

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

HERA 2.0: Más Funcionalidad para la Evaluación de Recursos Académicos

Ezequiel Carletti, Facultad de Informática, Universidad Nacional de La Plata, Argentina, carlettieze@gmail.com;

Enzo Rucci, III-LIDI, Facultad de Informática, Universidad Nacional de La Plata y Comisión de Investigaciones Científicas, Argentina, erucci@lidi.info.unlp.edu.ar;

Gonzalo Luján Villarreal, PREBI-SEDICI Universidad Nacional de La Plata y CESGI Comisión de Investigaciones Científicas, Argentina, gonzalo@prebi.unlp.edu.ar

Palabras claves

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

Eje temático

3. Evaluación y métricas alternativas

Resumen

En el marco de un escenario académico-científico donde la producción de información crece exponencialmente, la necesidad de herramientas que asistan en la evaluación de la calidad e impacto de los recursos disponibles se torna esencial. En esta línea, la aplicación HERA se erige como un recurso que busca agilizar y respaldar el proceso de valoración de artículos y revistas académicas. Sin embargo, reconociendo que las oportunidades de mejora siempre están presentes, este artículo tiene como objetivo presentar el desarrollo de una segunda versión de HERA, la cual contempla un conjunto de mejoras y extensiones. HERA 2.0 representa una versión sofisticada y extendida de su predecesora, al mejorar su rendimiento, escalabilidad, alcance y soporte. Considerando las características de esta nueva versión, se espera que los miembros de la comunidad académico-científica la encuentren de mayor utilidad para evaluar la calidad y el impacto de los recursos académicos y que contribuya a facilitar y acelerar dicha tarea.

Introducción

En la actualidad, los investigadores enfrentan un verdadero desafío al momento de tener que determinar la calidad y el impacto de los recursos académicos, debido a la combinación de dos factores. En primer lugar, el sostenido crecimiento de publicaciones científicas procedentes del desarrollo tecnológico (Smith, 2013; Sumit, 2024). En segundo lugar, la carencia de estándares y la disponibilidad de múltiples sistemas de evaluación y métricas, que aún cuando comparten objetivos, no siempre utilizan las mismas metodologías (Porto, 2021).

En este contexto, HERA surgió como una respuesta ante estos desafíos, siendo una aplicación web que enriquece recursos académicos (artículos o revistas) al integrar información de diferentes bases de datos académicas (Porto et al., 2022a, 2022b). Esta herramienta automatizada está dirigida principalmente a miembros de la comunidad académico-científica, y busca simplificar, agilizar y respaldar el proceso de determinar la calidad y el impacto de un recurso académico. Como parte de su funcionamiento, HERA consulta

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

múltiples fuentes en tiempo real para luego ofrecer información de un recurso determinado, como ser sus metadatos, su pertenencia a índices y bases de datos, sus indicadores de citas y menciones, e información de la publicación donde figura dicho recurso en caso que corresponda. En particular, HERA se nutre de múltiples bases de datos como CrossRef¹, DOAJ², Scopus³, REDIB⁴, Dimensions⁵, entre otras; y se encuentra disponible para la comunidad en forma abierta⁶.

En su primera versión, HERA resultó funcional y de gran utilidad para los usuarios. Recientemente, se ha lanzado una segunda versión, la cual contempla un conjunto de mejoras y extensiones. A partir de estas, HERA 2.0 busca optimizar su rendimiento, escalabilidad, alcance y soporte.

HERA 1.0

Propósito

HERA es una herramienta diseñada para simplificar, agilizar y apoyar el proceso de determinar la calidad y el impacto de un recurso académico, como un artículo o una revista. Sus características principales son:

- Permite al usuario visualizar de manera sencilla información relevante proveniente de múltiples bases de datos y criterios de expertos. Esto facilita la comprensión y análisis de los datos de una manera rápida y eficiente.
- Permite automatizar el proceso de recopilación de información, lo que evita que el usuario tenga que buscar en cada sitio de forma individual. Esto ahorra tiempo y esfuerzo, permitiendo enfocarse en la interpretación de los datos recabados.
- Permite perfilar rápidamente los datos obtenidos, como el número de citas, factores de impacto, información sobre licencias, ubicación de la publicación y más. Esto facilita la toma de decisiones y permite al usuario realizar un análisis personalizado según sus objetivos e intereses.

En resumen, HERA 1.0 brindaba a los usuarios la posibilidad de centrarse en el análisis y la interpretación de los datos, ahorrando tiempo y esfuerzo en la búsqueda de información.

Diseño y Desarrollo

En la implementación de HERA 1.0, se conformó en una aplicación web organizada en dos desarrollos: una aplicación front-end desarrollada con ReactJS⁷ y una aplicación back-end desarrollada con NodeJS⁸. El propósito

¹ <https://www.crossref.org/>

² <https://doaj.org/>

³ <https://www.scopus.com/>

⁴ <https://www.redib.org/>

⁵ <https://www.dimensions.ai/>

⁶ <http://hera.sedici.unlp.edu.ar>

⁷ <https://react.dev/>

⁸ <https://nodejs.org/en>

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

de la aplicación backend es actuar como intermediaria entre la aplicación frontend y los diferentes sistemas que suministran datos y métricas de recursos académicos, asumiendo así la mayor parte de la carga de trabajo.

La aplicación frontend, por otro lado, solicita información a la aplicación backend a través de una API REST, recopila los resultados y se encarga de generar visualizaciones en un sitio web.

Funcionamiento

El uso de HERA 1.0 es bastante sencillo: el usuario dispone de una barra de búsqueda en la que debe ingresar un identificador, que puede ser el ISSN para analizar una publicación o un DOI para analizar un artículo (ver Figura 1).

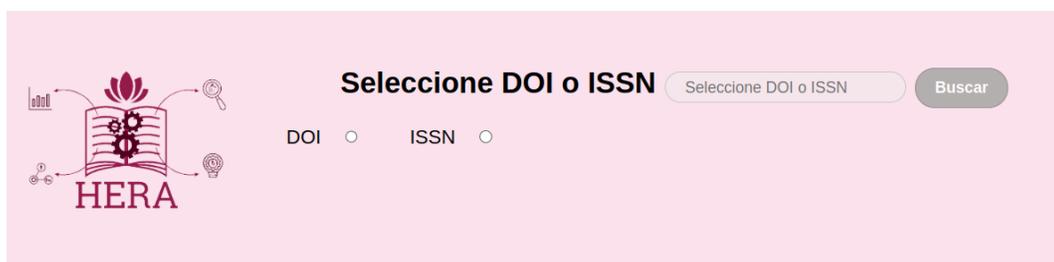


Figura 1: Interfaz de búsqueda de la herramienta con input de búsqueda y selectores de tipo de identificador (DOI e ISSN)

A continuación, se debe hacer clic en el botón "Buscar" para indicarle a HERA que comience la búsqueda de información del recurso asociado a dicho identificador. Luego, HERA 1.0 establece conexiones concurrentes con un conjunto de servicios en línea y sitios web (bases de datos en adelante) para intentar obtener información del recurso de interés. Además, HERA 1.0 determina la mejor estrategia para extraer información de cada base de datos (por ejemplo, una API REST o web scraping), así como también qué tipo de información puede obtener de cada una de ellas: pertenencia o no a la base de datos, citas, gráficos, altmetrics, etc. En (Porto et al., 2022b) se puede consultar el flujo detallado para las diferentes búsquedas posibles.

HERA 2.0: Mejoras y Extensiones

HERA 2.0 mantiene el propósito de su versión previa. Esta nueva versión busca extender y mejorar la funcionalidad, además de optimizar el rendimiento. Estas mejoras y extensiones tienen como meta principal ofrecer una experiencia mejorada tanto para los usuarios finales como para los desarrolladores involucrados en el proyecto.

Actualización del Banco de Bases de Datos

Es fundamental para el valor y utilidad de HERA mantener actualizado el repositorio de bases de datos. Este proceso consta de un análisis meticuloso de las fuentes de datos en uso, la identificación de nuevas fuentes de datos potenciales, y la implementación de medidas que aseguren un servicio continuo.

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

Evaluación de las Fuentes de Datos de HERA 1.0

Se realizaron revisiones de todos los servicios que emplea HERA 1.0, para asegurar que estos siguen proporcionando datos pertinentes, manteniendo sus API accesibles y que cualquier cambio en su estructura de datos o términos de uso no afecta adversamente la funcionalidad de la herramienta. En particular, se revisaron los servicios provistos por Crossref, DOAJ, Dimensions, Altmetrics, SemanticScholar, Scopus, WoS, SJR, Microsoft Academic y Google Scholar.

Durante el proceso anterior, se constató que Microsoft Academic dejó de estar operativo, tal como se anunció en un comunicado oficial en mayo de 2021⁹, indicando que su sitio web y las API subyacentes se retirarían el 31 de diciembre de 2021. Por tanto, Microsoft Academic fue eliminado de la lista de fuentes de datos utilizadas en HERA 2.0.

Adicionalmente, se identificó un problema temporal con la fuente de datos REDIB, la cual, según un comunicado oficial, podría tener interrupciones temporales o retrasos en la actualización de contenidos debido a cambios en la administración y la financiación del proyecto¹⁰. A pesar de estos inconvenientes, en HERA 2.0 se decidió mantener REDIB como fuente de datos, aunque actualmente no está respondiendo.

Identificación de Nuevas Fuentes de Datos

Además de la revisión de las fuentes de datos actuales, se identificó y añadió una nueva fuente de datos para mejorar la funcionalidad y cobertura de HERA 2.0. En particular, se trata de OpenAlex (Priem et al., 2022), la cual es una plataforma de datos académicos de acceso abierto desarrollada por la organización sin fines de lucro OurResearch¹¹. OpenAlex fue creada como una alternativa a Microsoft Academic Graph (MAG) y ofrece una base de datos amplia y actualizada de información académica que incluye publicaciones, autores, afiliaciones, citas y otros datos relacionados con la investigación.

Esta nueva versión de HERA emplea OpenAlex como su principal fuente de datos, en lugar de basarse principalmente en Crossref como en la versión 1.0, considerando que OpenAlex ya contiene la información de Crossref. Además, este cambio permitió ampliar las capacidades de HERA para incluir información y enlaces a PubMed¹² y PubMed Central¹³.

Por otra parte, Google Scholar (GS) reúne una vasta cantidad de información académica, facilitando el acceso a una diversidad de artículos científicos. Sin embargo, la integración de GS plantea una serie de desafíos técnicos, legales y éticos que tuvieron que ser considerados en el desarrollo de esta nueva versión de la herramienta. Debido a estas limitaciones, se decidió agregar un botón en la interfaz de usuario que redirige al usuario a la página de GS para el DOI buscado (ver Figura 4).

⁹ [Comunicado Microsoft Academic](#)

¹⁰ [Comunicado Oficial de REDIB](#)

¹¹ <http://ourresearch.org/>

¹² [PubMed - Sitio Web](#)

¹³ [PubMed Central - Sitio Web](#)

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

Por último, si bien Scopus estaba presente para revistas y congresos, se decidió también incluirla en la búsqueda de DOI. Aunque Scopus provee una API para esta funcionalidad, se decidió seguir el mismo enfoque que con GS por cuestiones operativas y de seguridad (Carletti, 2023).

Nueva Interfaz

Para la implementación de la nueva interfaz en HERA 2.0, se optó por incorporar el uso de React-Bootstrap¹⁴, una combinación de Bootstrap y ReactJS. Bootstrap es un marco de diseño popular y ampliamente utilizado que proporciona una serie de estilos y componentes predefinidos para facilitar el desarrollo de interfaces de usuario ricas y atractivas¹⁵. ReactJS, por otro lado, es una biblioteca de JavaScript para la construcción de interfaces de usuario de manera eficiente y flexible¹⁶.

La incorporación del paquete React-Bootstrap en la nueva versión de HERA ha producido resultados significativos en términos de diseño y usabilidad. React-Bootstrap, ha permitido una interfaz más intuitiva y atractiva. Mediante este cambio se mejora la apariencia de la aplicación, facilitando su uso en diferentes dispositivos y resoluciones de pantalla. Se considera que la integración de React-Bootstrap ha simplificado el proceso de desarrollo, proporcionando un conjunto coherente y estandarizado de componentes estilizados y responsivos. Esto ha mejorado la cohesión visual y la consistencia de diseño en toda la aplicación, proporcionando una experiencia de usuario más coherente.

En la aplicación, se presenta una interfaz de usuario única que se caracteriza por la presencia de una destacada barra de búsqueda. En esta barra, los usuarios deben ingresar el identificador del recurso que desean buscar, tal como se ilustra en la Figura 2.



Figura 2: Interfaz de búsqueda de la herramienta con input de búsqueda y selectores de tipo de identificador (DOI e ISSN)

Dentro de la interfaz, los usuarios deben elegir el tipo de búsqueda que desean realizar, ya sea mediante el uso del DOI o del ISSN. Una vez seleccionado el tipo de búsqueda, deberán ingresar el identificador correspondiente en la barra designada. Más allá de la nueva interfaz, el flujo de búsqueda es el mismo que en el caso de HERA 1.0 en general. El único cambio consiste en la posibilidad de introducir múltiples identificadores en la barra de búsqueda (se deben separar por comas) para permitir su búsqueda simultánea.

La Figura 3 exhibe los resultados para un artículo de ejemplo. Entre los resultados, se pueden observar varios metadatos del recurso, como el título, el tipo de recurso, el resumen, entre otros. Además, se muestra el

¹⁴ <https://react-bootstrap.netlify.app/>

¹⁵ <https://getbootstrap.com/>

¹⁶ <https://es.react.dev/>

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

nombre de la revista en la que se publicó el artículo, junto con un enlace que dirige a la página original del mismo. En caso de ser posible, la aplicación también intentará recopilar métricas del journal en el que se encuentra el artículo, y las mostrará en conjunto, como se muestra en la Figura 4. Para esto, se extrae el ISSN del journal de los metadatos recuperados y se realiza automáticamente una búsqueda por ISSN en segundo plano. De esta manera, se amplía el contexto de evaluación de un artículo, permitiendo ver sus métricas de forma aislada y también acceder a las métricas del journal, lo que ayuda a comprender qué influencia puede tener en las métricas del artículo. Al hacer clic en el botón "Ver más" que se muestra en esta última figura, se genera una vista expandida de las métricas, donde se puede obtener información más detallada (en los casos en que las bases de datos proporcionen datos adicionales), como se ilustra en las Figuras 5, 6 y 7.

The screenshot shows the HERA search interface. At the top, there is a search bar with the input '10.7717/peerj.4375' and buttons for 'DOI', 'ISSN', and 'Buscar'. Below the search bar, there are options for 'Sólo exportar' and two 'Exportar' buttons. The main content area displays the article title 'The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles' and the DOI 'DOI/10.7717/peerj.4375'. The authors listed are Heather A. Piwowar, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Matthias, Bree Norlander, Ashley Farley, Jevin D. West, and Stefanie Haustein, each with their respective institutions. The publication year is 2018, the journal is PeerJ, and the article type is journal-article. A summary section follows, discussing the need for large-scale, up-to-date studies on Open Access (OA) literature and the results of a study that investigated OA in three populations: all journal articles assigned a Crossref DOI, recent journal articles indexed in Web of Science, and articles viewed by users of Unpaywall.

Figura 3. Resultados generales de la búsqueda de un *paper* (vista resumida)

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

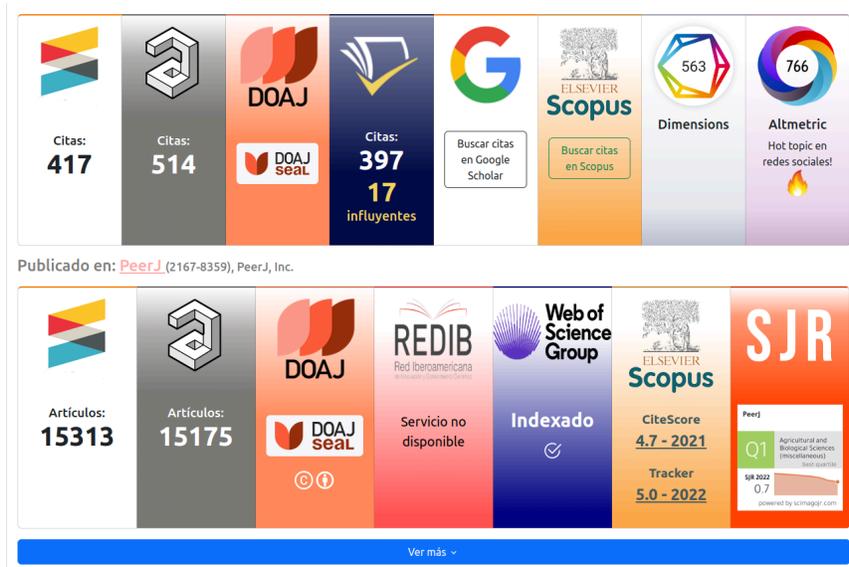
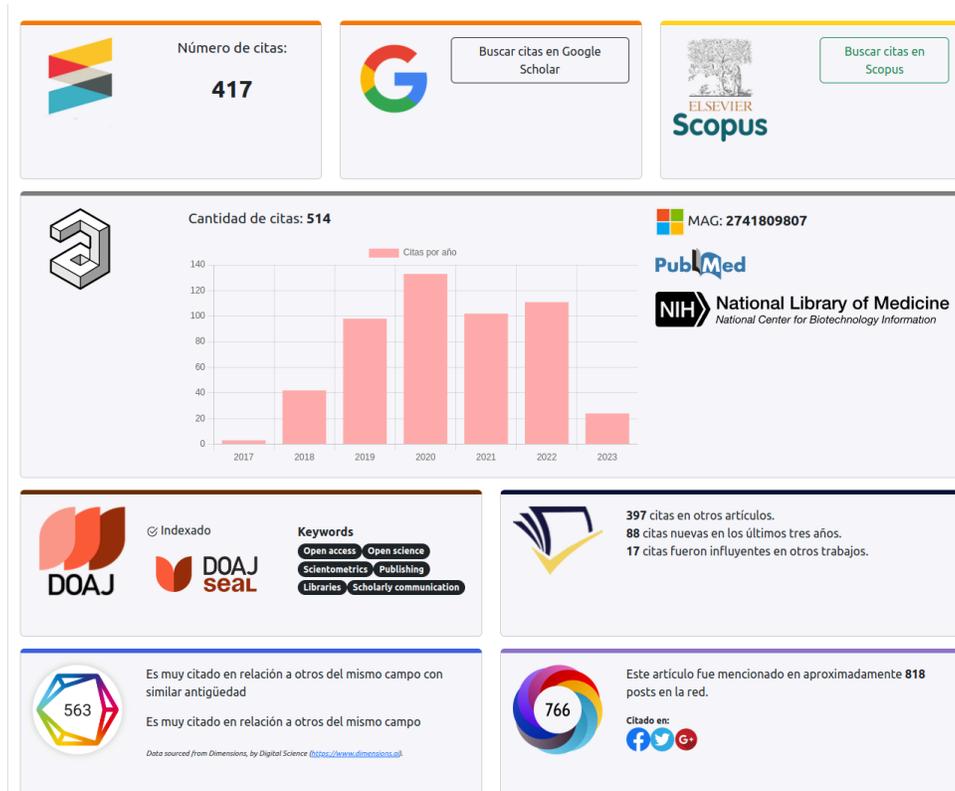


Figura 4. Métricas correspondientes al *paper* y *journal* donde se encuentra el *paper* (vista resumida en tarjetas)



BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

Figura 5. Vista expandida de métricas de un paper

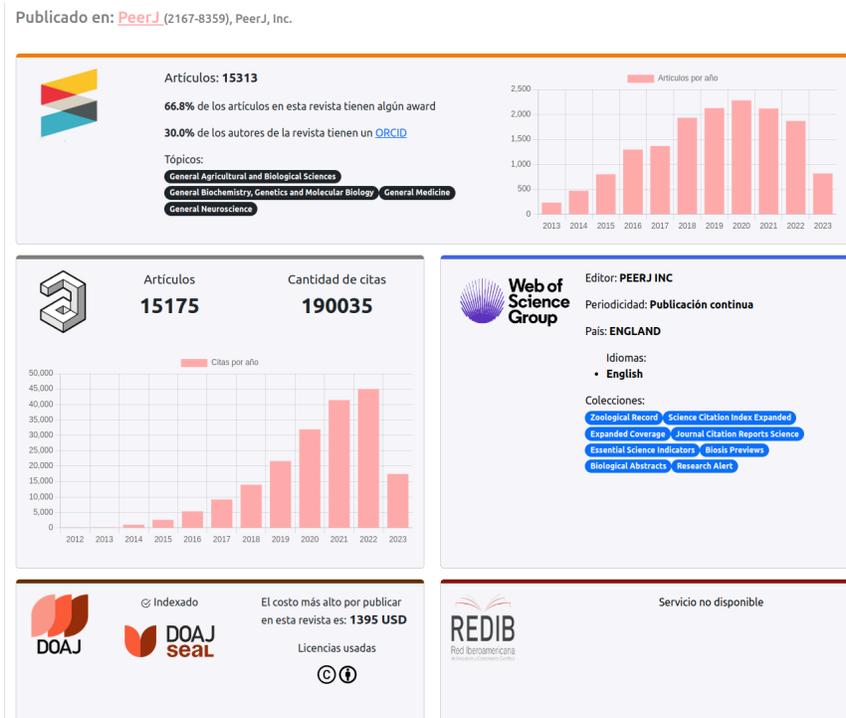


Figura 6. Vista expandida de métricas de un journal (parte 1)



Figura 7. Vista expandida de métricas de un journal (parte 2)

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

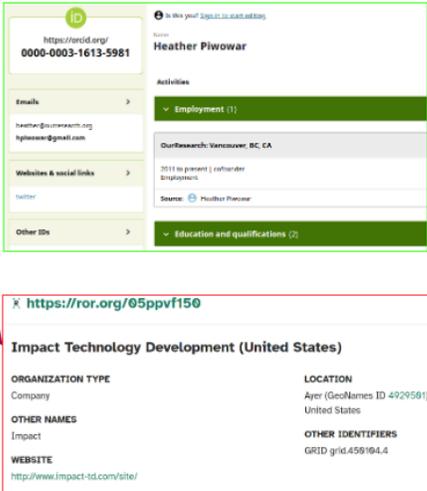
Integración de Información de Autores e Instituciones

Para integrar una mayor cantidad de información sobre los autores de una publicación en la aplicación HERA, se contempló la utilización de dos fuentes de datos adicionales: Research Organization Registry (ROR)¹⁷ y Open Researcher and Contributor ID (ORCID)¹⁸. Por un lado, con la integración de la información de ROR, HERA ahora puede mostrar datos detallados de la organización o institución a la que pertenecen los autores. Esto ofrece a los usuarios un contexto más completo y detallado sobre el entorno de investigación del autor. Por otro lado, al integrar la información de ORCID, HERA ahora puede proporcionar un perfil más completo de cada autor, incluyendo sus publicaciones anteriores, afiliaciones y contribuciones. Este nivel de detalle adicional proporciona a los usuarios una visión más profunda del trabajo y la trayectoria de cada autor, lo que mejora la utilidad y la riqueza de la información que HERA 2.0 puede proporcionar.

La Figura 8 ilustra los beneficios de la integración mencionada, aunque es importante destacar que esta información adicional solo se muestra cuando está disponible para un autor específico.

Autores:

- **Heather Piwowar**  [Institución: *Impact Technology Development \(United States\)*](#)
- **Jason Priem**  [Institución: *Impact Technology Development \(United States\)*](#)
- **Vincent Larivière**  [Institución: *Université de Montréal*](#)
- **Juan Pablo Alperin** [Institución: *Simon Fraser University*](#)
- **Lisa Matthias**  [Institución: *Simon Fraser University*](#)
- **Bree Norlander**  [Institución: *University of Washington*](#)
- **Ashley Farley**  [Institución: *University of Washington*](#)
- **Jevin D. West**  [Institución: *University of Washington*](#)
- **Stefanie Haustein**  [Institución: *University of Quebec at Montreal*](#)



<https://ror.org/05ppvf150>

Impact Technology Development (United States)

ORGANIZATION TYPE	LOCATION
Company	Ayer (GeoNames ID 4929501)
	United States
OTHER NAMES	OTHER IDENTIFIERS
Impact	GRID grid.459104.4
WEBSITE	
http://www.impact-td.com/site/	

Figura 8: Ejemplo de vista de autores en HERA 2.0.

Exportación a Formatos Procesables

La implementación de esta funcionalidad proporciona una considerable ventaja para los usuarios que necesiten realizar análisis de datos, investigaciones académicas, auditorías bibliográficas, entre otras aplicaciones. En particular, se consideraron dos formatos que son ampliamente usados en la actualidad, como CSV y JSON.

¹⁷ [ROR](#) es un registro global y comunitario que proporciona identificadores únicos para las organizaciones que participan en la producción de investigación.

¹⁸ [ORCID](#) proporciona un identificador digital persistente que distingue a un investigador de otros.

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

La exportación de datos permite a los usuarios obtener los principales indicadores y métricas para un recurso académico (tanto para DOI como para ISSN) de forma conjunta y estructurada. Cuando se realizan búsquedas simultáneas, la información recolectada de cada DOI o ISSN se presenta en filas separadas, facilitando su lectura y posterior procesamiento. En la Figura 9, se muestra un ejemplo de una descarga en formato CSV correspondiente a 10 DOIs.

	doi	type	title	authors	abstract	publication_year
1	10.3389/fenvs.2020.581591	journal-article	Analysis of Water Pollution Us	Rohit Sharma,Raghvendra Kuma	The Yamuna river has bec	2020
2	10.1109/LCOMM.2019.2898944	journal-article	Deep Learning-Based Channe	Mehran Soltani,Vahid Pourahmad	In this letter, we present a	2019
3	10.1016/j.telpol.2021.102261	journal-article	Internet of things and the econ	Günter Knieps,Johannes M. Bau	Fifth generation (5G) netw	2022
4	10.24215/26838559e28	journal-article	Diabetes Link: innovación tec	Enzo Rucci,Lisandro Nahuel Deli	La Diabetes Mellitus (DM)	2021
5	10.1007/978-3-031-07802-6_9	book-chapter	Migrating CUDA to oneAPI: A	Manuel Costanzo,Enzo Rucci,Ca	In order to tackle the prog	2022
6	10.1109/CLEI53233.2021.9640225	proceedings-article	Performance vs Programming	Manuel Costanzo,Enzo Rucci,Ma	Historically, Fortran and C	2021
7	10.1007/978-3-030-61702-8_25	book-chapter	Diabetes Link: Platform for Se	Enzo Rucci,Lisandro Nahuel Deli	Diabetes Mellitus (DM) is	2020
8	10.7717/peerj.4375	journal-article	The state of OA: a large-scale	Heather A. Piowar,Jason Priem	Despite growing interest i	2018
9	10.24215/18522971e087	journal-article	Desafíos de la videovigilancia	Andrés Herrera Esquivel	La videovigilancia automa	2021
10	10.5710/PEAPA.21.05.2020.298	journal-article	NEUROANATOMÍA DEL SAU	Ariana Paulina-Carabajal,Leonar	this taxon has uncertain p	2020
	crossref_cites	openalex_doi_cites	doaj_doi_presence	semantic scholar_cites	dimensions_cites	altmetric_cites
1	33	33	Si (Si)	19	51	1
2	261	281	No	278	278	2
3	7	6	No	8	11	1
4	0	0	No	0	0	2
5	2	2	No	5	6	17
6	1	1	No	1	2	
7	0	0	No	0	0	
8	434	541	Si (Si)	411	588	817
9	0	0	No		0	
10	1	1	No	1	2	1

Figura 9: Datos DOI en formato CSV sin la información de sus correspondientes ISSN (modificados para facilitar su visualización en una sola imagen)

Manejo de Errores y Recursos no Encontrados

Un aspecto crítico de cualquier sistema de software es cómo se manejan los errores y las situaciones en las que no se encuentran los recursos solicitados. En HERA 2.0 se ha robustecido la gestión de errores en combinación con una comunicación más clara hacia los usuarios cuando ocurren situaciones excepcionales.

Existen diferentes casos en los que puede surgir un error o no encontrarse un recurso en el contexto de la aplicación. Algunos ejemplos comunes son:

- **Recurso no encontrado:** ocurre cuando el usuario realiza una búsqueda de un DOI o un ISSN que no existe en ninguna de las bases de datos consultadas. En este caso, HERA 2.0 devuelve un mensaje claro al usuario indicando que el recurso solicitado no se encontró. En la Figura 10, la fuente de información DOAJ responde con el mensaje de recurso no encontrado:

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024



Figura 10: Métricas de DOAJ no encontradas

- Servidor no disponible: surge en situaciones en las que se intenta hacer una solicitud a una base de datos y el servidor de dicho banco de datos no responde o no está disponible en ese momento. En este caso, HERA 2.0 informa al usuario que el servidor está actualmente no disponible (ver Figura 11).



Figura 11: Tarjeta de REDIB informando: Servidor no disponible.

- Límite de solicitudes alcanzado: sucede en circunstancias donde se han realizado demasiadas solicitudes a una base de datos en un período corto de tiempo y se ha alcanzado el límite establecido por esa API externa. Este es un caso común con las API que tienen un límite en el número de solicitudes que se pueden hacer en un cierto período de tiempo. En este caso, HERA 2.0 informa al usuario que se ha alcanzado el límite de solicitudes para esa base de datos y que debe esperar un cierto tiempo antes de realizar más solicitudes (ver Figura 12).

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024



Figura 12: Tarjeta de OpenAlex informa: Límite de solicitudes alcanzado

Este sistema de manejo de errores y notificaciones no sólo mejora la experiencia del usuario, sino que también aumenta la robustez de la aplicación, ya que puede manejar situaciones excepcionales y proporcionar información útil al usuario en caso de que ocurran.

Búsqueda simultánea de múltiples recursos

El traslado de la lógica de negocio del *frontend* al *backend* en HERA se considera un avance en términos de rendimiento y tiempo de respuesta (ver sección siguiente). En la versión 1.0, para un recurso como un DOI, se debían realizar hasta siete consultas, y siete adicionales si el DOI contaba con un ISSN correspondiente. Estas consultas, se realizaban de manera secuencial por efectuarse en el frontend, lo que implicaba un consumo considerable de tiempo.

En HERA 2.0, estas peticiones se gestionan mediante el framework de aplicaciones web Express en el backend, permitiendo su ejecución concurrente (paralela si el hardware subyacente lo permite) en lugar de secuencial. Este cambio en la estrategia de ejecución ha permitido una optimización en la eficiencia y velocidad de las consultas (Carletti, 2023). Al realizar múltiples solicitudes en concurrente, se reduce el tiempo de espera asociado con las consultas secuenciales, mejorando así la experiencia del usuario y el rendimiento global de la herramienta.

Pese a que la búsqueda simultánea de recursos se optimizó a través de consultas concurrentes entre los DOI/ISSN, se enfrentó una limitación: cada API externa impone restricciones en cuanto a la cantidad de consultas que pueden realizarse en un período de tiempo determinado. Para superar este desafío, se implementó un sistema de búfer para regular la cantidad de consultas simultáneas. Por ejemplo, en lugar de buscar 30 artículos a la vez, se estableció un búfer para realizar las consultas en lotes de 5 a la vez. Este enfoque, a pesar de incorporar un grado de secuencialidad, permite mantener un alto rendimiento sin exceder las limitaciones impuestas por las API externas.

Refactorización de código

La refactorización de código es una técnica esencial para mantener la calidad del software, permitiendo mejorar la estructura interna del código sin cambiar su comportamiento externo (Méndez, 2010; Team, 2021). En HERA 1.0 se encontraron varios problemas o posibles oportunidades de optimización, incluyendo:

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

- Falta de modularización: la estructura del código de HERA 1.0 carece de modularización, lo cual puede complicar la legibilidad y mantenimiento del sistema.
- Manejo de múltiples fuentes de datos: la gestión de múltiples fuentes de datos presenta margen para optimización, mejorando así la eficiencia en su procesamiento.
- Lógica de negocio: la implementación actual de la lógica de negocio en HERA 1.0 muestra áreas de mejora, con posibles optimizaciones para incrementar la funcionalidad de la aplicación.

Es por ello que llevó a cabo una refactorización integral del código de HERA para mejorar su eficiencia, mantenibilidad y escalabilidad. Más detalles pueden ser consultados en (Carletti, 2023).

Implementación de API REST

La construcción de la API REST en HERA constituye una actualización estratégicamente significativa, la cual contribuye a una mayor integración con aplicaciones y servicios externos (ver Figura 13), a la optimización del rendimiento en las solicitudes y al aumento de la escalabilidad de la plataforma.

- Interoperabilidad y flexibilidad: la adopción de una API REST ofrece a HERA 2.0 la capacidad de interactuar con mayor eficiencia con las diferentes fuentes de datos que utiliza, como OpenAlex, Crossref y DOAJ. Este enfoque facilita la incorporación de nuevas fuentes de datos en el futuro y la flexibilidad requerida para adaptarse a los cambios y evolución en el ecosistema de datos. Al adherirse a los principios de la arquitectura REST, la API permite la interacción de otras aplicaciones con HERA, potencialmente generando nuevas vías para la utilización y crecimiento de la plataforma.
- Eficiencia: la implementación de una API REST mejora la eficiencia de la plataforma. Las API REST emplean métodos HTTP estándar (GET, POST, PUT, DELETE), los cuales son ampliamente adoptados y optimizados en la industria. Adicionalmente, una API REST es capaz de manejar múltiples tipos de datos (como JSON o CSV), lo que permite a los clientes solicitar los datos en el formato que mejor se adapte a sus requerimientos y a la naturaleza de sus operaciones.
- Escalabilidad: en términos de escalabilidad, una API REST es capaz de manejar una gran cantidad de solicitudes y sustentar el crecimiento de la plataforma sin afectar su rendimiento. Las API REST son stateless¹⁹, lo que significa que cada solicitud es autónoma y no depende de la información de solicitudes previas. Esta característica facilita la distribución de las solicitudes entre múltiples servidores, mejorando así el rendimiento y la capacidad de respuesta de la plataforma.
- Gestión de múltiples consultas: una característica adicional de la API REST de HERA es su capacidad de generar múltiples búsquedas y proporcionar múltiples resultados simultáneamente. Esta funcionalidad es fundamental para el perfilamiento eficiente de los datos, permitiendo obtener una visión amplia y detallada de los conjuntos de datos con los que se trabaja.

La incorporación de la API REST permitió el diseño del frontend y de un complemento para el navegador Google Chrome como las primeras aplicaciones clientes livianas y externas al core de HERA.

¹⁹ Protocolo sin estado.

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

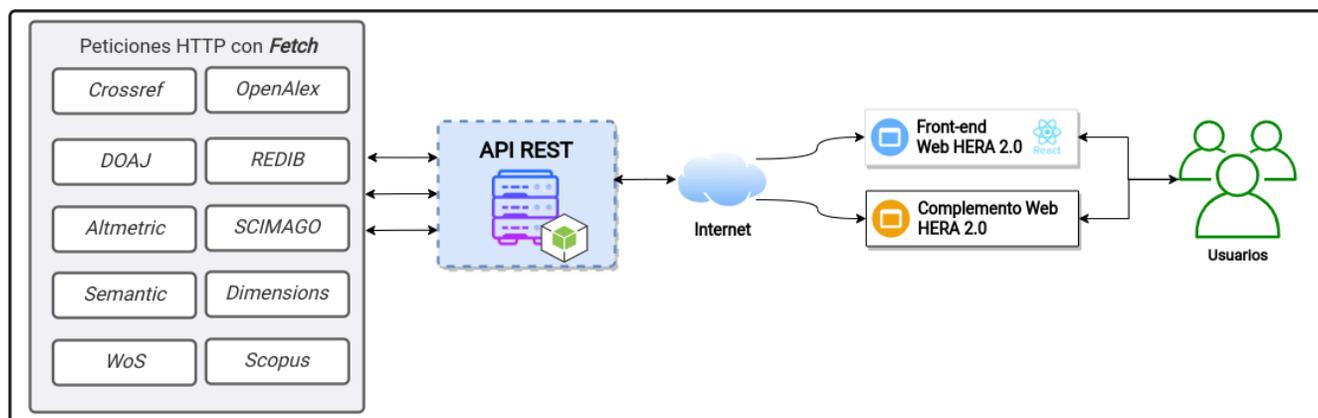


Figura 13: Esquema de comunicación de la API REST

Gestión de las fuentes de información

En el proceso de reestructuración y mejora del código de HERA, una cuestión de relevancia era garantizar la flexibilidad del sistema frente a posibles cambios en las fuentes de información. Con este fin, se diseñó un archivo de configuración que almacena información detallada de cada fuente de datos externa, entre la que se incluye un identificador único de cada fuente (por ejemplo, "CrossrefDOI"), su correspondiente URL y un indicador booleano "enabled" para señalar si la fuente está actualmente activa o no.

En la Tabla 1, se presenta un ejemplo simplificado del archivo de configuración. Dicha tabla sirve como guía para la interacción del sistema con las diferentes fuentes de datos y componentes, especificando los parámetros necesarios para su correcta operación.

Fuente de datos	URL	enabled
OpenAlexDOI	https://api.openalex.org/works?filter=doi:https://doi.org/	true
CrossrefDOI	https://api.crossref.org/works/	true
DoajDOI	https://doaj.org/api/v2/search/articles/doi:	true
Microsoft Academic	https://www.microsoft.com/en-us/research/project/academic/	false
ScimagoISSN	https://www.scimagojr.com/journalsearch.php?q=	true

Tabla 1: Representación resumida del archivo de configuración.

En primer lugar este enfoque externaliza los detalles específicos de cada fuente de datos del código base, proporcionando un alto grado de modularidad. Así, la inclusión o exclusión de fuentes de datos se simplifica enormemente, requiriendo solo la modificación del valor del campo "enabled" en la configuración correspondiente. Esta flexibilidad probó su valor en situaciones como fue el caso con Microsoft Academic, que cesó su funcionamiento durante esta etapa de desarrollo de HERA. En lugar de requerir una revisión exhaustiva del código para eliminar referencias a dicha fuente, fue suficiente con cambiar su estado a "inhabilitado" en el archivo de configuración.

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

En segundo lugar, esta implementación aumenta la robustez del sistema. Si una API deja de estar disponible o experimenta interrupciones, se puede desactivar fácilmente en la configuración sin que esto afecte a la funcionalidad general de HERA. En lugar de causar errores o problemas de rendimiento, la falta de disponibilidad de una API se traduce simplemente en que esa fuente de datos no se incluye en el contexto generado.

Conclusiones y trabajos futuros

HERA proporciona una herramienta de gran utilidad para simplificar y agilizar el proceso de recolección de indicadores actualizados sobre la calidad de los recursos académicos y el impacto que estos están generando en la comunidad académica internacional. HERA 2.0 representa una versión sofisticada y extendida de su predecesora, al mejorar su rendimiento, escalabilidad, alcance y soporte. Considerando las características de esta nueva versión, se espera que los miembros de la comunidad académico-científica la encuentren de mayor utilidad para evaluar la calidad y el impacto de los recursos académicos y que contribuya a facilitar y acelerar dicha tarea.

Las tareas de refactorización de código, el rediseño de la arquitectura de la aplicación y la incorporación de la API REST, realizadas en la última versión, sientan las bases para considerar nuevos desarrollos que permitan ampliar el alcance y brindar mayor escalabilidad para soportar un mayor número de usuarios. Entre los desarrollos a futuro se están considerando extensiones similares a la ya desarrollada para Google Chrome, para los navegadores web Firefox, Edge y Safari. Asimismo, los indicadores recuperados por HERA cobran una gran relevancia en contextos donde se gestionan documentos académicos, como por ejemplo repositorios institucionales, portales de revistas científicas o sitios web institucionales. En la siguiente etapa del desarrollo se propondrán integraciones para sistemas de gestión de contenidos, como por ejemplo Wordpress o Drupal, de gestión editorial, como por ejemplo Open Journal Systems, y plataformas para repositorios digitales, como por ejemplo Dspace. Finalmente, a fin de soportar una mayor carga en cuanto a la cantidad de consultas que deberá atender HERA en el futuro, así como también de evitar o minimizar las penalizaciones por superar los límites de consultas a las fuentes de datos, se evaluará la incorporación de un módulo de almacenamiento de indicadores en una caché propia. Este módulo permitirá a HERA enviar indicadores de recursos para los que ya había recibido consultas, sin tener que volver a consultar las fuentes de datos primarias. La implementación de este módulo requerirá evaluar los tiempos de expiración de los datos en caché para cada fuente de datos, así como también la manera de almacenar y gestionar el volumen de información que será necesario alojar.

Referencias

- Carletti, E. (2023). *HERA 2.0: Extensión de alcance y funcionalidad* [Tesina de Licenciatura en Informática, Universidad Nacional de La Plata]. <http://sedici.unlp.edu.ar/handle/10915/157417>
- Méndez, M. (2010). *Refactoring de código estructurado* [Tesis, Universidad Nacional de La Plata].

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

<http://sedici.unlp.edu.ar/handle/10915/4184>

Porto, J. F. (2021). *HERA: Herramienta para Enriquecimiento de Recursos Académicos* [Tesina de Licenciatura en Sistemas, Universidad Nacional de La Plata]. <http://sedici.unlp.edu.ar/handle/10915/129874>

Porto, J. F., Rucci, E., & Villarreal, G. L. (2022a). HERA, una Herramienta para la Evaluación de Recursos Académicos. *Actas del XXVIII Congreso Argentino de Ciencias de la Computación (CACIC 2022)*, 546-557. <http://sedici.unlp.edu.ar/handle/10915/149623>

Porto, J. F., Rucci, E., & Villarreal, G. L. (2022b). HERA-Herramienta para Enriquecimiento de Recursos Académicos. *Actas de la XI Conferencia Internacional de Bibliotecas y Repositorios Digitales (BIREDIAL-ISTEC)*, 251-266. <http://sedici.unlp.edu.ar/handle/10915/148922>

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts* (arXiv:2205.01833). arXiv. <https://doi.org/10.48550/arXiv.2205.01833>

Smith, R. (2013, julio 1). *How Technology is Changing Academic Research* | WIRED.

<https://www.wired.com/insights/2013/07/how-technology-is-changing-academic-research/>

Sumit. (2024, marzo 22). *How Does Technology Help In Doing Academic Research?* | Scoop Byte.

<https://www.scoopbyte.com/how-does-technology-help-in-doing-academic-research/>

Team, L. (2021, diciembre 3). *What Is Code Refactoring? Definition, Benefits and Why It's Important*. Lvivity.

<https://lvivity.com/what-is-code-refactoring>

ANEXO 1

BIREDIAL-ISTEC 2024

XIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 22 al 24 de octubre de 2024

Resumen biográfico de los autores

Ezequiel Carletti es Analista Programador Universitario y estudiante de Licenciatura en Informática por la Facultad de Informática de la Universidad Nacional de La Plata. En el año 2023 presentó su tesis de grado “HERA 2.0: Extensión de alcance y funcionalidad”, proyecto realizado bajo la dirección del Dr. Enzo Rucci y la codirección del Dr. Gonzalo Villarreal, y en colaboración con la Facultad de Ciencias Económicas, la Facultad de Informática y PREBI-SEDICI, dependientes de la UNLP.

Enzo Rucci (<https://orcid.org/0000-0001-6736-7358>) es Doctor en Ciencias Informáticas por la Facultad de Informática de la Universidad Nacional de La Plata, docente-investigador de la UNLP en las áreas relacionadas con procesamiento concurrente y paralelo, e investigador adjunto de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires. Forma parte del Instituto de Investigación en Informática LIDI (III-LIDI, UNLP-CIC), donde realiza sus actividades de investigación en temáticas vinculadas a Cómputo de Alto Rendimiento y Aplicación de TICs a Ciencias de la Vida y Bibliometría. Es co-editor del Journal of Computer Science and Technology.

Gonzalo Luján Villarreal (<https://orcid.org/0000-0002-3602-8211>) es Doctor en Ciencias Informáticas por la Facultad de Informática de la Universidad Nacional de La Plata, es docente-investigador de la UNLP y director del Centro de Servicios en Gestión de Información (CESGI) de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires. Desarrolla su actividad docente en grado y posgrado de la Facultad de Informática de la UNLP. Trabaja con la plataforma OJS desde el año 2008, cuando en el marco de PREBI-SEDICI la UNLP lanzó el Portal de Revistas de la UNLP. En la actualidad es coordinador técnico de revistas científicas de la UNLP, y brinda asesoramiento a organizaciones y equipos editoriales temas vinculados a publicaciones científicas, circuitos editoriales y gestión de OJS.

ANEXO 2

Requerimientos de equipo técnico para la presentación de la ponencia

Para la presentación se requiere proyector y computadora con conexión a internet.