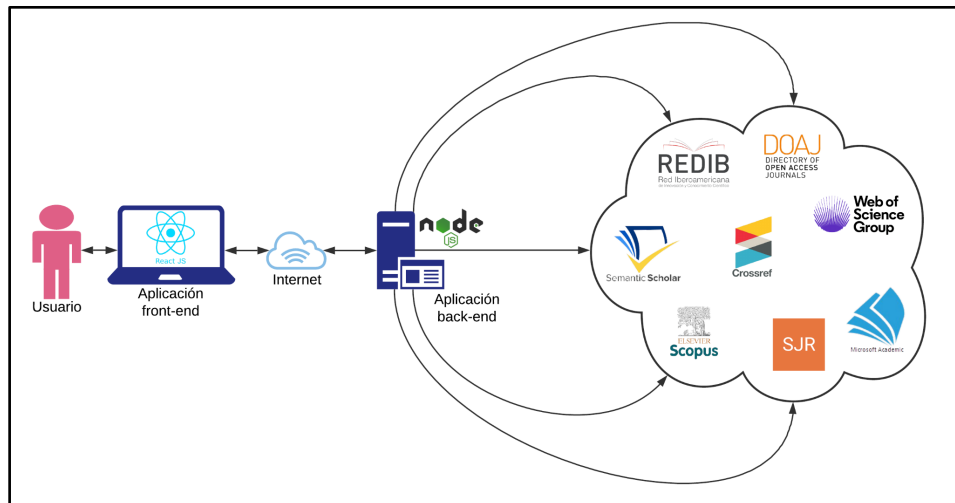


FIGURA 2. Esquema de comunicación entre las aplicaciones

Descripción de la aplicación back-end

Esta aplicación fue desarrollada con NodeJS y utilizando el framework `express`² 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.

² <https://expressjs.com/>

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>

TABLA 2. Endpoints de la API

Descripción de la 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 uno³. 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 cuales 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 Fetch⁴, similar al uso de XMLHttpRequest.

HERA en funcionamiento

Búsqueda de *journal*

Al introducir un ISSN en la barra de búsqueda, se realiza la búsqueda de un *journal* o publicación periódica. Se obtiene primero la información sobre la editorial y enlace al recurso, así como la vista del resumen de métricas recopiladas de los diferentes sitios, visualizado en la FIGURA 3. 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 visualizar la totalidad de su contenido, como mostramos en las FIGURAS 4, 5 y 6.

³ ReactJS: An Open Source JavaScript Library for Front-end Development - <https://www.theseus.fi/handle/10024/130495>

⁴ Fetch API - Web APIs | MDN - https://developer.mozilla.org/en-US/docs/Web/API/Fetch_API



FIGURA 3. Resultados generales de la búsqueda de un journal

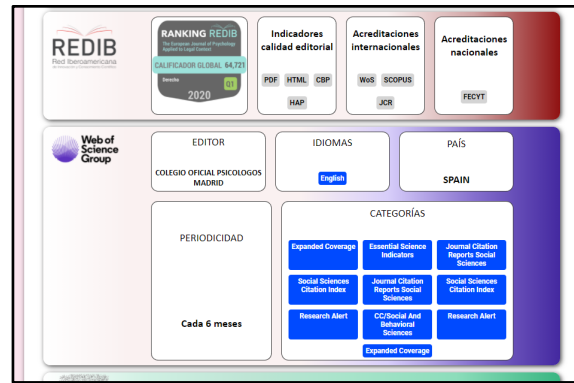


FIGURA 4. Vista expandida de métricas de un journal (1)

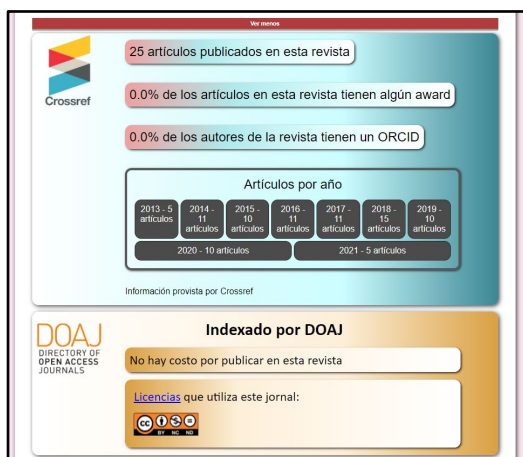


FIGURA 5. Vista expandida de métricas de un journal (2)

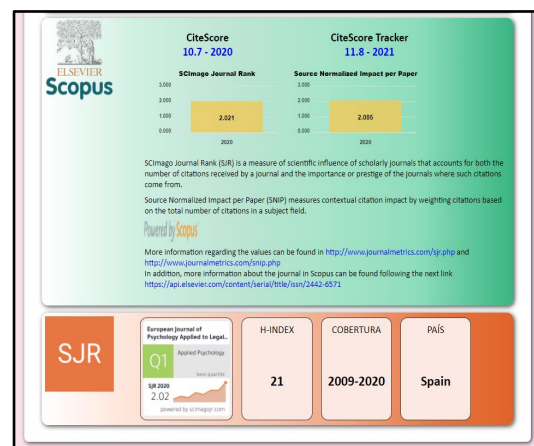


FIGURA 6. Vista expandida de métricas de un journal (3)

Búsqueda de paper

Al 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 7. Entre los resultados se pueden apreciar algunos metadatos de los recursos, como título, tipo de recurso, resumen, entre otros, así como el nombre de la publicación donde se publicó el *paper* y un enlace para dirigirse a su página original. 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 8. 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 trabajo, 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 9.

The screenshot shows the HERA website interface. At the top, there are navigation links for 'HERA', 'Buscar', and 'Preguntas frecuentes'. The main heading is 'Herramienta para Enriquecimiento de Recursos Académicos'. Below this is a search bar with the text 'Ingrese DOI del contenido a analizar' and the input field containing '10.3389/fenvs.2020.561591'. A 'Buscar' button is to the right. Below the search bar, there are radio buttons for 'DOI' (selected) and 'ISSN'. The main content area displays the following information:

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

Tipo de recurso: Journal article

Autores: Sharma Rohit, Kumar Raghendra, Satapathy Suresh Chandra, Al-Ansari Nadhir, Singh Krishna Kant, Mahapatra Rajendra Prasad, Agarwal Anuj Kumar, Le Hai Van, Pham Binh Thai.

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 Coliform, 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, Chlorides, Alkalinity, DO, Temperature, COD, BOD, pH, Magnesium, Hardness, Total Coliform, and Calcium. As per the results and trend analysis, the value of total coliform, 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

Below the abstract, there is a row of six metric cards:

- Crossref:** 7 citas
- DOAJ (Directory of Open Access Journals):** 7 citas
- Microsoft Academic:** Predicciones de citas
- Scilit:** 7 citas
- Scopus:** Sin información de tendencias
- Se habla del tema:** 1

FIGURA 7. Resultados generales de la búsqueda de un paper (vista resumida)

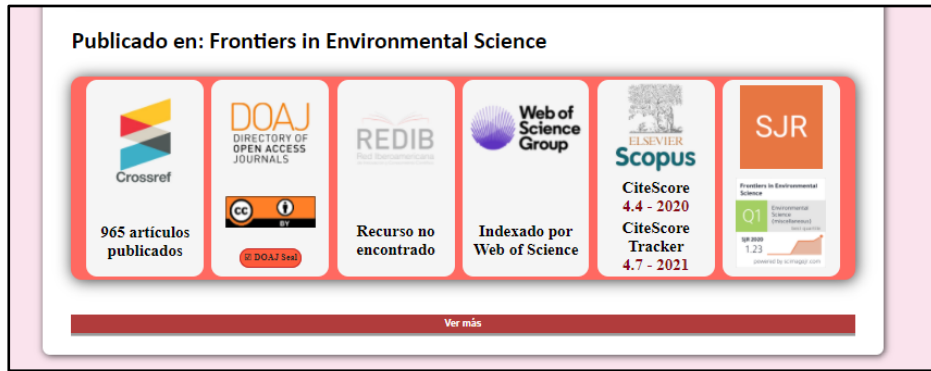


FIGURA 8. Métricas correspondientes al journal donde se encuentra el paper (vista resumida)



FIGURA 9. Vista expandida de métricas de un paper

Conclusiones y trabajos futuros

Determinar la calidad y el impacto de una revista o un artículo científico se torna un desafío cada vez más dificultoso debido a dos factores: (1) el crecimiento exponencial en la cantidad de recursos para analizar gracias al desarrollo tecnológico; y (2) la existencia de múltiples indicadores y métricas de diversa índole para cada clase de recurso. En este contexto, 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 su análisis integral.

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.

Por otro lado, el diseño modular de esta aplicación permite y alienta su crecimiento hacia nuevas y mejores funciones para los usuarios. Entre las nuevas funciones que están siendo consideradas, se destacan:

Expandir el banco de bases de datos académicas: se buscará incorporar nuevos índices y fuentes de datos en el futuro para ser añadidos en la herramienta. Opciones atractivas incluyen la incorporación de índices regionales, como por ejemplo Redalyc, 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 incorporar la funcionalidad de poder seleccionar de qué bases de datos interesa obtener información.

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 directamente desde la página de dicho recurso.
- Generación de archivos de datos que recopilen métricas de varios recursos académicos a la vez. Esta funcionalidad puede resultar ideal para evaluar recursos dentro de un contexto específico, como un proyecto particular, un área temática o una institución.
- Integración con OJS para visualizar métricas desde allí y obtener datos a través de terceros de forma más transparente.

Optimización de tiempo de búsquedas de métricas: sería de utilidad poseer mecanismos de caché para optimizar los tiempos en caso de búsquedas recurrentes. Además, algunas de las API de sitios externos utilizadas tienen cuotas máximas específicas de uso, que se podría evitar alcanzar al tener la información ya almacenada.

Bibliografía

- 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>
- 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.
- REPISO, R. (2015). Cómo identificar una revista de calidad. *Cardiocre*, 50 (2): 46-48. <https://www.redalyc.org/articulo.oa?id=277041630002>
- RHAIEM, M. (2017). Measurement and determinants of academic research efficiency: a systematic review of the evidence. *Scientometrics*, 110: 581-615. <https://doi.org/10.1007/s11192-016-2173-1>