



UNIVERSIDAD
NACIONAL
DE LA PLATA

FACULTAD DE INFORMÁTICA

TESINA DE LICENCIATURA

TÍTULO: HERA 2.0: Extensión de Alcance y Funcionalidad

AUTORES: Ezequiel Carletti

DIRECTOR: Enzo Rucci

CODIRECTOR: Gonzalo Villarreal

ASESOR PROFESIONAL:

CARRERA: Licenciatura en informática

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, esta tesina tiene como objetivo principal el desarrollo de una segunda versión de HERA, aspirando a optimizar su rendimiento, escalabilidad, alcance y soporte.

Palabras Clave

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

Conclusiones

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.

Trabajos Realizados

Se realizó un análisis exhaustivo de diversas bases de datos académicas para conocer el contenido disponible, sus métricas y medios de acceso. Luego, se emprendió un estudio de la herramienta HERA 1.0, buscando identificar sus puntos fuertes y las áreas de mejora, con el fin de proponer y realizar modificaciones que pudieran extender su funcionamiento y mejorar su rendimiento y usabilidad. A partir del trabajo anterior se diseñó y desarrolló HERA 2.0, considerando extensiones y mejoras como rediseño de la interfaz de usuario, integración de información adicional de autores, actualización del banco de bases de datos, refactorización del código, normalización de la gestión de las bases de datos de consulta, desarrollo de una API REST, búsqueda simultánea de recursos, exportación de la información, extensión para navegadores y manejo de errores y recursos no encontrados.

Trabajos Futuros

*Implementación de una base de datos local para un mejor control y rendimiento del sistema.
Ampliación de las funcionalidades de búsqueda en Integración de un sistema de caché para acelerar las operaciones de búsqueda y recuperación de datos.
Desarrollo de un sistema de monitoreo de la disponibilidad de las API para gestionar de manera eficiente los problemas de disponibilidad.
Ampliación de la funcionalidad de la extensión web a otros navegadores para aumentar la accesibilidad de HERA.
Desarrollo de un plugin para la integración con OJS.
Creación de un sistema de configuración dinámica de las API para adaptar la configuración en tiempo real según el rendimiento y disponibilidad.*

Fecha de la presentación: AGOSTO 2023

**FACULTAD DE INFORMÁTICA
UNIVERSIDAD NACIONAL DE LA PLATA**



HERA 2.0: Extensión de Alcance y Funcionalidad

Tesina de Grado

Licenciatura en Informática

Autor: *Ezequiel Carletti*

Director: *Enzo Rucci*

Codirector: *Gonzalo Luján Villarreal*

2023

Agradecimientos

Me gustaría comenzar este apartado con un fuerte agradecimiento a mi director, Enzo. Siempre supo dar buenos consejos, muy predispuesto, un guía que contribuyó enormemente en esta tesis.

A Gonzalo, mi codirector. Por haberme presentado y propuesto este proyecto, siempre dispuesto a resolver cualquier duda y ayudarme a superar mis inquietudes, y no solo con respecto a la tesis, sino a lo largo de toda la carrera. Un referente tanto en lo profesional como en lo personal, y solo espero que este trabajo esté a la altura de sus expectativas.

Y, por supuesto, no puedo dejar de agradecer a mi familia, mi novia y mis amigos, que han sido un gran soporte en todo momento, durante la elaboración de esta tesina y durante toda mi carrera.

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, esta tesina tiene como objetivo principal el desarrollo de una segunda versión de HERA, aspirando a optimizar su rendimiento, escalabilidad, alcance y soporte.

Inicialmente, se realizó un análisis exhaustivo de diversas bases de datos académicas para conocer el contenido disponible, sus métricas y medios de acceso. Luego, se emprendió un estudio minucioso de la herramienta HERA 1.0, buscando identificar sus puntos fuertes y las áreas de mejora, con el fin de proponer y realizar modificaciones que pudieran extender su funcionamiento y mejorar su rendimiento y usabilidad. A partir del trabajo anterior se diseñó y desarrolló una segunda versión de HERA, considerando extensiones y mejoras como rediseño de la interfaz de usuario, integración de información de autores desde ORCID y de instituciones desde ROR, actualización del banco de bases de datos, refactorización del código actual, normalización de la gestión de las bases de datos de consulta, implementación de una API REST, búsqueda simultánea de recursos, exportación de la información, extensión para navegadores y manejo de errores y recursos no encontrados.

HERA 2.0 representa una versión extendida y más sofisticada que su predecesora, con mejoras en 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.

Índice

Introducción	6
1.1 Motivación	6
1.2 Objetivos y metodología.	8
1.3 Contribuciones	9
1.4 Organización del documento	9
Marco teórico	11
2.1 Métricas e Indicadores de Calidad e Impacto Bibliográfico	11
2.2 Bases de Datos Bibliográficas	20
2.2.1 Scopus	21
2.2.2 Web of Science	22
2.2.3 Directory of Open Access Journals (DOAJ)	24
2.2.4 REDIB	25
2.2.5 Crossref	27
2.2.6 SCimago Journal & Country Rank	28
2.2.7 Semantic Scholar	29
2.2.8 Servicios de métricas alternativas (Altmetric & Dimensions)	31
2.2.9 Google Scholar	34
2.2.10 Microsoft Academic	36
2.2.11 Resumen de bases de datos analizadas	37
2.3 HERA	39
2.3.1 Propósito	39
2.3.2 Diseño y desarrollo	39
2.3.3 Funcionamiento	39
2.3.4 Comparación con herramientas similares	44
Propuestas	45
3.1 Nueva Interfaz	45
3.2 Integración de información de autores e instituciones	45
3.3 Actualización del banco de bases de datos	46
3.4 Refactorización de código	46
3.5 Gestión de las fuentes de información	48
3.6 Implementación de API REST	48
3.7 Búsqueda simultánea de múltiples recursos	48
3.8 Exportación a formatos descargables	50
3.9 Extensión para navegadores	50
3.10 Manejo de errores y recursos no encontrados	51
3.11 Base de datos local	51
3.12 Resumen	52
Resultados Obtenidos	54
4.1 Resultados de: Nueva Interfaz	54

4.2 Resultados de: Integración de información de autores e instituciones	62
4.3 Resultados de: Actualización del banco de bases de datos	63
4.4 Resultados de: Refactorización de código	73
4.5 Resultados de: Gestión de las fuentes de información	84
4.6 Resultados de: Implementación de API REST	87
4.7 Resultados de: Búsqueda simultánea de múltiples recursos	89
4.8 Resultados de: Exportación a formatos descargables	93
4.9 Resultados de: Complemento web	96
4.10 Resultados de: Manejo de errores y recursos no encontrados.	100
4.11 Resultados en: Base de datos local	101
4.12 Resumen	103
Conclusiones y Trabajo Futuro	105
Conclusiones	105
Trabajos futuros	107
Referencias	109

Capítulo 1

Introducción

En esta tesina se presenta inicialmente la motivación para el desarrollo de HERA, la cual se divide en dos partes: la primera de ellas abarca una breve reseña histórica sobre la investigación científica y tecnológica, mientras que la segunda se centra en el análisis de las problemáticas relacionadas a la evaluación de calidad e impacto de la producción de HERA. A continuación, se describen los objetivos y la metodología utilizada, para posteriormente detallar las contribuciones realizadas en esta tesina. Por último, se proporciona una descripción de la organización del resto del documento.

1.1 Motivación

En el campo de la investigación científica y tecnológica, es usual para los investigadores leer, acceder y usar información extraída de los artículos publicados por sus pares. En las últimas dos décadas, la globalización y el avance tecnológico provocaron un aumento exponencial en el volumen de información disponible y en su difusión [1], influyendo principalmente en la forma en que estos investigadores acceden a dichos recursos¹ [2].

Antes de este cambio de paradigma, la exigencia de originalidad y el proceso de evaluación por pares eran elementos suficientes para certificar la calidad del material con el que se trabajaba. No obstante, la aparición de Internet y de sistemas automatizados cambió radicalmente la evaluación y publicación de documentos científicos. Para tomar dimensión se puede considerar un informe del Banco Mundial de datos [3], quien reveló que 2.5 millones de artículos fueron publicados en 2018, en aproximadamente 33100 revistas científicas.

El impacto de una revista y/o artículo es un factor fuertemente buscado, siendo habitualmente vinculado con la calidad del recurso. En efecto, es un error frecuente confundir calidad con impacto y considerarlos sinónimos. La calidad de una revista se relaciona con los procesos que sigue y los requisitos que impone para la publicación definitiva de un artículo. Por otro lado, el impacto está asociado a la incidencia que la revista o artículo tiene sobre el mundo académico, generalmente medido en número de citas [4].

El avance tecnológico y el consecuente incremento en la información disponible volvió necesario el desarrollo y extensión de sistemas de evaluación, indicadores y métricas bibliográficas tanto de revistas como artículos. El propósito de estas herramientas es poder determinar la calidad y el impacto de los diferentes recursos académicos. Dicho de otra manera, lo que se busca es dar respuestas a preguntas como: este artículo o revista, ¿ha seguido estándares aceptados en el área?, ¿puede considerarse de calidad?, ¿ha tenido un impacto alto en la comunidad? A pesar del surgimiento de estos esfuerzos, desafortunadamente aún

¹ Se entiende como recurso a cualquier publicación individual o seriada que resulte habitual en la ciencia.

no se puede responder en forma certera a los interrogantes mencionados, ya que muchos de estos valores tienen un marcado carácter subjetivo por no ser un estándar y por existir diferentes formas de calcularlos y valorarlos en los diferentes repositorios [5].

En síntesis, los investigadores enfrentan actualmente un verdadero desafío al momento de tener que determinar la calidad y el impacto de los recursos académicos, debido a la combinación de los factores mencionados. En primer lugar, el sostenido crecimiento de publicaciones científicas procedentes del desarrollo tecnológico. 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.

HERA surge como una respuesta ante estos desafíos, siendo una aplicación web que enriquece recursos académicos (artículos o revistas²) al integrar información de diferentes bases de datos académicas. 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 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 CrossRef, DOAJ, Scopus, Web of Science, REDIB, Dimensions, entre otras; y se encuentra disponible para la comunidad en forma abierta en <http://hera.sedici.unlp.edu.ar>

En su primera versión, HERA resultó funcional y de gran utilidad para los usuarios. Sin embargo, en esta tesina se propone mejorarla y extenderla a partir de un conjunto de modificaciones y nuevas funcionalidades con el objetivo de optimizar su rendimiento, escalabilidad, alcance y soporte. Entre ellas, se pueden mencionar:

- Nueva interfaz de usuario: se planifica una modificación completa de la interfaz de usuario. Este rediseño buscará ser más intuitivo y accesible, facilitando así la interpretación y navegación de los datos para el usuario.
- Integración de información de autores e instituciones: como parte esencial en la evaluación de recursos académicos, se planea la integración de datos de autores desde ORCID y de instituciones desde ROR.
- Refactorización del código actual: la refactorización siempre tiene el sencillo y claro propósito de mejorar el código. Con un código más óptimo, puede facilitarse la integración de nuevos elementos sin incurrir en errores nuevos. Se propone implementar métodos como la encapsulación, patrones de diseño, el reformateo o la extracción [6].
- Normalización de la forma en que HERA gestiona las bases de datos académicas: considerando que existe un alto dinamismo en la disponibilidad

² En este informe se utilizará el término “revista” para referirse no sólo a revistas científicas sino a cualquier publicación seriada que contenga artículos científicos

de bases de datos académicas, se modificará la aplicación web para obtener sus datos de una manera más eficiente. En particular, se propone estandarizar la manera en que HERA incorpora o retira bases de datos, y que a su vez resulte una forma práctica.

- Implementación de API REST: con la intención de ampliar la accesibilidad e interoperabilidad de HERA, se propone la implementación de una API REST. Esta API permitirá a otras aplicaciones interactuar con HERA de forma estandarizada y eficiente.
- Búsqueda de más de un recurso simultáneo: se incorporará una nueva característica que permitirá buscar más de un recurso académico de forma simultánea para evitar la búsqueda de un recurso por vez. Esta característica resulta útil en situaciones en las que se necesita evaluar la producción científica de un investigador, proyecto u organización, por mencionar algunos casos posibles.
- Simplificación del uso y aumento del alcance: se incorporará un complemento (extensión web³), el cual evitará que el usuario deba ingresar los identificadores de los recursos por su cuenta. Al mismo tiempo, permitirá aumentar el alcance de HERA, ya que el usuario final podrá utilizarlo sin necesidad de visitar el sitio web.
- Exportación de la información a otros formatos: se agregará la capacidad de exportar la información generada por la herramienta en archivos de tipo .CSV y .JSON para que el usuario pueda descargar y realizar un post-procesamiento de estos.
- Manejo de errores y recursos no encontrados: para asegurar una mayor robustez del sistema y una comunicación eficiente con los usuarios, se prevé mejorar el manejo de errores y situaciones en las que no se encuentran los recursos solicitados. Así, se podrá informar al usuario de manera clara y útil cuando un recurso no esté disponible o cuando se produzcan errores en el sistema.
- Incorporación de una base de datos local: para optimizar la gestión de los datos y mejorar el rendimiento, se considera la implementación de una base de datos local en HERA. Esta incorporación permitirá almacenar información relevante y reducirá la dependencia de las constantes solicitudes a bases de datos externas, agilizando así la obtención de la información requerida.

1.2 Objetivos y metodología.

El objetivo consiste en desarrollar una segunda versión de HERA, con el propósito de sofisticarla y extenderla a partir de un conjunto de modificaciones y nuevas funcionalidades, buscando mejorar su rendimiento, escalabilidad, alcance y soporte.

Para cumplir con el objetivo propuesto, se llevaron a cabo las siguientes actividades, de acuerdo con lo establecido en la propuesta:

³ Pequeño módulo de software para personalizar un navegador web.

- Se estudió el proceso de generación de conocimiento y sus medios de publicación.
- Se relevaron nuevas bases de datos académicas existentes y sus métricas bibliográficas.
- Se analizaron diferentes alternativas para poder extraer información de las diferentes bases de datos académicas existentes.
- Se diseñó y desarrolló una segunda versión de HERA, considerando extensiones y mejoras como rediseño de la interfaz de usuario, integración de información de autores desde ORCID y de instituciones desde ROR, actualización del banco de bases de datos, refactorización del código actual, normalización de la gestión de las bases de datos de consulta, implementación de una API REST, búsqueda simultánea de recursos, exportación de la información, extensión para navegadores y manejo de errores y recursos no encontrados.
- Se verificó que esta nueva versión alcance la funcionalidad especificada.

1.3 Contribuciones

- Un análisis de métricas, granularidad, información relacionada con la calidad, nivel de estandarización y otros parámetros proporcionados por diversos índices y bases de datos de publicaciones periódicas.
- La documentación de un conjunto de técnicas y herramientas que permiten recuperar, procesar e integrar métricas de revistas y artículos provenientes de múltiples fuentes.
- La implementación de la versión 2.0 de HERA, la cual contempla 9 modificaciones y extensiones diferentes orientadas a mejorar su rendimiento, escalabilidad, alcance y soporte. La versión Beta de HERA 2.0 se encuentra disponible en <https://testing.sedici.unlp.edu.ar>
- El desarrollo de un prototipo de extensión web para el navegador Chrome que permite a los usuarios obtener información sobre métricas e indicadores de diferentes publicaciones directamente desde el sitio que están navegando, sin necesidad de realizar el proceso de búsqueda en HERA. Si bien consiste de un prototipo, su desarrollo sirve a modo de prueba de concepto de la propuesta. El complemento se encuentra disponible en <https://chrome.google.com/webstore/detail/hera-browser-extension/cdmmeadgfiafkgpdicpepbmngbbfghcm/related?hl=es&authuser=0>

1.4 Organización del documento

A continuación, se describe la organización de los capítulos restantes del documento:

En el Capítulo 2, se proporciona un marco histórico de las tendencias en el uso de indicadores bibliométricos, abordando sus fortalezas, debilidades y distintas posturas. Se continúa con una descripción detallada de las bases de datos académicas más reconocidas, haciendo hincapié en las métricas disponibles y los métodos de acceso a las mismas. Finalmente, se introduce HERA 1.0, discutiendo su propósito, diseño, desarrollo y funcionamiento, y se compara con herramientas similares.

En el Capítulo 3, se abordarán propuestas concretas destinadas a mejorar la aplicación HERA, la cual se denominará HERA 2.0, con el objetivo primordial de extender y mejorar su funcionalidad, además de optimizar su rendimiento. Estas propuestas tienen como meta principal ofrecer una experiencia mejorada tanto para los usuarios finales como para los desarrolladores involucrados en el proyecto.

En el Capítulo 4, se discute la posible implementación de cada una de las mejoras propuestas en el Capítulo 3 y se describe el trabajo realizado. Luego, se presentan y analizan los resultados obtenidos a partir de las mejoras realizadas. Estos resultados muestran cómo las mejoras propuestas han impactado la funcionalidad, el rendimiento y la usabilidad de la aplicación, y cómo han contribuido a lograr los objetivos establecidos.

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

Capítulo 2

Marco teórico

En este capítulo se exponen las principales tendencias en tipos de indicadores utilizados para la evaluación de la producción científica. Posteriormente, se realiza un análisis de las bases de datos académicas utilizadas, señalando la información que aporta cada una en la valoración de la calidad y el impacto de un recurso académico. Además, se definen las métricas usuales encontradas en la web y los métodos disponibles para su recopilación. Por último, se describen las características funcionales y técnicas de la versión 1.0 de HERA.

2.1 Métricas e Indicadores de Calidad e Impacto Bibliográfico

Las métricas e indicadores de calidad e impacto bibliográfico son herramientas cuantitativas y cualitativas utilizadas para evaluar el valor y la relevancia de las publicaciones académicas. Estas métricas sirven para medir la calidad, el impacto y la influencia de una investigación en el ámbito académico y, en algunos casos, en la sociedad en general. Estas métricas pueden ser útiles para la toma de decisiones en la asignación de recursos, la promoción de académicos y la selección de proyectos de investigación, entre otros aspectos [7].

Las métricas e indicadores de calidad e impacto pueden clasificarse en tres categorías principales (se detallarán a lo largo de esta sección):

1. **Métricas basadas en citas:** métricas que evalúan la calidad y el impacto de las publicaciones académicas a través del análisis de las citas recibidas por un artículo, revista o autor. Ejemplos de métricas basadas en citas incluyen el factor de impacto, el índice h y *CiteScore*.
2. **Métricas alternativas (Altmetrics):** estas métricas evalúan el impacto de la investigación más allá de las citas académicas, considerando aspectos como menciones en redes sociales, blogs, descargas y otros tipos de interacción en línea. Las Altmetrics ofrecen una visión más amplia y diversa del impacto de la investigación.
3. **Indicadores cualitativos:** además de las métricas cuantitativas, también existen enfoques cualitativos para evaluar la calidad y el impacto de las publicaciones académicas, como la revisión por pares y la evaluación por expertos. Estos enfoques se centran en la calidad intrínseca del contenido de la investigación y su relevancia en el campo de estudio.

Las métricas de calidad e impacto bibliográfico juegan un papel crucial en la evaluación de la investigación académica debido a las siguientes razones:

1. **Comparabilidad:** permiten comparar de manera sistemática el rendimiento de diferentes investigadores, instituciones, revistas y artículos. Además, facilitan la identificación de trabajos, autores y grupos destacados en un campo específico, lo cual es útil para la asignación de recursos y la toma de decisiones. Sin embargo, el hecho de que haya un sistema no necesariamente garantiza una comparación objetiva.
2. **Responsabilidad:** pueden ayudar a las instituciones académicas y a los organismos financiadores a evaluar el rendimiento de sus investigadores y a asegurarse de que los recursos se utilicen de manera eficiente.
3. **Visibilidad:** pueden aumentar la visibilidad de la investigación y los investigadores, atrayendo la atención de otros académicos, financiadores y público en general. Esta visibilidad puede conducir a más colaboraciones y oportunidades de financiamiento.
4. **Incentivo:** pueden servir como incentivos para que los investigadores produzcan investigaciones de alta calidad y con un impacto significativo en su campo de estudio.

Sin embargo, es importante tener en cuenta que las métricas no deben ser el único criterio para evaluar la calidad e impacto de la investigación académica. Deben utilizarse en conjunto con indicadores cualitativos, como la revisión por pares y la evaluación por expertos, para obtener una comprensión más completa y equilibrada del valor y la relevancia de la investigación.

Además, es esencial reconocer las limitaciones y sesgos asociados con las métricas mencionadas anteriormente. Por ejemplo, las métricas basadas en citas pueden estar influenciadas por factores como la autopromoción, las citas obligatorias y las diferencias en las prácticas de citación entre disciplinas [8]. Por otro lado, las Altmetrics pueden verse afectadas por el "efecto de novedad" y la popularidad en línea, lo que no necesariamente refleja la calidad académica.

Métricas basadas en citas

Las métricas basadas en citas son aquellas que evalúan la calidad y el impacto de las publicaciones académicas a través del análisis de las citas recibidas por un artículo, revista o autor. Estas métricas ofrecen una visión cuantitativa del impacto de la investigación al medir la frecuencia con la que un trabajo académico ha sido citado por otros investigadores [9].

Entre las métricas basadas en citas más conocidas se encuentran el factor de impacto, el índice h y CiteScore.

Factor de impacto (FI):

El factor de impacto es una métrica que mide la frecuencia promedio con la que los artículos publicados en una revista específica han sido citados durante un periodo determinado [10]. Fue desarrollado por Garfield (2006) en la década de 1960 y es ampliamente utilizado para evaluar la calidad e impacto de las revistas académicas. El factor de impacto es calculado y publicado anualmente por *Clarivate Analytics*⁴ a través de su base de datos *Journal Citation Reports (JCR)*⁵.

Cálculo del factor de impacto:

El factor de impacto de una revista en un año específico se calcula dividiendo el número total de citas recibidas por los artículos publicados en esa revista durante los dos años anteriores, entre el número total de artículos publicados en la revista durante esos mismos dos años. [12]

La fórmula es la siguiente: (Citas en el año X a los artículos publicados en los años X-1 y X-2) / (Número total de artículos publicados en los años X-1 y X-2)

Ventajas:

1. **Comparabilidad:** permite comparar de una manera objetiva el rendimiento de diferentes revistas dentro de un campo específico, y facilita la identificación de las revistas más influyentes en su área de investigación.
2. **Reconocimiento:** las revistas con un factor de impacto alto son consideradas publicaciones de prestigio, lo que aumenta la visibilidad y reconocimiento de los investigadores que publican en ellas.
3. **Estandarización:** es una métrica ampliamente conocida y aceptada en la comunidad académica, lo que facilita su uso y comprensión.

Desventajas:

1. **Sesgo de campo:** puede variar significativamente entre diferentes campos de investigación debido a diferencias en las prácticas de citación y en el tamaño de la comunidad investigadora, lo que dificulta la comparación entre disciplinas.
2. **Sesgo de artículo:** no refleja la calidad o el impacto de un artículo individual, ya que se basa en el promedio de citas de todos los artículos publicados en una revista.
3. **Susceptibilidad a manipulación:** puede ser manipulado por prácticas editoriales, como la autopromoción y la citación obligatoria, lo que puede afectar su validez como indicador de calidad e impacto.

⁴ Clarivate Analytics es una empresa que posee y opera una colección de servicios enfocados principalmente al análisis de datos, incluyendo la investigación científica y académica de Google Analytics. Sitio web: <https://clarivate.com/>

⁵ Base de datos que proporciona factores de impacto y clasificaciones de muchas revistas de ciencias sociales y de la vida basadas en millones de citas. Sitio web <https://jcr.clarivate.com/jcr/home>

4. Sesgo Lingüístico: existe una clara desventaja en términos de factor de impacto para las revistas que publican en idiomas distintos al inglés. Según un estudio [13], los científicos que no son hablantes nativos de inglés tienden a publicar exclusivamente en este idioma, asumiendo que esto hará que sus artículos sean más visibles y citados. Su estudio reveló que los artículos publicados en inglés reciben un número mayor de citas que aquellos publicados en otros idiomas, cuando se controla estadísticamente el efecto de la revista, el año de publicación y la longitud del artículo. Este fenómeno puede deberse a que los artículos en inglés son accesibles a una audiencia más amplia, aunque también podrían existir otros factores.

Además, la representatividad del factor de impacto de ISI-Thomson y la relación existente entre los idiomas nacionales de los países y la difusión de las publicaciones científicas se ve afectada por el dominio del inglés. Según un estudio [14], la adopción del factor de impacto como criterio de valoración para las actividades científicas ha favorecido la consolidación de las revistas en inglés en la difusión del conocimiento científico. Los idiomas vernáculos sólo conservan parte de su importancia en ciertas disciplinas, como la Medicina Clínica o las Ciencias Sociales y Humanidades. El factor de impacto, inventado hace más de 50 años, podría ser una limitación para los autores y revistas científicas no anglófonos, y no considera algunas prácticas ampliamente utilizadas entre la comunidad científica en relación con el desarrollo de Internet como medio para la difusión del conocimiento.

Índice H

El índice h, desarrollado por Hirsch (2005), es una métrica que evalúa la productividad y el impacto de un investigador en función del número de citas que sus artículos han recibido. El índice h intenta combinar en una sola cifra la cantidad y calidad de la producción científica de un investigador, proporcionando una medida robusta de su desempeño académico [16].

Cálculo del índice h:

1. Ordenar los artículos del investigador por número de citas en orden descendente.
2. Identificar el mayor número (h) tal que el investigador tenga al menos h artículos con h o más citas cada uno.

Por ejemplo, si un investigador tiene un índice h de 10, esto significa que ha publicado al menos 10 artículos que han sido citados al menos 10 veces cada uno.

Utilizando el cálculo explicado anteriormente se ejemplifica el índice h:

En el ejemplo a continuación [17] (ver Figura 2.1), un autor tiene 8 trabajos que han sido citados 33, 30, 20, 15, 7, 6, 5 y 4 veces. Esto nos indica que el índice h del autor es 6.

<u>Articles</u>	<u>Citation numbers</u>
1	33
2	30
3	20
4	15
5	7
6	6 = h-index
7	5
8	4

Figura 2.1: Ejemplo cálculo del índice h: artículos y número de citaciones

Como se comentó anteriormente un índice h de 6 significa que este autor ha publicado al menos 6 trabajos que han recibido al menos 6 citas cada uno.

Explicación:

- El primer artículo ha sido citado 33 veces, y nos da un 1 (hay un artículo que ha sido citado al menos una vez).
- El segundo artículo ha sido citado 30 veces, y nos da un 2 (hay dos artículos que han sido citados al menos dos veces).
- El tercer artículo nos da un 3 y así sucesivamente hasta 6 con el sexto artículo más citado.
- Los dos últimos artículos no tienen efecto en este caso ya que han sido citados menos de seis veces [16].

Ventajas:

1. **Cantidad y calidad:** considera tanto el número de publicaciones como el número de citas, proporcionando una medida más equilibrada del impacto de un investigador.
2. **Robustez:** es menos susceptible a la influencia de artículos atípicos (con un número extremadamente alto o bajo de citas) que otras métricas basadas en citas.
3. **Comparabilidad:** como el factor de impacto permite comparar de manera objetiva el rendimiento de diferentes investigadores dentro de un mismo campo.

Desventajas:

1. **Sesgo de tiempo:** favorece a los investigadores con extensa trayectoria, ya que la acumulación de citas se consigue a lo largo del tiempo.

2. **Sesgo de campo:** las prácticas de citación varían entre disciplinas, lo que puede afectar el índice h y dificultar las comparaciones entre campos de investigación.
3. **Coautoría:** no distingue entre las contribuciones individuales en artículos coautorados, lo que puede resultar en una sobreestimación o subestimación del impacto de un investigador.

Si bien el índice h fue creado originalmente para medir la producción científica de investigadores, en la actualidad también es utilizado para revistas.

CiteScore

Desarrollado por Scopus (Elsevier)⁶, es una métrica que mide el impacto promedio de todos los artículos publicados en una revista durante un periodo de tres años. Es similar al factor de impacto, pero con algunas diferencias en el cálculo y en el periodo de análisis [18].

Cálculo de CiteScore:

Se calcula dividiendo el número total de citas recibidas por los artículos publicados en una revista durante un periodo de tres años, entre el número total de artículos publicados en la revista durante esos mismos tres años.

La fórmula es la siguiente:

$$\text{CiteScore} = (\text{Citas en el año } X \text{ a los artículos publicados en los años } X-1, X-2 \text{ y } X-3) / (\text{Número total de artículos publicados en los años } X-1, X-2 \text{ y } X-3)$$

Ventajas:

1. **Cobertura:** abarca un rango más amplio de revistas en comparación con el factor de impacto, ya que incluye todas las revistas indexadas en Scopus.
2. **Periodo de análisis más largo:** utiliza un periodo de análisis de tres años en lugar de dos, lo que puede proporcionar una visión más estable del impacto de una revista.
3. **Transparencia:** es de acceso gratuito y su metodología de cálculo está abierta al público, lo que facilita su uso y comprensión por parte de la comunidad académica.

⁶ Scopus es una base de datos bibliográfica de resúmenes y citas de artículos de revistas científicas propiedad de Elsevier lanzada en 2004. Sitio web: <https://www.elsevier.com/solutions/scopus>

Desventajas:

1. **Sesgo de campo:** al igual que el factor de impacto e índice H, CiteScore puede verse afectado por diferencias en las prácticas de citación entre disciplinas, lo que dificulta la comparación entre campos de investigación.
2. **No refleja el impacto individual de los artículos:** al igual que el factor de impacto, CiteScore se basa en el promedio de citas de todos los artículos publicados en una revista y no refleja necesariamente la calidad o el impacto de un artículo individual.

Métricas alternativas

Altmetrics

Las Altmetrics [19] (métricas alternativas) son un conjunto de indicadores que evalúan el impacto de la investigación académica a través de su presencia y alcance en entornos digitales y redes sociales. A diferencia de las métricas tradicionales basadas en citas, las Altmetrics consideran una amplia gama de interacciones en línea, como menciones en blogs, tweets, descargas, marcadores sociales y menciones en plataformas de noticias. Estas métricas ofrecen una perspectiva complementaria sobre el impacto y la relevancia de la investigación académica al considerar su alcance en la sociedad y su capacidad para generar conversaciones y debates [20].

Las altmetrics, aunque proporcionan una visión complementaria del impacto de la investigación, todavía carecen de estandarización y pueden variar significativamente entre diferentes plataformas y redes sociales, dificultando su interpretación y uso en la evaluación del impacto académico [21].

Métricas basadas en la web, como menciones en blogs, tweets y descargas:

Algunos ejemplos de Altmetrics incluyen:

- **Menciones en blogs:** el número de veces que un artículo ha sido mencionado en blogs académicos o de divulgación científica.
- **Tweets:** el número de veces que un artículo ha sido compartido o mencionado en Twitter.
- **Descargas:** el número de veces que un artículo ha sido descargado desde la plataforma de la editorial o un repositorio.
- **Marcadores sociales:** el número de veces que un artículo ha sido guardado en servicios de marcadores sociales, como Mendeley o CiteULike.
- **Menciones en plataformas de noticias:** el número de veces que un artículo ha sido mencionado en sitios web de noticias y medios de comunicación.

Las Altmetrics desempeñan un papel importante en la evaluación del impacto académico por varias razones:

Ventajas:

- **Visión más amplia del impacto:** capturan aspectos del impacto de la investigación que las métricas basadas en citas pueden pasar por alto, como la difusión de conocimientos fuera del ámbito académico y la participación pública en la investigación.
- **Velocidad:** pueden proporcionar información temprana sobre el impacto de un artículo, ya que las interacciones en línea suelen ocurrir mucho antes que las citas en publicaciones académicas.
- **Diversidad de fuentes:** consideran una amplia gama de fuentes y medios, lo que permite una evaluación más integral del alcance y la influencia de la investigación.
- **Complementariedad:** pueden complementar las métricas tradicionales basadas en citas al ofrecer información adicional sobre la calidad y el impacto de la investigación, lo que facilita una evaluación más equilibrada y completa.

Desventajas:

1. **Falta de estandarización:** a diferencia de las métricas basadas en citas, las Altmetrics aún carecen de un conjunto de estándares ampliamente aceptados y metodologías de cálculo, lo que puede generar confusión y dificultar su interpretación y uso en la evaluación del impacto académico.
2. **Variabilidad entre plataformas:** pueden variar significativamente entre diferentes plataformas y redes sociales, lo que dificulta la comparación y la agregación de métricas.
3. **Popularidad versus calidad:** un artículo muy mencionado o compartido en línea no necesariamente es sinónimo de alta calidad académica. Un tema de interés o polémico puede generar muchas interacciones en línea, pero esto no siempre refleja la calidad de la investigación en sí.

Indicadores cualitativos

Es un enfoque que considera aspectos no cuantificables de la calidad e impacto bibliográfico. A diferencia de las métricas cuantitativas, la evaluación cualitativa se basa en la experiencia, el conocimiento y el juicio de expertos para evaluar la calidad y el impacto de la investigación académica. Uno de los principales métodos de evaluación cualitativa en la academia es la revisión por pares.

Revisión por pares (Peer review)

La revisión por pares es un proceso mediante el cual los expertos en un campo de investigación evalúan críticamente el trabajo de otros investigadores antes de su publicación en revistas académicas, conferencias o libros. Este proceso tiene como objetivo asegurar que la investigación publicada sea rigurosa, original y relevante, y que cumpla con los estándares de calidad académica [22].

El proceso de revisión por pares puede variar según la disciplina y la revista, pero en general sigue los siguientes pasos:

1. Envío del manuscrito
2. Revisión preliminar por parte del editor
3. Asignación de revisores
4. Evaluación de los revisores
5. Decisión editorial
6. Revisión adicional y aceptación

Evaluación por expertos (Expert review)

Es un proceso en el que un especialista en un área de investigación específica evalúa críticamente el trabajo de otros investigadores, en términos de contenido, metodología, originalidad y relevancia. Aunque la revisión por pares también involucra la evaluación por parte de expertos, la evaluación por expertos puede extenderse más allá del proceso de revisión previo a la publicación e incluir evaluaciones de trabajos ya publicados, como libros, informes técnicos y capítulos de libros. La evaluación por expertos generalmente sigue los siguientes pasos:

1. Selección del experto.
2. Evaluación crítica.
3. Comentarios y sugerencias.
4. Aplicación de los comentarios.

La revisión por pares y la valoración por expertos son fundamentales en la evaluación de la calidad académica por las siguientes razones:

1. Garantiza la calidad y rigurosidad: asegura que la investigación publicada sea rigurosa, original y relevante para el campo, lo que contribuye a mantener altos estándares de calidad en la academia.

2. Mejora la investigación: los comentarios y sugerencias de los revisores pueden ayudar a los autores a mejorar y perfeccionar sus trabajos, lo que resulta en investigaciones más sólidas y de mayor calidad.
3. Confianza y credibilidad: otorga confianza y credibilidad a la investigación publicada, ya que indica que el trabajo ha sido evaluado y validado por expertos en el campo.
4. Tanto la revisión por pares como la evaluación por expertos puede ayudar a identificar y prevenir la publicación de investigaciones fraudulentas, plagio o prácticas poco éticas, lo que contribuye a mantener la integridad y la confiabilidad de la investigación académica.

A pesar de su importancia, la evaluación cualitativa tiene sus limitaciones y no es infalible. Puede estar sujeta a sesgos personales y subjetividad. Además, el proceso de revisión por pares puede ser lento y, a veces, no detecta todos los errores o problemas en un artículo. Aun así, sigue siendo un componente esencial en la evaluación cualitativa de la calidad e impacto bibliográfico y es ampliamente reconocido como un pilar en la academia.

2.2 Bases de Datos Bibliográficas

Existen cientos de banco de datos bibliográficos disponibles, y su prestigio y popularidad pueden variar según el área de estudio. Considerando el marco de este estudio y para garantizar una cobertura amplia y representativa de diferentes disciplinas y regiones, se decidió centrarse en las bases de datos multidisciplinarias y regionales más populares y ampliamente utilizadas. Estas fuentes de información ofrecen varias ventajas:

- **Cobertura amplia:** las fuentes de datos multidisciplinarias y regionales proporcionan acceso a una amplia gama de disciplinas y temas, lo que facilita la realización de investigaciones interdisciplinarias y comparativas.
- **Mayor visibilidad:** las bases de datos populares suelen ser consultadas con mayor frecuencia por investigadores, lo que puede aumentar la visibilidad y el impacto de los trabajos publicados en ellas.
- **Confiabilidad y calidad:** las bases de datos de mayor prestigio tienden a tener estándares de calidad más rigurosos y procesos de revisión más estrictos, lo que garantiza la calidad de los trabajos incluidos.
- **Facilita la comparabilidad:** utilizar banco de datos populares y ampliamente utilizados permite comparar y analizar de manera más efectiva los resultados de diferentes investigaciones, ya que es más probable que utilicen criterios y métricas similares.

Sin embargo, es importante tener en cuenta que centrarse sólo en las bases de datos más divulgadas también puede tener algunas desventajas, como pasar por alto investigaciones relevantes en bases de datos especializadas o menos conocidas.

2.2.1 Scopus

Scopus es una de las bases de datos bibliográficas más grandes y respetadas a nivel mundial. Es una base de datos multidisciplinaria y de citas desarrollada por Elsevier que abarca una amplia gama de disciplinas, incluyendo ciencias, tecnología, medicina, ciencias sociales y artes y humanidades. Scopus proporciona acceso a resúmenes, citas e información bibliográfica de más de 27.100 revistas académicas y miles de conferencias, libros y otras publicaciones académicas [23].

Scopus es conocida por su riguroso proceso de selección de revistas y criterios editoriales, lo que garantiza la calidad y confiabilidad de las publicaciones incluidas en la base de datos. Aunque no se especifican aquí, Scopus cuenta con varios criterios editoriales y de impacto que las revistas deben cumplir para ser incluidas en la base de datos. Estos criterios abarcan aspectos como la calidad del contenido, el proceso de revisión por pares, la diversidad del comité editorial y el cumplimiento de las normas éticas [24].

El uso de Scopus en la investigación académica ofrece varias ventajas, como la capacidad de realizar búsquedas amplias y específicas en diferentes disciplinas, la posibilidad de rastrear y analizar citas y métricas de impacto, y el acceso a información de calidad y confiable.

Scopus ofrece una gran cantidad de métricas y metadatos a través de sus diversas API, lo cual facilita el análisis y la comparación de la información en HERA. A pesar de que algunas API requieren una suscripción paga o están sujetas a restricciones de uso, la API de *Journal Metrics* resultó ser de particular interés para nuestros propósitos.

La API de *Journal Metrics* proporciona acceso a métricas clave, como **CiteScore**, **Source Normalized Impact per Paper (SNIP)** y **SCImago Journal Rank (SJR)**, mediante métodos *HTTP GET* siempre que se cuente con una *API Key*, la cual se puede obtener de forma gratuita al registrarse en el sitio web de Elsevier, aunque con una cuota de uso limitada.

CiteScore ([desarrollado en el capítulo 2.1](#)) mide el impacto de las citas y los documentos publicados en un período de cuatro años. Además, calcula un coeficiente llamado *CiteScoreTracker*, que permite estimar la posición futura de un recurso en el ranking en función de sus citas actuales (ver Figura 2.2). Estas métricas resultaron útiles para nuestros objetivos y se incorporaron en la herramienta para ser visualizadas y comparadas con las métricas de otros sitios donde el recurso también puede estar indexado y valorado de manera diferente.

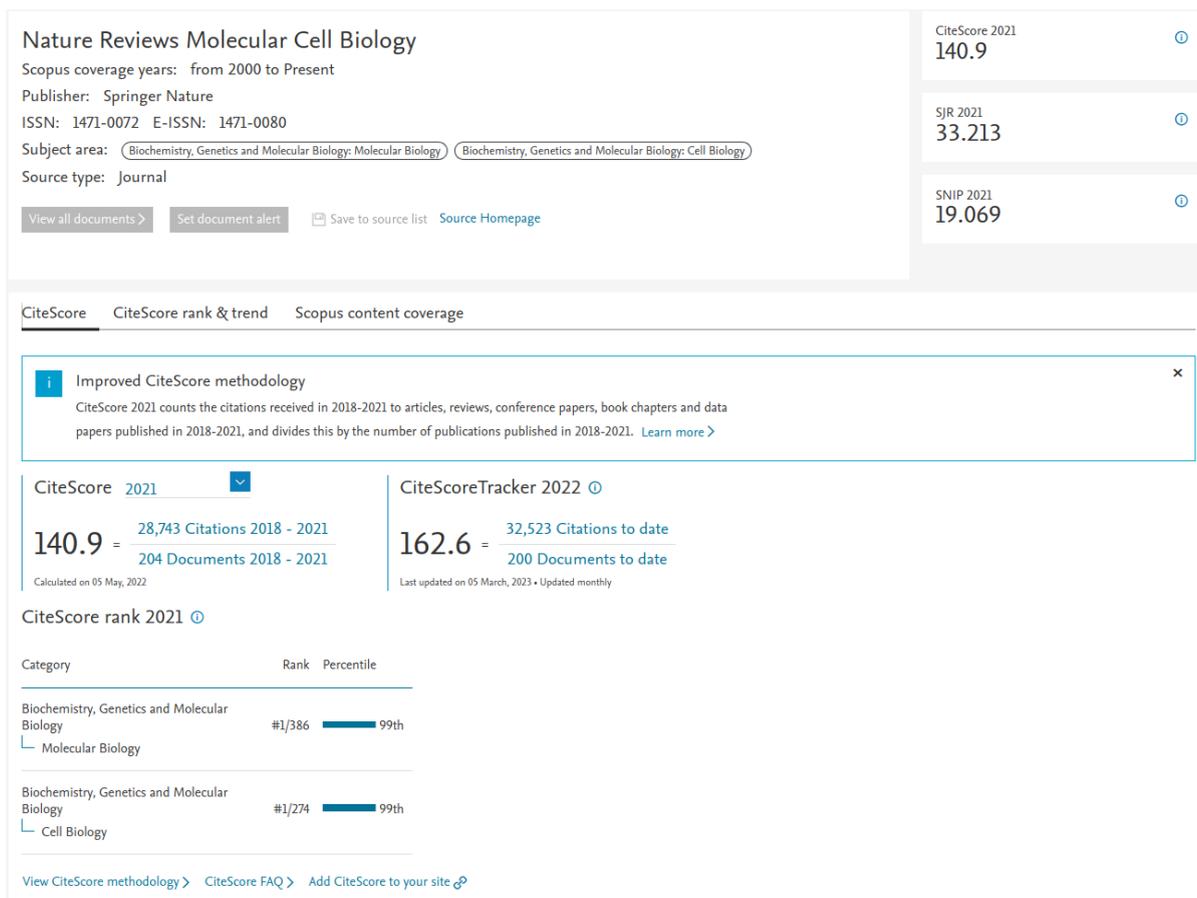


Figura 2.2: Captura de revista “Nature Reviews Molecular Cell Biology” en Scopus

2.2.2 Web of Science

Web of Science (WoS) es otra base de datos bibliográfica de renombre mundial y ampliamente utilizada en la investigación académica. Desarrollada por *Clarivate Analytics*, WoS es una base de datos multidisciplinaria que abarca diversas disciplinas, como ciencias, tecnología, medicina, ciencias sociales y artes y humanidades. La plataforma ofrece acceso a citas, resúmenes e información bibliográfica de miles de revistas académicas, conferencias, libros y otras publicaciones (ver Figura 2.3) [25].

The screenshot shows the Web of Science Master Journal List interface. At the top, there is a navigation bar with the Web of Science Group logo, 'Master Journal List', and links for 'Search Journals', 'Match Manuscript', 'Downloads', and 'Help Center'. On the right, there are 'Login' and 'Create Free Account' buttons. The main content area is titled 'Refine Your Search Results' and shows a search for 'Nature Reviews Molecular Cell Biology' with 7,121 results found. A filter sidebar on the left includes options like 'Web of Science Coverage', 'Open Access', 'Category', 'Country / Region', 'Language', 'Frequency', and 'Journal Citation Reports'. The search results section highlights an 'Exact Match Found' for 'NATURE REVIEWS MOLECULAR CELL BIOLOGY', providing details such as the publisher (NATURE PORTFOLIO), ISSN (1471-0072 / 1471-0080), and various Web of Science indices. There are also buttons for 'Share This Journal' and 'View profile page'.

Figura 2.3: Captura de la revista “Nature Reviews Molecular Cell Biology” en Web of Science.

WoS es conocida por sus rigurosos criterios de selección y estándares editoriales, lo que garantiza la calidad y confiabilidad de las publicaciones incluidas en la base de datos. Tiene varios criterios editoriales y de impacto que las revistas deben cumplir para ser consideradas en la base de datos. Estos criterios incluyen factores como la excelencia del contenido, la implementación del proceso de revisión por expertos, la variedad en la composición del comité editorial y la adhesión a las pautas éticas establecidas [26].

En WoS, se puede acceder a varias API a través del sitio web de Clarivate para obtener información sobre si un *journal* está indexado y sus métricas. Desafortunadamente, las API específicas de WoS no proporcionan una API key pública para su uso. No obstante, descubrimos que es de utilidad la API pública del sitio web *Master Journal List* (Web of Science Group), que permite recuperar un número limitado, pero adecuado, de métricas mediante solicitudes *HTTP POST* a la API.

Es importante mencionar que hay casos en los que un recurso puede ser obtenido a través de la API, pero al realizar una búsqueda manual del mismo en la página de *Master Journal List*, el recurso no aparece en los resultados. La API proporciona varias métricas, y algunas de las que consideramos más destacadas incluyen:

- Frecuencia de publicación
- Número de ediciones por año
- País de origen y año de inicio de publicación
- Idiomas del recurso
- Categorías (incluyendo a qué nivel de índices pertenece el recurso en WoS)

2.2.3 Directory of Open Access Journals (DOAJ)

Para abordar las limitaciones de Scopus y WoS [27], [28], se han creado varias bases de datos por consorcios académicos que están estrechamente relacionados con el movimiento de acceso abierto. Una de las bases de datos más destacadas a nivel mundial es DOAJ [29].

DOAJ fue fundado en 2003 y se estableció como una base de datos independiente que incluye publicaciones de acceso abierto de todas las áreas, países e idiomas. La base de datos es financiada por individuos y organizaciones que comparten la visión de DOAJ de proporcionar acceso abierto al contenido académico.

DOAJ es administrado por una Junta Asesora y un Consejo, ambos constituidos por académicos de numerosas universidades y partes del mundo. Para ser indexado en DOAJ, un *paper* o *journal* debe tener acceso sin restricciones y puede incluso recibir el Sello DOAJ (*DOAJ Seal* [30]) si cumple con requisitos más específicos que demuestran su calidad bajo los criterios del sitio. Aunque este sello no es necesario para ser indexado en la base de datos, destaca aquellos *journals* que ejercen las mejores prácticas de acceso abierto.

DOAJ ofrece diferentes métodos para extraer información de sus recursos. Uno de estos métodos es el acceso a grandes volúmenes de metadatos a través del protocolo OAI-PMH [31] (*Open Archives Initiative Protocol for Metadata Harvesting*). Además, DOAJ también proporciona una API [32] pública que permite extraer información de recursos de forma individual o grupal mediante búsquedas utilizando identificadores como DOI⁷ o ISSN⁸. Esto facilita la obtención de datos específicos sobre los recursos disponibles en la plataforma. Además desde su página web se pueden visualizar resultados de artículos científicos (Ver figura 2.4).

Al realizar consultas a la API de DOAJ, se pueden extraer numerosos datos de los recursos. Nos enfocamos en los siguientes aspectos:

- **Tipo de licencia del recurso:** esta información es útil para determinar los derechos de uso y distribución de un recurso académico en particular.
- **Costos de publicación del *journal*:** conocer los costos asociados con la publicación en un *journal* específico puede ser relevante para los autores que buscan opciones de publicación de acceso abierto.
- **Existencia de un *DOAJ Seal* asociado al recurso:** el *DOAJ Seal* es un distintivo que indica que el *journal* cumple con ciertos estándares de calidad y

⁷ El DOI es una cadena alfanumérica única para identificar contenido y proveer un vínculo persistente a su localización en la internet.

⁸ El ISSN es un código numérico único que identifica de manera internacional y unívoca las publicaciones seriadas, como revistas académicas, periódicos, anuarios, informes técnicos, boletines y otros recursos similares, tanto en formato impreso como electrónico.

buenas prácticas. La presencia de este sello puede ser un indicador de la calidad del recurso.

DOAJ SUPPORT APPLY SEARCH

Medical Laboratory Journal

2538-4449 (ONLINE)

Website ISSN Portal

About Articles

PUBLISHING WITH THIS JOURNAL

There are **NO PUBLICATION FEES** (article processing charges or APCs) to publish with this journal.

There is a [waiver policy](#) for these charges.

BEST PRACTICE

This journal began publishing in **open access** in 2007.

This journal uses a **CC BY-NC** license.

→ Look up their [open access statement](#) and their [license terms](#).

The author **retains unrestricted** copyrights and publishing rights.

→ Learn more about their [copyright policy](#).

JOURNAL METADATA

Publisher
[Golestan University of Medical Sciences](#),
Iran, Islamic Republic of

Manuscripts accepted in
English

LCC subjects
Medicine

Keywords
medical laboratory
clinical pathology parasitology
hematology microbiology
virology

Figura 2.4: Captura de revista “Medical Laboratory Journal” en Web of Science.

2.2.4 REDIB

Red Iberoamericana de Innovación y Conocimiento Científico es una plataforma de recursos científicos y académicos en acceso abierto que tiene como objetivo principal promover la difusión, visibilidad e impacto de la producción científica en idiomas español y portugués. REDIB [33] es el resultado de una colaboración entre el Consejo Superior de Investigaciones Científicas (CSIC [34]) de España y la Universidad de Tecnología de Monterrey en México.

Esta plataforma aglutina contenidos de revistas científicas iberoamericanas de diversas disciplinas, facilitando el acceso a la investigación y el conocimiento producidos en la región. REDIB también contribuye a mejorar la calidad y alcance de las revistas iberoamericanas mediante el seguimiento y la aplicación de criterios y estándares de calidad, así como el fomento de buenas prácticas editoriales.

REDIB ofrece una serie de herramientas y servicios, como búsquedas avanzadas, que permiten a los usuarios encontrar fácilmente artículos y publicaciones de interés (ver Figura 2.5). Además, la plataforma proporciona métricas e indicadores de

calidad e impacto bibliográfico, como el Índice REDIB, que ayuda a los usuarios a evaluar la relevancia y calidad de las revistas científicas incluidas en la base de datos (ver Figura 2.6).

Inicio / Ranking

Ranking de Revistas REDIB

Metodología del Ranking REDIB

Powered by Clarivate Analytics

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 2.5: Captura de Ranking de Revistas REDIB.

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

Documentos de esta revista (541)

Datos bibliográficos

ISSN:	1666-6038
Tipo de recurso:	Revista
Web de la revista:	http://journal.info.unlp.edu.ar/JCST
Entidad editora:	Universidad Nacional de La Plata
Idioma:	Inglés

Widget exportable

[<a href="https://redib.org/Record/"](https://redib.org/Record/)

Números publicados

- 2020

+ Volumen 20

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

Figura 2.6: Captura de revista "Journal of Computer Science & Technology"

No obstante, recientemente REDIB ha publicado un comunicado oficial en el que informa de una etapa de transición en la que podría haber interrupciones temporales

y retrasos en la actualización de contenidos. Según el comunicado, la Agrupación Índice Iberoamericano de Investigación y Conocimiento, A.I.E., responsable del desarrollo del proyecto REDIB, ha agotado la financiación prevista para el mismo. El CSIC, como socio fundador del proyecto, asumirá la responsabilidad de mantener y valorar los resultados obtenidos en el proyecto, y considerará su integración futura en nuevas iniciativas y proyectos de explotación. Aunque este cambio puede generar ciertas dificultades a corto plazo, representa una oportunidad para reconsiderar y redefinir la gestión y las estrategias de diseminación de la producción científica iberoamericana en el mundo digital.

2.2.5 Crossref

Crossref es una organización sin fines de lucro que se dedica a mejorar la colaboración y accesibilidad en la investigación académica y científica. Fundada en 2000 [35], su principal función es proporcionar servicios de infraestructura digital para la comunicación académica, incluyendo la asignación de identificadores únicos (DOI) para artículos de investigación, libros, capítulos de libros, informes técnicos y otros tipos de contenido académico.

Crossref ofrece una variedad de servicios y herramientas, como la API REST pública, que permite a investigadores y desarrolladores acceder a metadatos de millones de registros académicos. Estos servicios facilitan la búsqueda, descubrimiento, vinculación y evaluación de la literatura académica, así como el seguimiento de citas y la generación de métricas de impacto.

El impacto de Crossref en la investigación académica es significativo, ya que permite a los investigadores encontrar y acceder fácilmente a recursos relevantes, aumentando la visibilidad y el alcance de la producción científica. Además, Crossref promueve la colaboración entre investigadores, editores y otras partes interesadas al estandarizar formatos y procedimientos, lo que contribuye a una comunicación académica más eficiente y transparente.

Dado que solo se suben metadatos en Crossref y el objetivo principal del investigador es facilitar el descubrimiento de su contenido, es responsabilidad del investigador cargar y mantener actualizados estos metadatos. Crossref ofrece una amplia documentación [36] con guías y recomendaciones para realizar esta tarea de manera efectiva. Además, es necesario mantener un formato adecuado para mostrar el DOI del recurso según las pautas de Crossref [37].

En cuanto a los artículos, entre los metadatos encontramos métricas como la cantidad de citas que recibe el artículo, así como información útil como el título, el resumen, la lista de autores, el ISSN del *journal* al que pertenece el artículo y la URL original del mismo. En el caso de los *journals*, podemos encontrar información sobre la cantidad de artículos publicados y cuántos se han incorporado cada año. También se incluyen metadatos de cobertura, como el porcentaje de autores que publican en

el *journal* y que tienen un ORCID, lo cual es importante para monitorear el impacto académico de un investigador [38].

En la Figura 2.7 se visualiza el resultado de realizar una búsqueda de un artículo en base a un DOI en Crossref. La información que presenta es limitada debido a que el objetivo principal de Crossref es el almacenamiento de metadatos y no está diseñado específicamente para el consumo directo de los usuarios finales.



Figura 2.7: Captura de búsqueda de artículo en Crossref.

2.2.6 SCImago Journal & Country Rank

SCImago Journal & Country Rank (SCImago) [39] es una plataforma en línea que desarrolla métricas e indicadores de calidad para revistas académicas y países, utilizando la información contenida en el servicio Scopus. El enfoque principal del grupo SCImago [40] es el análisis de información y su representación mediante técnicas de visualización.

En cuanto al ranking de revistas, se puede observar la siguiente información en sus métricas (ver Figura 2.8):

- Categoría del recurso
- Cuartil
- Índice H
- Total de artículos del año anterior y de los últimos 3 años
- Cantidad de referencias
- Citas totales en los últimos 3 años
- Métricas de citas por artículo y referencias por artículo

Además, es posible acceder individualmente a una revista y visualizar información bibliográfica adicional, así como una breve descripción y gráficos que ilustran la evolución de la producción científica del *journal* a lo largo del tiempo.

Title	Type	↓ SJR	H index	Total Docs. (2021)	Total Docs. (3years)	Total Refs. (2021)	Total Cites (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc. (2021)
1 Ca-A Cancer Journal for Clinicians	journal	56.204 Q1	182	41	121	4006	17959	78	186.75	97.71
2 Nature Reviews Molecular Cell Biology	journal	33.213 Q1	452	111	338	9025	13797	161	38.55	81.31
3 Quarterly Journal of Economics	journal	31.348 Q1	272	48	111	3406	2241	110	16.30	70.96
4 Cell	journal	25.716 Q1	814	517	1727	33658	73240	1639	45.00	65.10
5 MMWR Recommendations and Reports	journal	25.045 Q1	148	124	17	2900	663	17	33.79	23.39
6 New England Journal of Medicine	journal	24.907 Q1	1079	1453	4498	14767	143343	1891	35.41	10.16

Figura 2.8: Captura de Ranking de SJR.

2.2.7 Semantic Scholar

Semantic Scholar es un motor de búsqueda académico gratuito y en línea desarrollado por la organización sin fines de lucro Allen Institute for Artificial Intelligence (AI2)[41]. Fue lanzado en 2015 [42] y su objetivo principal es ayudar a los investigadores a encontrar información relevante y de calidad en el ámbito científico y académico, utilizando técnicas de inteligencia artificial y procesamiento del lenguaje natural.

Semantic Scholar indexa millones de documentos académicos de diversas disciplinas, incluyendo ciencias de la computación, biología, medicina, física, matemáticas, entre otras. Su enfoque en el aprendizaje automático y la inteligencia artificial le permite ofrecer resultados de búsqueda más precisos y contextualizados, en comparación con otros motores de búsqueda académicos tradicionales.

Algunas de las características distintivas de Semantic Scholar incluyen:

1. **Filtros avanzados de búsqueda:** permiten a los usuarios refinar sus búsquedas por autor, año de publicación, tipo de documento, área temática, entre otros.
2. **Extracción de información clave:** Semantic Scholar utiliza algoritmos de aprendizaje automático para extraer datos relevantes de los documentos, como citas, tablas, figuras, resultados experimentales y enlaces a conjuntos de datos.

3. **Resúmenes mejorados:** ofrece resúmenes automáticos y enriquecidos que ayudan a los investigadores a comprender rápidamente el contenido y la relevancia de un documento.
4. **Detección de relaciones entre documentos:** Semantic Scholar identifica y destaca las relaciones entre documentos, lo que facilita a los usuarios encontrar trabajos relacionados y comprender el contexto en el que se encuentran.
5. **Métricas de impacto e influencia:** presenta métricas de impacto, como el número de citas y el índice h, para ayudar a evaluar la importancia de un documento en su campo de estudio.

Semantic Scholar también ofrece una API REST [43] que permite a los desarrolladores acceder a la información indexada en la plataforma y utilizarla para crear aplicaciones y herramientas de investigación personalizadas, que permite extraer las siguientes métricas:

- Número de citas
- Número de citas en los últimos 3 años
- Número de citas influyentes
- Información bibliográfica (título, resumen, autores, referencias)
- URL del recurso en Semantic Scholar

Al explorar el sitio web de Semantic Scholar y acceder a la página específica de un artículo, tal como se ilustra en la Figura 2.9, se puede apreciar una detallada información bibliográfica del mismo, además de visualizar las citas que recibe, las referencias que realiza, y los enlaces a los recursos correspondientes. Semantic Scholar ofrece la posibilidad de aplicar filtros a las citas, facilitando la identificación de aquellas que usan el artículo como base teórica, que citan sus métodos, o que aluden a sus resultados. Dentro de las citas, algunas son catalogadas como “altamente influyentes”, identificadas por el algoritmo de Semantic Scholar como aquellas que poseen un significativo impacto en el artículo que las cita, esto puede ser debido a que abordan una temática y área muy similar o porque fundamentan sus contenidos en las mismas referencias.

Figura 2.9: Captura del artículo “The Nature of Statistical Learning Theory” en Semantic Scholar.

2.2.8 Servicios de métricas alternativas (Altmetric & Dimensions)

Altmetric es un servicio de métricas alternativas que ha adquirido una importancia creciente en la evaluación del impacto y la relevancia de la investigación científica en la actualidad. Esta herramienta permite medir la atención que reciben los artículos de investigación en los medios sociales y otros canales en línea, proporcionando así una manera más amplia y compleja de evaluar la influencia de la investigación en la comunidad científica y en la sociedad en general.

Principales métricas de un documento académico:

1. Puntaje de atención basado en menciones con peso dependiendo el sitio.
2. Presencia de menciones en redes sociales.

El *badge* de Altmetric muestra un "puntaje de atención", que es una acumulación de menciones identificadas en diversas redes sociales, periódicos, blogs, etc., tal como se puede apreciar en la Figura 2.10. Esta acumulación no es una simple suma, sino que cada valor obtenido se pondera según el lugar donde se haya mencionado el recurso. No es lo mismo que un recurso sea citado en una red social de uso común, como Twitter o Facebook, que en un blog científico o un sitio de noticias reconocido. Además, el algoritmo encargado de calcular la suma intenta reconocer publicaciones duplicadas con la finalidad de restringir la manipulación de los valores.



Figura 2.10: Captura de Altmetric Badge.

Tal como se ilustra en la Figura 2.11, al ingresar a la página de detalles de las métricas también se puede observar información geográfica y demográfica en relación con el origen de las citas.

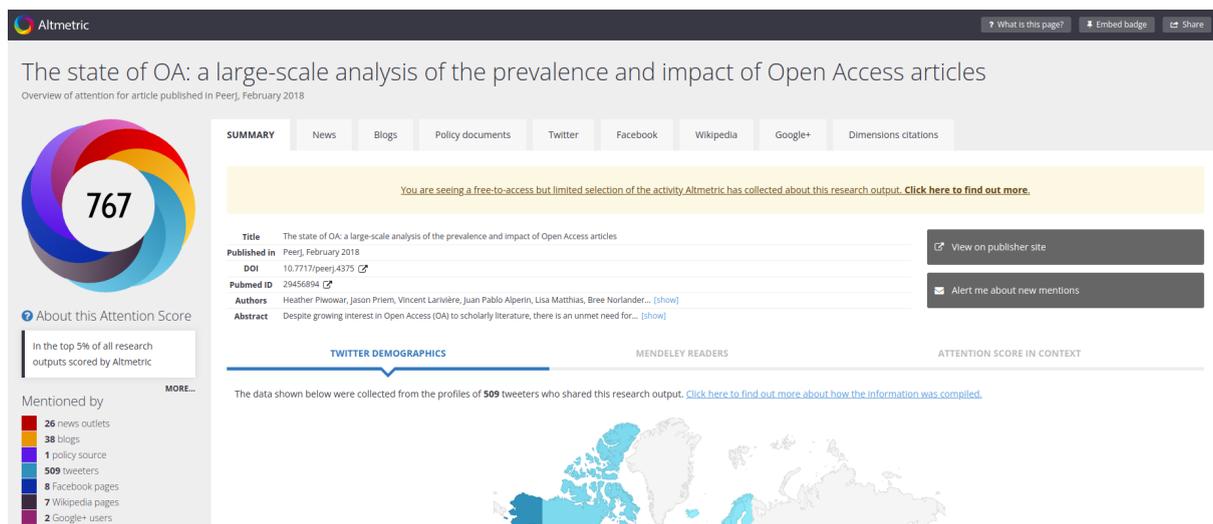


Figura 2.11: Captura del sitio web Altmetric.

Por su parte, **Dimensions** (desarrollado por la misma compañía que Altmetric, Digital Science)[44], es otro servicio de métricas de investigación que ha adquirido importancia en los últimos años. A diferencia de Altmetric, Dimensions ofrece una amplia gama de datos y herramientas para evaluar el impacto y la relevancia de la investigación en varios aspectos, como las citas de las publicaciones, los datos y las métricas de las colaboraciones, los financiamientos, las patentes, las políticas y las menciones en medios sociales.

La Dimensions Badge [45] nos presenta esta información de manera gráfica y fácilmente importable a cualquier página web⁹. En dicha *badge* podemos ver

⁹ Dimensions Badge Builder - <https://badge.dimensions.ai/>

información de número de citas a las publicaciones, y además realiza cálculos sobre las mismas para determinar otras métricas como lo ilustra la Figura 2.12:

1. Citas Recientes, en los últimos 2 años
2. 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
3. Proporción de Citas Relativa, que indica el desempeño relativo de un *paper* comparado con *papers* del mismo área de estudio

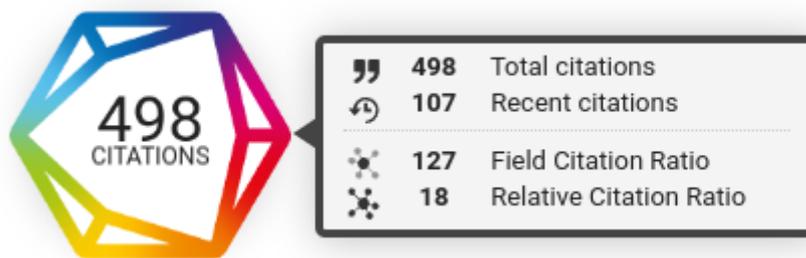


Figura 2.12: Captura de Dimensions Badge.

Al ingresar al *badge*, se ingresa al sitio web de Dimensions (Figura 2.13), donde se proporciona información más detallada sobre la obra o recurso específico. La plataforma dispone de datos como, por ejemplo, qué proporción de las citas totales son recientes. También se brinda información sobre las áreas de estudio en las que se cita con mayor frecuencia el *paper* en consideración.

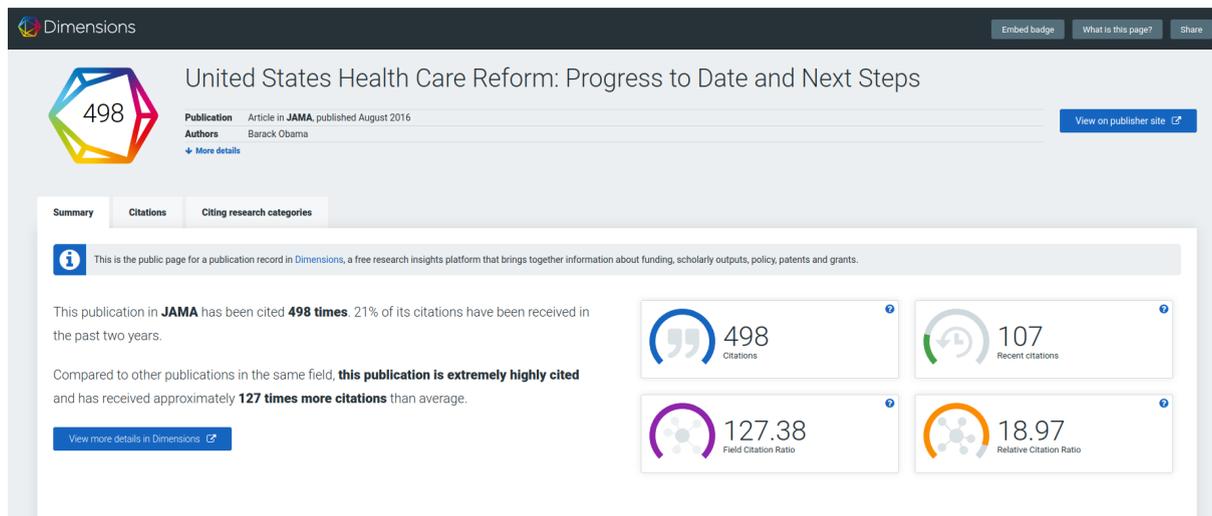


Figura 2.13: Captura del sitio web Dimensions.

La importancia tanto de Altmetric como de Dimensions radica en que las métricas tradicionales de investigación, como el factor de impacto de una revista, no siempre reflejan la verdadera influencia o relevancia de un artículo o investigación. Actualmente, hay un debate en curso sobre el papel de las altmetrics en la

evaluación de la ciencia y el uso de las redes sociales para difundir la investigación y generar un impacto más amplio.

El estudio de Ebrahimi, Setareh, y HosseinChari [46] encontraron correlaciones entre las altmetrics y las medidas tradicionales de impacto, sugiriendo que las altmetrics pueden ofrecer una visión complementaria del impacto de la investigación. Sin embargo, hay opiniones divididas sobre este tema. Mientras que Lockwood [47] respalda el uso de altmetrics como un complemento valioso para las métricas tradicionales, afirmando que los artículos con títulos que captan la atención reciben más atención en línea, Ortega [44] crítica su validez y confiabilidad, argumentando que las métricas de Publons presentan varios problemas como indicadores de evaluación de la investigación. Mientras tanto, et al. [49] proporcionan un punto de vista más neutral, demostrando que las altmetrics y las métricas bibliográficas proporcionan perspectivas importantes pero diferentes sobre el impacto de los artículos.

Además, Luc et al [46] demostraron que el uso de Twitter puede mejorar significativamente las citas de artículos, lo que sugiere que las redes sociales pueden desempeñar un papel importante en la difusión y el impacto de la investigación académica.

Por lo tanto, es esencial tener en cuenta este debate al considerar el uso de altmetrics en la evaluación de la investigación. Aunque las altmetrics pueden proporcionar información adicional valiosa, también es crucial entender sus limitaciones y aplicarlas de manera crítica y reflexiva en el contexto de la evaluación académica.

Los investigadores pueden utilizar estas herramientas para evaluar el impacto y la relevancia de la investigación, los investigadores y editores pueden tener en cuenta un conjunto más amplio de factores, como la difusión de la investigación en las redes sociales, blogs, noticias y otros canales en línea, y evaluar la influencia y el impacto de la investigación en un contexto más amplio.

Las herramientas y datos proporcionados por Dimensions y las puntuaciones de Altmetric se basan en un algoritmo que recopila y analiza los datos de diferentes fuentes en línea, y asigna una puntuación a cada artículo según el nivel de atención que recibe. Estos datos pueden ser útiles para los investigadores y editores a la hora de evaluar la calidad y la relevancia de la investigación, identificar nuevas oportunidades de colaboración, financiamiento y difusión de investigación, y promover un mayor impacto de la investigación en la sociedad.

2.2.9 Google Scholar

La evolución de la tecnología ha permitido la indexación automática de contenido en Internet por parte de motores de búsqueda como Google y Bing. De esta manera, emergen bases de datos como Google Scholar (GS) y Microsoft Academic, que

reúnen información que sus algoritmos de búsqueda identifican como contenido académico. Resulta destacable que estos sistemas son capaces de encontrar prácticamente cualquier documento que se ajuste a su formato. Sus algoritmos de machine learning son extremadamente potentes, capaces incluso de indexar videoconferencias grabadas, que no son artículos en sí mismos. Un buen número de investigadores prefiere estas herramientas para presentar sus descubrimientos [51].

Estas bases de datos han revolucionado la forma en que los académicos acceden al contenido, a diferencia de bases de datos más selectivas y específicas como Scopus y WoS. Según un estudio comparativo [8], GS no solo tiene una mayor cobertura, sino que también encuentra una mayor cantidad de citas en comparación con sus contrapartes. Sin embargo, se aconseja usar GS con precaución, ya que las citas únicas que se encuentran en ella suelen tener un impacto promedio menor que las encontradas en WoS y Scopus.

Además dicho análisis que incluyó a GS, Scopus y WoS en múltiples disciplinas encontró que GS no solo posee una cobertura más extensa, sino que también localiza un mayor número de citas en comparación con sus competidores. Este estudio también mostró que la proporción de citas presentes en Scopus y WoS pero ausentes en GS es relativamente pequeña. Además, GS resulta ser considerablemente más efectivo en la búsqueda de materiales como tesis, disertaciones, conferencias y libros, y se confirma que GS, en general, rastrea un volumen mayor de citas sin importar el campo de estudio o el tipo de recurso. No obstante, es importante resaltar que aproximadamente la mitad de las citas que solo GS localiza son de este tipo de recursos, incluyendo materiales no publicados como preprints, informes técnicos y otros documentos de formato similar a un artículo. Se recomienda prudencia al usar GS, ya que estas citas exclusivas de GS suelen tener un impacto promedio menor en comparación con las que se encuentran en WoS y Scopus.

Respecto a su uso, aunque GS se presenta como una herramienta precisa para la búsqueda de recursos, no proporciona una cantidad suficiente de metadatos útiles para nuestro propósito. Solo ofrece datos de citas a nivel de artículo e información de autores. El número de citas que se muestra al realizar una búsqueda no está presente cuando el recurso no tiene ninguna, y las demás métricas que ofrece GS 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 (ver Figura 2.14).

The screenshot shows the Google Scholar interface. At the top, the search bar contains the query "The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles". Below the search bar, the results are displayed in a list. The first result is highlighted in blue and includes the title, authors (JP Alperin, L Matthias, B Norlander, A Farley), the year (2018), and the journal (PeerJ). The abstract snippet mentions "the citation impact of OA articles, corroborating the so-called open-access citation advantage: accounting for age and ...". To the right of the search bar, there are filters for "Artículos" and "14 resultados (0,08 s)". On the left side, there are various filters such as "Cualquier momento" (with sub-options for "Desde 2023", "Desde 2022", "Desde 2019", and "Intervalo específico..."), "Ordenar por relevancia", "Ordenar por fecha", "Cualquier idioma" (with "español" selected), "Cualquier tipo" (with "Artículos de revisión" selected), and checkboxes for "Incluir patentes", "Incluir citas", and "Crear alerta".

Figura 2.14: Captura del sitio web Google Scholar.

2.2.10 Microsoft Academic

En el pasado, Microsoft Academic (MA)¹⁰ fue una alternativa viable a Google Scholar. Este motor de búsqueda académica, que aprovechaba la potencia de Bing y de los algoritmos de aprendizaje automático, así como técnicas de análisis del lenguaje natural y búsqueda semántica, no sólo era capaz de identificar recursos académicos pertinentes, sino que también se convirtió en una potente herramienta de búsqueda que asistía a los investigadores en la laboriosa tarea de recopilar información.

Para que una publicación fuera detectada e indexada en MA, primero tenía que ser indexable por Bing. Además, se recomendaba que siguiera los estándares web de etiquetas HTML para contenido académico, que eran los mismos que usaba Google Scholar, para facilitar su descubrimiento.

MA también contaba con un sistema de clasificación de autores que reflejaba aspectos relacionados con la publicación continua y el impacto futuro de sus trabajos. Un autor que aparecía más arriba en la clasificación no solo era más citado, sino que también lo era en su campo de estudio específico. Estos campos eran en su mayoría generados por la herramienta de aprendizaje automático de Microsoft.

¹⁰ [En una entrada de blog, se anunció que el sitio web de Microsoft Academic y las API subyacentes fueron retiradas el 31 de diciembre de 2021.](#)

En términos de datos relevantes para la evaluación de la calidad del contenido académico, MA proporcionaba información sobre el número de citas y las áreas o campos de estudio relacionados con el recurso. Además, ofrecía predicciones del número de citas, indicadas con un asterisco si se trataba de un valor estimado (Ver Figura 2.15).

The screenshot shows a Microsoft Academic page for the paper "The Google scholar experiment: How to index false papers and manipulate bibliometric indicators". The page includes the following information:

- Title:** The Google scholar experiment: How to index false papers and manipulate bibliometric indicators
- Publication Info:** 2014 Journal of the Association for Information Science and Technology | Volume: 65, Issue: 3, pp 446-454 | DOI: 10.1002/ASL.23056
- Authors:** Emilio Delgado López-Cózar ¹, Nicolás Robinson-García ¹, Daniel Torres-Salinas ²
- Affiliations:** ¹ University of Granada, ² University of Navarra
- Metrics:** 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
- Other Versions:** The Google Scholar Experiment: how to index false papers and manipulate bibliometric indicators 2013 arXiv: Digital Libraries | DOI: 10.1002/ASL.23056
- Related Topics:** Citation analysis, Bibliometrics, Information science, View More (5+)

Figura 2.15: Captura del sitio web antiguo de Microsoft Academic en funcionamiento.

2.2.11 Resumen de bases de datos analizadas

La siguiente tabla muestra las bases de datos consultadas por HERA 1.0 y las métricas correspondientes que se obtienen en cada caso:

Base de Datos	Método de selección de contenido	Principales métricas e indicadores	Recurso
Crossref	Metadatos estandarizados por miembros de la plataforma.	<ul style="list-style-type: none"> • Número de citas • Estadísticas de papers por año y autores con ORCID • Información bibliográfica 	Artículo; Revista
Scimago Journal & Country Rank	Recursos contenidos en Scopus	<ul style="list-style-type: none"> • Posición en ranking global • Cuartil • H-index del <i>journal</i> • Información bibliográfica 	Revista
Scopus	Proceso de evaluación por junta asesora	<ul style="list-style-type: none"> • Número de citas • Predicción de citas • Factores de impacto 	Revista
Web of Science	Proceso de evaluación por editores expertos	<ul style="list-style-type: none"> • Índice de WoS • Información bibliográfica 	Revista
DOAJ	Proceso de evaluación de acceso abierto y calidad	<ul style="list-style-type: none"> • Licencias • Sello de DOAJ 	Artículo; Revista
Semantic Scholar	Indexación mediante algoritmos de machine learning ejecutados sobre bases de datos afiliadas	<ul style="list-style-type: none"> • Número de citas • Frecuencia de citas • Número de citas influyentes 	Artículo
Altmetric	Realiza cálculos reconociendo menciones en redes sociales, foros, etc. Mediante inteligencia artificial	<ul style="list-style-type: none"> • Puntuación de atención basada en menciones ponderadas según el sitio • Presencia de menciones en redes sociales 	Artículo
Google Scholar	Indexación por Google mediante algoritmos de machine learning	<ul style="list-style-type: none"> • Citas 	Artículo; Revista
Dimensiones	Recopilación de contenidos de bases abiertas (Crossref y PubMed)	<ul style="list-style-type: none"> • Número de citas • Proporción de citas por campo y relativa 	Artículo
Microsoft Academic	Indexación por Bing mediante algoritmos de machine learning	<ul style="list-style-type: none"> • Número de citas • Predicción de citas • Información bibliográfica 	Artículo
Red Iberoamericana	Proceso de evaluación de calidad	<ul style="list-style-type: none"> • Posición de ranking • Información bibliográfica • Indicadores de calidad editorial y acreditaciones 	Revista

Tabla 1: Resumen de bases de datos consultadas.

2.3 HERA

2.3.1 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, esta herramienta proporciona una solución integral para simplificar, acelerar y apoyar el proceso de evaluación de recursos académicos, permitiendo a los usuarios centrarse en el análisis y la interpretación de los datos, ahorrando tiempo y esfuerzo en la búsqueda de información.

2.3.2 Diseño y desarrollo

A continuación, se procede a describir el diseño y desarrollo de la primera versión de HERA, denominada en adelante como HERA 1.0 [52]. En la implementación de HERA 1.0, se conformó en una aplicación web organizada en dos desarrollos: una aplicación front-end desarrollada con ReactJS y una aplicación back-end desarrollada con NodeJS. El propósito de la aplicación back-end es actuar como intermediaria entre la aplicación front-end y los diferentes sistemas que suministran datos y métricas de recursos académicos, asumiendo así la mayor parte de la carga de trabajo.

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

2.3.3 Funcionamiento

HERA 1.0 es una herramienta web accesible en <http://hera.sedici.unlp.edu.ar>. Su uso 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 2.16).

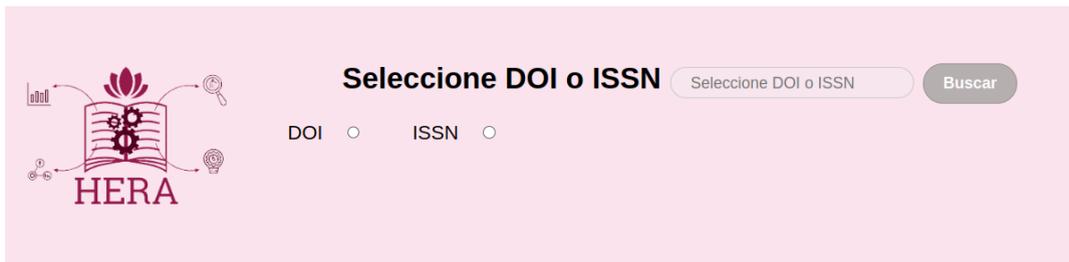


Figura 2.16: Interfaz de búsqueda de la herramienta con input de búsqueda y selectores de tipo de identificador (DOI e ISSN)

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.

Al comenzar la búsqueda, 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.

Búsqueda de un *journal*

Al ingresar un ISSN en la barra de búsqueda, se efectúa la búsqueda de un *journal* o publicación periódica. Primero se obtiene información acerca de la editorial y el enlace al recurso, así como una vista resumida de las métricas recopiladas de diferentes sitios, como se muestra en la Figura 2.17. Desplazándose hacia abajo, se puede ver un resumen de las métricas del *journal* recopiladas en las bases de datos.



Figura 2.17: Resultados generales de la búsqueda de un *journal*

Existe un botón (Ver más) para ampliar la información y visualizar todo su contenido, como se muestran en las Figuras 2.18, 2.19 y 2.20:



Figura 2.18: Resultados en detalle de Crossref y DOAJ

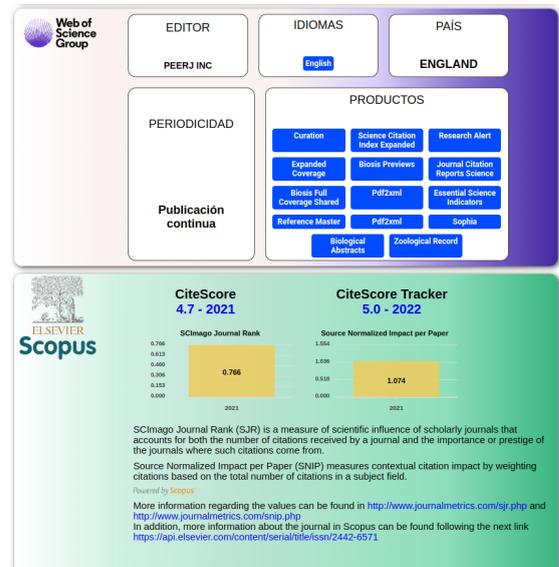


Figura 2.19: Resultados en detalle de WoS y Scopus

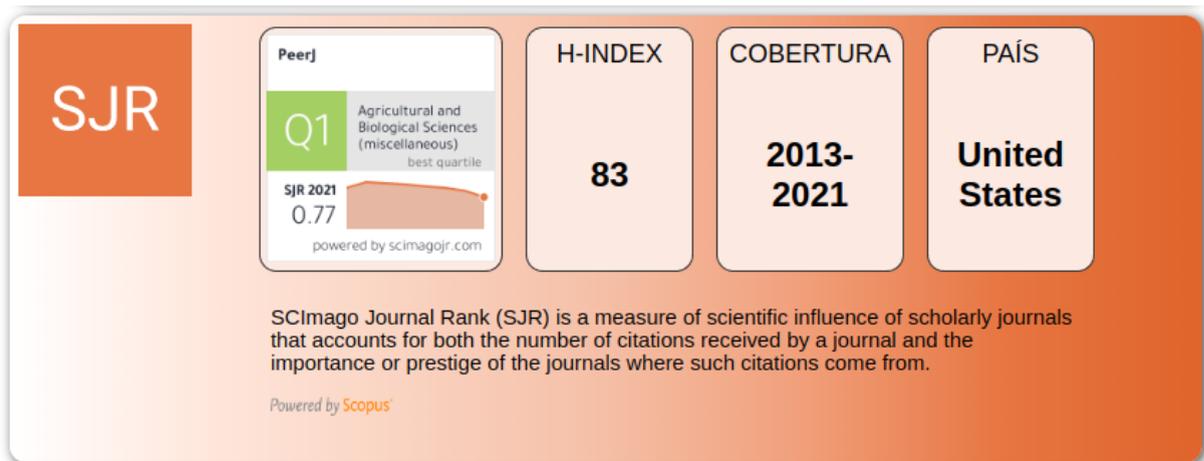


Figura 2.20: Resultados en detalle de Scimago

Búsqueda de un artículo

Al introducir un DOI en la barra de búsqueda y pulsar el botón buscar, las métricas se recopilan de diferentes bases de datos y se muestran en pantalla, como se puede observar en la Figura 2.21.

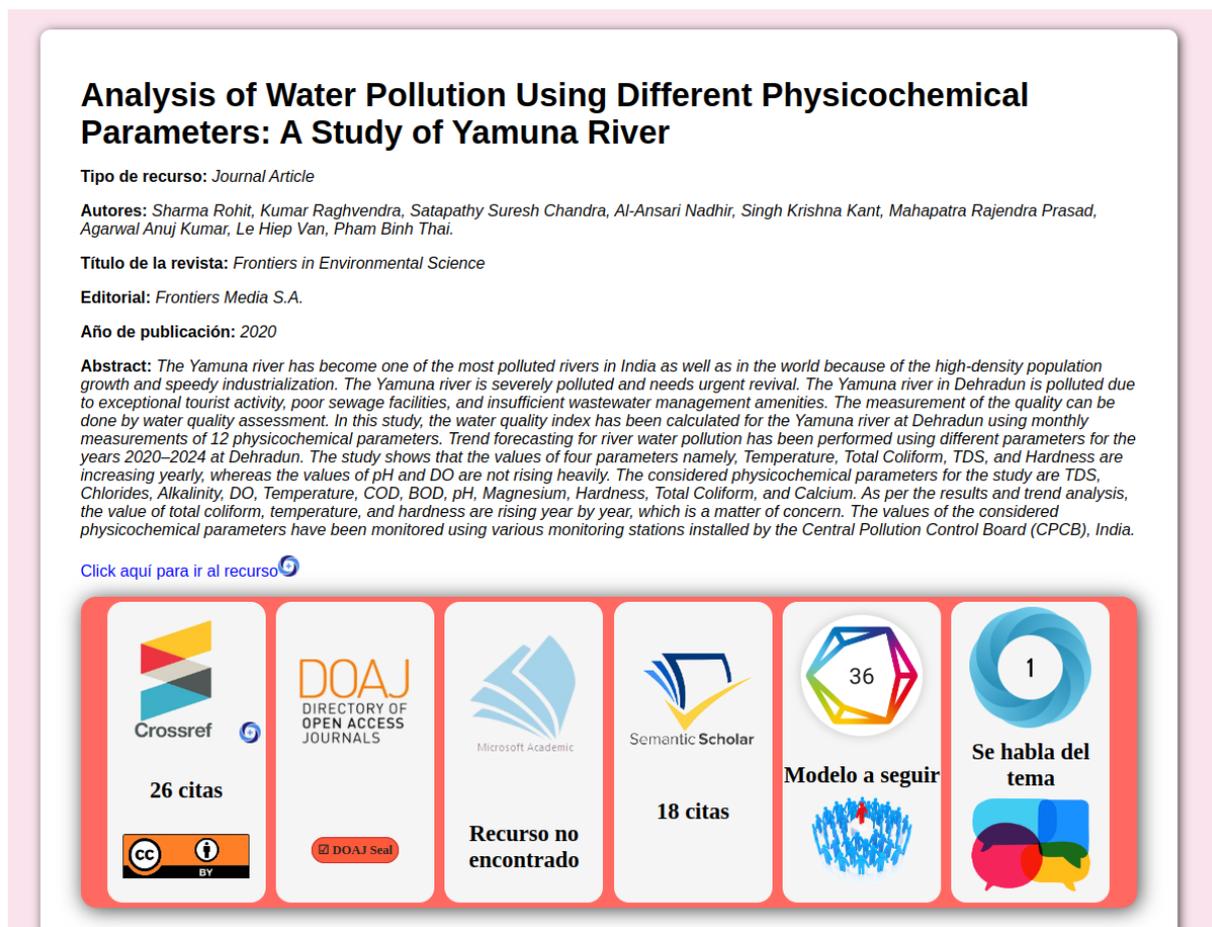


Figura 2.21: Vista resumida - Resultados generales de la búsqueda de un paper

Entre los resultados, se pueden apreciar metadatos del recurso, como título, tipo de recurso, resumen, entre otros, así como el nombre de la publicación donde se publicó el artículo y un enlace a su página original. Además, si es posible, la aplicación intentará recolectar métricas del *journal* en el que se encuentra el artículo para mostrarlas en conjunto (ver Figura 2.22).

Para ello, se extrae el ISSN del *journal* de los metadatos recuperados y se ejecuta automáticamente y en segundo plano una búsqueda por ISSN. De esta manera, se amplía el contexto de evaluación de un trabajo, al poder visualizar sus métricas de forma individual y también tener acceso a las del *journal*, para así comprender qué influencia podría tener en las del artículo. En general, la calidad de un artículo está asociada a la del *journal* en el que está publicado.

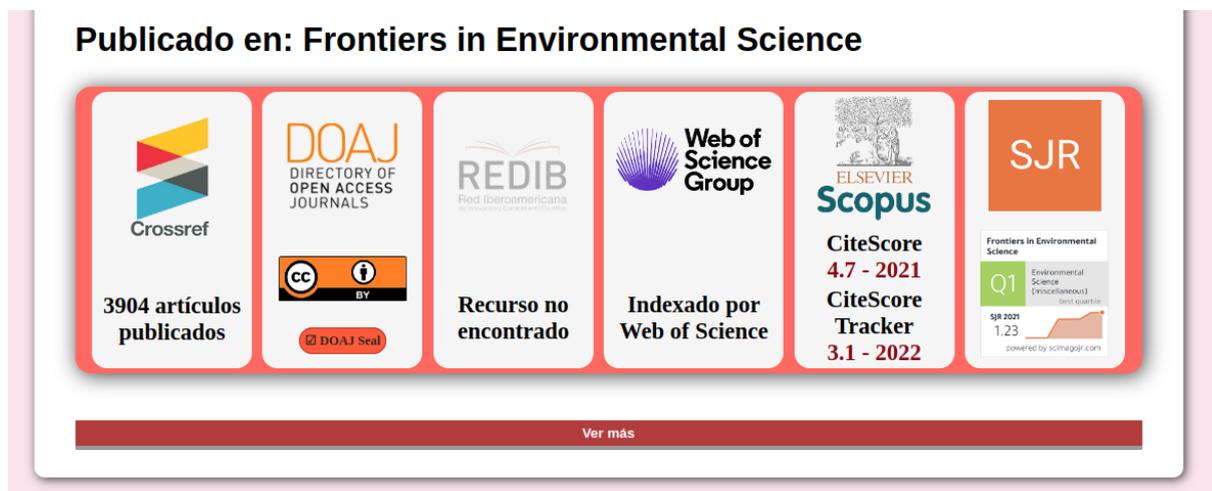


Figura 2.22: Vista Resumida - Métricas correspondientes al journal donde se encuentra el paper

Al presionar el botón "Ver más" que se muestra en esta última imagen, se genera una vista ampliada de las métricas en la que se presenta información más detallada (en aquellos casos donde las bases de datos proporcionan datos adicionales), como se puede ver en la Figura 2.23.

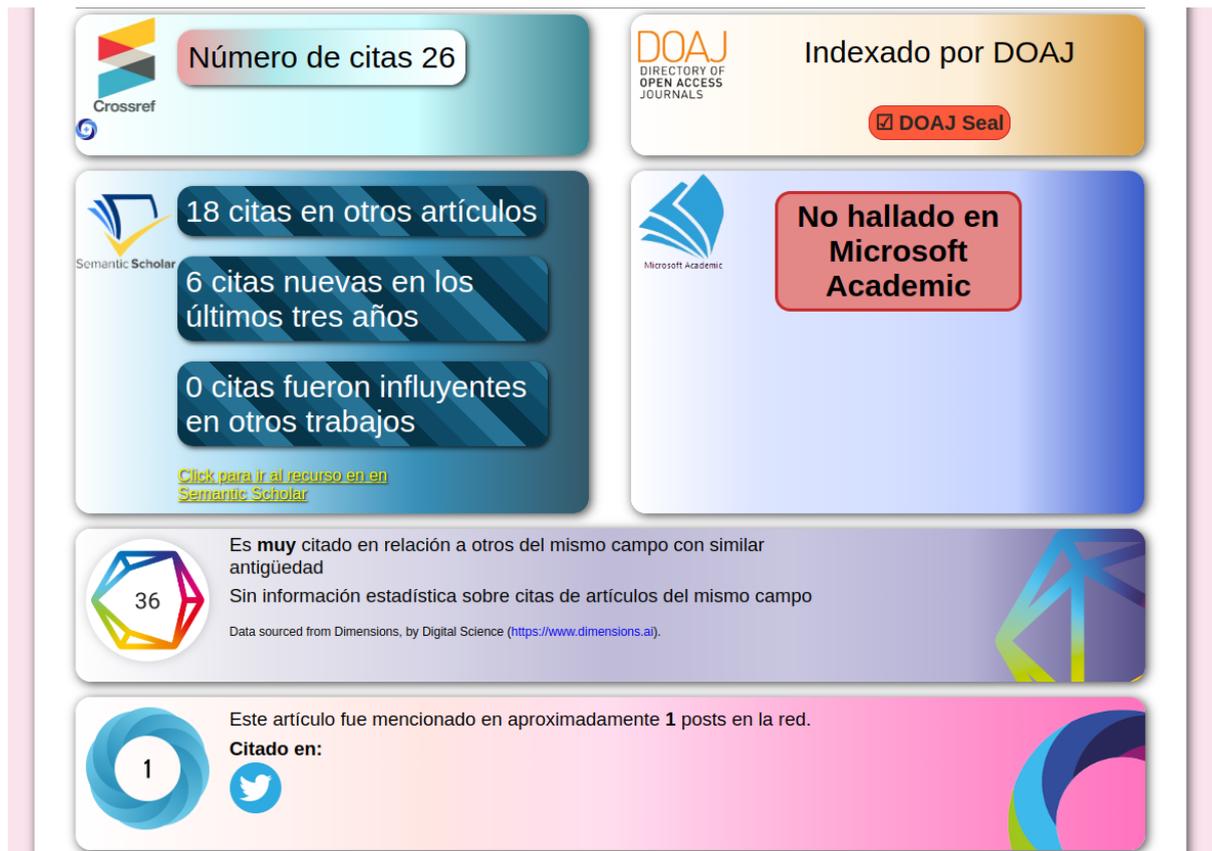


Figura 2.23: Vista Expandida - Métricas de un *paper*

2.3.4 Comparación con herramientas similares

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 Público? (DLP)¹¹ 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)¹², que da una vista amplia, detallada e integral de métricas de los revistas que posee en su base de datos. Cabe aquí mencionar que MIAR no provee información a nivel de artículo y que el proyecto DLP ha sido discontinuado.

¹¹ www.dondelopublico.com

¹² Urbano, C., Somoza-Fernández, M., Rodríguez-Gairín, J.M., Ardanuy, J., Guardiola, E., Pons, A., Borrego, , Brucart, J.M., Coscolluela, A.: MIAR : una base de datos para la identificación y la evaluación de la difusión secundaria de revistas de humanidades y ciencias sociales (2005), <http://eprints.rclis.org/6267/>

Capítulo 3

Propuestas

En el presente capítulo, se abordarán propuestas concretas destinadas a mejorar la aplicación HERA, la cual se denominará HERA 2.0, con el objetivo primordial de extender y mejorar su funcionalidad, además de optimizar su rendimiento. Estas propuestas tienen como meta principal ofrecer una experiencia mejorada tanto para los usuarios finales como para los desarrolladores involucrados en el proyecto.

En primer lugar, se propone una serie de mejoras con la perspectiva del usuario final en mente, lo cual implica abordar y evaluar aspectos relacionados con la usabilidad y la interfaz de usuario. Se plantea una nueva interfaz gráfica cuyo diseño se centra en mejorar la experiencia del usuario. Se desarrollarán propuestas detalladas que describirán los cambios específicos y la manera en que estos se traducirán en una mejor experiencia para el usuario.

Desde la perspectiva del desarrollador, se explorarán propuestas que busquen optimizar el rendimiento y la eficiencia de HERA 2.0. Esto implica evaluar el código existente y realizar mejoras en la arquitectura y la estructura del software. Se considerarán técnicas de optimización, como la reducción de la complejidad algorítmica, el uso eficiente de los recursos del sistema y la implementación de técnicas de almacenamiento en caché.

3.1 Nueva Interfaz

La propuesta para la nueva interfaz de la aplicación HERA se centra en mejorar la experiencia del usuario y optimizar la usabilidad del sistema. Para ello, se plantea el diseño de una interfaz que resulte más intuitiva y atractiva, y que permita a los usuarios acceder de manera sencilla a todas las funcionalidades de HERA. Se hará especial hincapié en la organización y estructura de los elementos visuales, utilizando un enfoque centrado en el usuario para garantizar que las tareas comunes sean fáciles de realizar. Además, se explorará la posibilidad de incorporar íconos claros y representativos, así como un esquema de colores coherente y agradable a la vista. También se prestará especial atención a la legibilidad de los textos, asegurando que el contenido se presente de forma clara y concisa. En definitiva, la propuesta para la nueva interfaz de HERA busca mejorar la experiencia de los usuarios al interactuar con la aplicación, facilitando su uso y maximizando su eficiencia.

3.2 Integración de información de autores e instituciones

La propuesta de integrar información adicional sobre los autores en HERA permitirá a los usuarios obtener una visión más amplia y contextualizada de las publicaciones

académicas que están buscando. La inclusión de información detallada sobre los autores puede ayudar a los usuarios a entender mejor el contexto y la relevancia de las publicaciones. Esto podría incluir información como la afiliación institucional del autor, su área de especialización, sus publicaciones anteriores y su índice de citación. Este nivel de detalle puede ser especialmente útil para los usuarios que están buscando expertos en un campo específico o que quieren evaluar la fiabilidad y relevancia de una publicación.

3.3 Actualización del banco de bases de datos

La actualización y expansión del banco de bases de datos es fundamental para mantener y mejorar la relevancia y utilidad de HERA. Esto implica un análisis exhaustivo de las fuentes de datos actuales, la identificación de nuevas fuentes de datos potenciales y la implementación de medidas para garantizar la continuidad del servicio.

3.3.1 Evaluación de fuentes de datos actuales

Es esencial evaluar las fuentes de datos actuales para determinar si siguen siendo útiles, si aún están disponibles y son compatibles con HERA. Por ejemplo, servicios como Crossref, DOAJ y Microsoft Academic serían revisados para asegurar que siguen proporcionando datos relevantes, que su API sigue siendo accesible y que los cambios en su estructura de datos o términos de uso no afectan a la funcionalidad de HERA.

3.3.2 Identificación de nuevas fuentes de datos

Además de evaluar las fuentes de datos actuales, se propone identificar y agregar nuevas fuentes de datos que puedan proporcionar información de artículos y revistas. Esto implica, por ejemplo, la exploración de bases de datos académicas, repositorios institucionales y otras fuentes de información académica. La incorporación de nuevas fuentes de datos apunta a aumentar la cantidad y diversidad de los resultados de búsqueda en HERA, permitiendo a los usuarios acceder a una gama más amplia y variada de información académica.

3.3.3 Robustez del servicio

Es importante implementar medidas para garantizar la continuidad del servicio de HERA, dado que las fuentes de datos pueden cambiar o dejar de estar disponibles. Este punto será detallado más adelante como una propuesta aparte en 3.5 Gestión de las fuentes de información.

3.4 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 [53] [6].

En HERA 1.0 se encontraron varios problemas o posibles oportunidades de optimización los cuales son:

- **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.

Se propone una refactorización integral del código de HERA para mejorar su eficiencia, mantenibilidad y escalabilidad.

3.4.1 Mejora de la eficiencia

La refactorización puede eliminar redundancias, optimizar las operaciones y mejorar la eficiencia general del código. Este proceso incluiría la revisión y optimización de las funciones de búsqueda y recopilación de datos, con el objetivo de mejorar el rendimiento de la aplicación y la velocidad de respuesta para los usuarios.

3.4.2 Mejora de la mantenibilidad

Un código limpio y bien estructurado es más fácil de entender y de mantener. La refactorización permitiría organizar el código de manera más coherente, con una estructura más modular y una nomenclatura más consistente. Esto facilita la detección y corrección de errores, así como la implementación de nuevas funcionalidades.

3.4.3 Mejora de la escalabilidad

La refactorización también es una oportunidad para mejorar la escalabilidad del código, preparándolo para futuras mejoras y expansiones. Esto podría implicar, por ejemplo, la introducción de patrones de diseño que permitan una mayor flexibilidad y reutilización de código, o la implementación de pruebas automáticas para garantizar la robustez del código a medida que se añaden nuevas funcionalidades.

3.4.4 Técnicas de refactorización

La refactorización implicaría técnicas como la extracción de métodos, la consolidación de fragmentos de código duplicado, la simplificación de condicionales y la creación de abstracciones para encapsular complejidad. Asegurarse de que cada método y clase tenga una responsabilidad única y clara también sería una parte importante de este proceso.

3.5 Gestión de las fuentes de información

A medida que HERA crece en términos de fuentes de datos y funcionalidades, surge la necesidad de una gestión más eficiente y mantenible. Los cambios inesperados en las fuentes de datos externas pueden resultar en la necesidad de revisar y modificar manualmente el código de la aplicación, una tarea habitualmente laboriosa y propensa a errores. Por lo tanto, se propone la implementación de una solución que permita una gestión más ágil y flexible de las fuentes de datos. La solución propuesta consiste en un archivo de configuración que almacena información relevante sobre cada fuente de datos, incluyendo su URL, su grado de prioridad, y si está habilitada o no. De esta manera, en lugar de codificar estos detalles particulares de cada fuente de datos directamente en la aplicación, se especificarán en el archivo de configuración. Gracias a esto, en caso que una fuente de datos deje de estar disponible o, por el contrario, en caso de añadir una nueva fuente de datos, los cambios necesarios se podrían hacer directamente en el archivo de configuración sin necesidad de modificar el código de la aplicación.

Además, se propone que la configuración permita establecer un nivel o grado de prioridad para cada fuente de datos, lo que conduciría a una construcción más eficiente del contexto. En particular, los datos de las fuentes con mayor prioridad se considerarán primero, asegurando que se utilice la información más completa y precisa disponible.

Finalmente, también se propone la inclusión de un atributo *'enabled'* en la configuración que permita activar o desactivar fácilmente las fuentes de datos, según sea necesario. De esta manera, se proveería a HERA de una rápida adaptación a las circunstancias cambiantes, como la desactivación de una fuente de datos debido a problemas de disponibilidad o la adición de una nueva fuente de datos a la lista de consulta.

Se considera que esta propuesta de configuración representa un gran avance en términos de la flexibilidad y mantenibilidad de HERA, lo que permitirá el continuo crecimiento y desarrollo de la plataforma.

3.6 Implementación de API REST

La implementación de una API *Representational State Transfer* (REST) [54] en HERA representa una mejora estratégica significativa, que permitirá una mayor interacción con otras aplicaciones y servicios, incrementará la eficiencia de las solicitudes y mejorará la escalabilidad de la plataforma.

3.7 Búsqueda simultánea de múltiples recursos

La implementación de una búsqueda simultánea en múltiples recursos en HERA 2.0 surge como una mejora significativa sobre su versión anterior, especialmente útil en

situaciones de evaluación exhaustiva de la producción científica de un investigador o de una institución.

Por ejemplo, si se desea hacer un seguimiento detallado del trabajo de un científico en particular o analizar la producción académica de una universidad durante un determinado período, el proceso con HERA 1.0 puede resultar largo y tedioso. Este proceso implicaría buscar y copiar cada DOI de los artículos del investigador, pegarlo en la interfaz de HERA, esperar a que se recuperen los datos de todas las bases de datos disponibles y finalmente, revisar y anotar la información relevante. Luego, el usuario tendría que repetir este proceso para el siguiente DOI y así sucesivamente. Para un investigador con decenas o incluso cientos de publicaciones, este proceso sería claramente ineficiente y propenso a errores, sin mencionar el consumo excesivo de tiempo.

El desarrollo de una funcionalidad que permita la búsqueda simultánea en múltiples recursos es una mejora significativa para HERA. Esta función aumentaría la eficiencia de las búsquedas, maximizaría la exhaustividad de los resultados y proporcionaría a los usuarios una visión más completa y diversificada de la información académica disponible.

3.7.1 Mayor eficiencia

Actualmente, la búsqueda de DOI o ISSN en diferentes recursos debe realizarse de forma individual, lo que puede ser un proceso lento y laborioso. La implementación de una búsqueda simultánea de múltiples recursos permitiría a los usuarios obtener resultados de todos los recursos solicitados con una sola consulta, mejorando significativamente la eficiencia del proceso de búsqueda.

3.7.3 Mejor visión general

Además, la capacidad de buscar simultáneamente en múltiples recursos puede dar a los usuarios una visión más completa y diversificada de la información académica disponible. Esto puede ser especialmente útil para los usuarios que están realizando investigaciones interdisciplinarias o que buscan información en campos de estudio con diferentes tradiciones de publicación.

3.7.4 Implementación

La implementación de esta funcionalidad implicaría el desarrollo de un sistema que pueda enviar consultas a múltiples API de bases de datos de manera simultánea, recopilar y procesar los resultados, para luego presentarlos al usuario de manera coherente y fácil de entender. Resulta claro que se requerirán mejoras en las capacidades de recopilación y procesamiento de datos de HERA, así como ajustes en la interfaz de usuario para acomodar los resultados de múltiples recursos.

3.8 Exportación a formatos descargables

HERA 1.0 es capaz de proveer al usuario de una amplia variedad de indicadores y métricas para un recurso académico. Sin embargo, esta información siempre se muestra en pantalla, sin que sea posible descargarla para realizar algún tipo de análisis o procesamiento particular.

Por lo tanto, se propone la posibilidad de exportar los resultados de la búsqueda a formatos descargables, como JSON y CSV. Al permitir la exportación de los resultados de la búsqueda en formatos descargables, los usuarios de HERA tendrán una mayor flexibilidad en cuanto al uso de los datos. Por ejemplo, podrán guardar los resultados de la búsqueda para su uso futuro, compartirlos con colegas o analizarlos utilizando herramientas externas.

La capacidad de exportar los resultados en formatos como JSON y CSV también facilitará la integración de HERA con otras herramientas y flujos de trabajo. Los usuarios podrían, por ejemplo, importar los datos en una hoja de cálculo para realizar análisis adicionales, o utilizarlos en aplicaciones de programación para desarrollar nuevos análisis o visualizaciones.

3.9 Extensión para navegadores

La propuesta de desarrollar una extensión para navegadores representa una mejora estratégica clave para HERA. La extensión propuesta proporcionará una interfaz más accesible y conveniente para interactuar con HERA, permitiendo una integración más fluida en el flujo de trabajo del usuario y mejorando la experiencia del usuario.

Para agilizar y ampliar el uso de HERA como herramienta para la evaluación de publicaciones, se propone el desarrollo de una extensión para navegadores que sea capaz de identificar las publicaciones mencionadas en páginas de buscadores o bases de datos académicas.

3.9.1 Funcionalidad adaptativa de la búsqueda

La extensión del navegador ofrecerá funcionalidades de búsqueda adaptativas, dependiendo del tipo de consulta. Si la búsqueda es sobre un DOI específico, la extensión mostrará información completa relacionada con ese DOI. En caso de que la búsqueda sea sobre un tema que incluya varios DOI, la extensión proporcionará un resumen de información para cada DOI. El objetivo es que los usuarios tengan una visión rápida y detallada de la información académica relevante directamente desde el sitio que están navegando, sin necesidad de realizar las búsquedas manualmente en HERA.

3.9.2 Enlace directo a HERA

Además, todos los componentes que muestren información sobre un DOI en la extensión tendrán un botón que enlazará directamente a la página completa de ese DOI en HERA. Esta característica permitirá a los usuarios acceder a una vista más detallada y completa de la información de un DOI con solo un clic, agilizando el acceso a información más completa, ya que se evita de esta manera el proceso de buscar manualmente en HERA.

3.9.3 Mejora de la integración y la experiencia del usuario

La extensión del navegador permitirá a los usuarios integrar HERA en su flujo de trabajo de una manera más cómoda y eficiente, mejorando la utilidad y la experiencia del usuario. Por ejemplo, los usuarios podrían usar la extensión para buscar información adicional sobre un artículo mientras leen un resumen en otra pestaña del navegador, mejorando la facilidad y la velocidad con la que pueden obtener y utilizar la información académica.

3.10 Manejo de errores y recursos no encontrados

Una propuesta adicional para mejorar HERA es optimizar el manejo de errores y la respuesta ante un DOI o ISSN no encontrado. Actualmente, cuando HERA no encuentra información de un recurso, simplemente muestra un mensaje genérico como "no se encontró el recurso". Esta experiencia puede ser confusa y frustrante para los usuarios, especialmente cuando no está claro si el problema se debe a que el recurso no existe, si hay algún problema con la API que se está consultando, o si hay un exceso de solicitudes que está impidiendo una respuesta adecuada.

Por lo tanto, se propone mejorar la precisión de los mensajes de error para proporcionar más claridad a los usuarios. Esto podría incluir la distinción entre cuando un recurso no se encuentra y cuando el servicio no está disponible.

3.11 Base de datos local

HERA 1.0, depende en gran medida de fuentes de datos externas para recolectar y presentar información académica. Sin embargo, esta dependencia puede presentar problemas, ya que si una fuente de datos externa no está disponible por cualquier razón (por ejemplo, mantenimiento del servidor, cambios en la estructura de la web, políticas de acceso, etc.), la herramienta no puede mostrar la información relacionada. Esta situación podría limitar la efectividad de HERA en su objetivo de proporcionar una visión completa y precisa de los recursos académicos.

En este contexto, surge la propuesta de adoptar una base de datos local para HERA, con el fin de superar estos desafíos. La implementación de una base de datos local permitiría a HERA almacenar y manejar datos de recursos académicos de manera más directa y controlada, mejorando la eficiencia y la escalabilidad del

sistema. Para llevar a cabo esta mejora, será necesario realizar un estudio exhaustivo de los snapshots disponibles, los cuales son instantáneas de bases de datos tomadas en un punto específico en el tiempo y podrían proporcionar un recurso valioso para la creación de esta base de datos local.

Ventajas de una base de datos local

Existen varias ventajas potenciales asociadas con la implementación de una base de datos local. Primero, permitiría a HERA tener un mayor control sobre los datos que maneja, lo que podría resultar en una mayor precisión y consistencia de los resultados. Segundo, podría mejorar la velocidad de respuesta global del sistema, ya que se reduciría la necesidad de realizar solicitudes repetidas a las mismas fuentes de datos externas. Tercero, se podría mejorar la escalabilidad del sistema, permitiendo que HERA 2.0 maneje una mayor cantidad de datos de manera más eficaz.

Desventajas de una base de datos local

Lamentablemente, la adopción de una base de datos local también implicaría desafíos y limitaciones potenciales asociados con su implementación. Primero, la creación y mantenimiento de una base de datos local no es una tarea trivial y requeriría de recursos significativos. Además, habría que tener en cuenta la posibilidad de que los datos almacenados en la base de datos local puedan volverse obsoletos o inexactos con el tiempo, lo que obliga a contemplar estrategias para la actualización regular de los datos.

3.12 Resumen

En este capítulo, se han sugerido diversas propuestas de extensiones y mejoras para la aplicación HERA, con la finalidad de potenciar y expandir la funcionalidad de la herramienta. Dichas propuestas están diseñadas para mejorar el rendimiento y la escalabilidad de HERA, expandir su alcance al incorporar más bases de datos académicas, y fortalecer el soporte a través de la refactorización del código y la gestión eficiente de las bases de datos de consulta.

1. **Nueva Interfaz:** se propuso una interfaz de usuario más simple, que resulte más atractiva y accesible, y que facilite la interacción con la aplicación.
2. **Integración de información de autores e instituciones:** se sugirió la inclusión de información detallada de los autores en las publicaciones, lo que proporcionaría un contexto más completo y facilitaría la búsqueda de artículos relacionados.
3. **Actualización del banco de bases de datos:** se propuso la revisión regular y actualización de las bases de datos consultadas por HERA, añadiendo nuevas fuentes de información y asegurando la disponibilidad de las actuales.

4. **Refactorización de código:** se propuso realizar un proceso de refactorización de código para mejorar la eficiencia, la escalabilidad y la mantenibilidad de la aplicación.
5. **Gestión de las fuentes de información:** en cuanto a la gestión de las fuentes de información, se plantea la implementación de un sistema más eficiente y mantenible a través de un archivo de configuración. Esta solución facilitará la adaptación a cambios inesperados en las fuentes de datos, evitando la modificación manual del código de la aplicación. Adicionalmente, se propone la asignación de grados de prioridad a cada fuente de datos y la posibilidad de activar o desactivarlas según sea necesario, aumentando la flexibilidad de HERA y facilitando su evolución continua.
6. **Implementación de API REST:** se planteó la implementación de una API REST para facilitar la integración de HERA con otras aplicaciones y permitir una mayor flexibilidad en el acceso a sus funcionalidades.
7. **Búsqueda Simultánea de Múltiples Recursos:** se propuso el desarrollo de una funcionalidad para la búsqueda simultánea en múltiples bases de datos, aumentando la eficiencia y exhaustividad de las búsquedas.
8. **Exportación a formatos descargables:** se propuso la capacidad de exportar los resultados de búsqueda en formatos descargables (JSON y CSV), proporcionando mayor flexibilidad para el uso de la información y facilitando la integración de HERA con otras herramientas.
9. **Extensión para navegador:** se propuso el desarrollo de una extensión de navegador para agilizar y ampliar el uso de HERA, permitiendo la búsqueda de DOI directamente desde el sitio que estén navegando, sin necesidad de realizar la búsqueda a través del sitio de HERA.
10. **Manejo de errores y recursos no encontrados:** se sugiere diferenciar los casos de ausencia de información; en particular, la propuesta es diferenciar cuando el servicio está disponible pero el recurso no es encontrado de cuando el servicio directamente no está disponible. De esta manera, se proporcionarán mensajes de error más claros y útiles para los usuarios.
11. **Implementación de una base de datos:** se planteó la implementación de una base de datos local para superar los desafíos de la dependencia de HERA a fuentes de datos externas, lo que podría mejorar su eficiencia y escalabilidad.

Estas mejoras propuestas tienen como objetivo mejorar la experiencia del usuario y la utilidad de HERA.

Capítulo 4

Resultados Obtenidos

En este capítulo, se discute la posible implementación de cada una de las mejoras propuestas en el Capítulo 3 y se describe el trabajo realizado. Luego, se presentan y analizan los resultados obtenidos a partir de las mejoras realizadas. Estos resultados muestran cómo las mejoras propuestas han impactado la funcionalidad, el rendimiento y la usabilidad de la aplicación, y cómo han contribuido a lograr los objetivos establecidos. Por último, la versión BETA de HERA 2.0 se encuentra disponible en <https://hera.testing.sedici.unlp.edu.ar>

4.1 Resultados de: Nueva Interfaz

Para la implementación de la nueva interfaz en HERA 2.0, se optó por incorporar el uso de React-Bootstrap [55], una combinación de Bootstrap y React. 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 atractivas [56]. React, por otro lado, es una biblioteca de JavaScript para la construcción de interfaces de usuario de manera eficiente y flexible [57].

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 4.1.

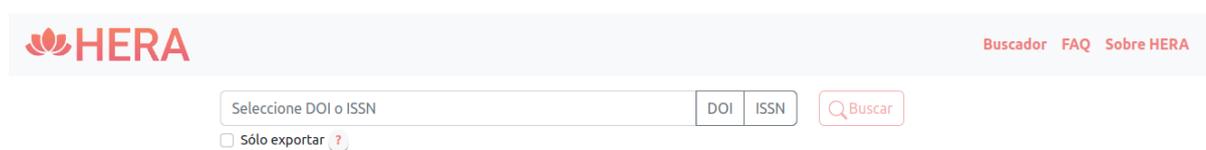


Figura 4.1: Interfaz de búsqueda de la herramienta con input de búsqueda y selectores de tipo de identificador (DOI e ISSN)

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. En las secciones posteriores ([ver sección 4.7](#)) se presentará el flujo de búsqueda detallado para cada tipo de recurso.

Búsqueda de un *paper*

Después de ingresar un DOI en la barra de búsqueda y hacer clic en el botón de búsqueda, las métricas se recopilan de diversas bases de datos y se muestran en la pantalla, tal como se ilustra en la Figura 4.2. 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 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 4.3. 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 4.4, 4.5 y 4.6.

10.7717/peerj.4375 DOI ISSN Buscar
 Sólo exportar ?

Exportar Exportar

DOI/10.7717/peerj.4375

The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles

Autores:

- **Heather A. 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](#)

Año de publicación: 2018

Título de la revista: PeerJ

Editorial: PeerJ, Inc.

Tipo de recurso: journal-article

Resumen:

Despite growing interest in Open Access (OA) to scholarly literature, there is an unmet need for large-scale, up-to-date, and reproducible studies assessing the prevalence and characteristics of OA. We address this need using oaDOI, an open online service that determines OA status for 67 million articles. We use three samples, each of 100,000 articles, to investigate OA in three populations: (1) all journal articles assigned a Crossref DOI, (2) recent journal articles indexed in Web of Science, and (3) articles viewed by users of Unpaywall, an open-source browser extension that lets users find OA articles using oaDOI. We estimate that at least 28% of the scholarly literature is OA (19M in total) and that this proportion is growing, driven particularly by growth in Gold and Hybrid. The most recent year analyzed (2015) also has the highest percentage of OA (45%). Because of this growth, and the fact that readers disproportionately access newer articles, we find that Unpaywall users encounter OA quite frequently: 47% of articles they view are OA. Notably, the most common mechanism for OA is not Gold, Green, or Hybrid OA, but rather an under-discussed category we dub Bronze: articles made free-to-read on the publisher website, without an explicit Open license. We also examine the citation impact of OA articles, corroborating the so-called open-access citation advantage: accounting for age and discipline, OA articles receive 18% more citations than average, an effect driven primarily by Green and Hybrid OA. We encourage further research using the free oaDOI service as a way to inform OA policy and practice.

Figura 4.2. Resultados generales de la búsqueda de un *paper* (vista resumida)

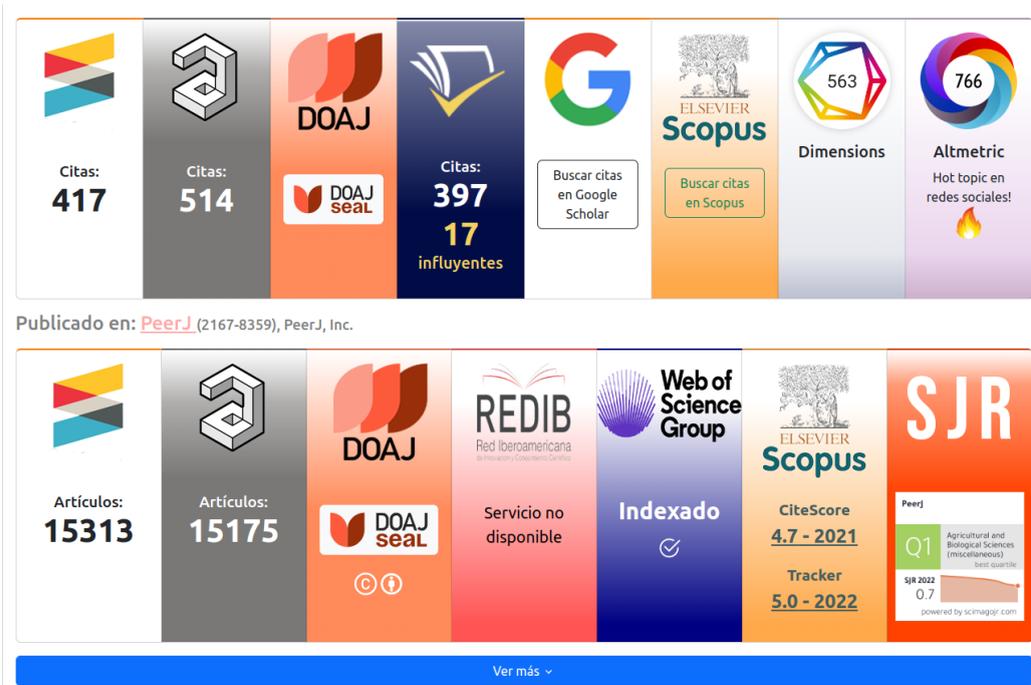


Figura 4.3. Métricas correspondientes al *paper* y *journal* donde se encuentra el *paper* (vista resumida en tarjetas)

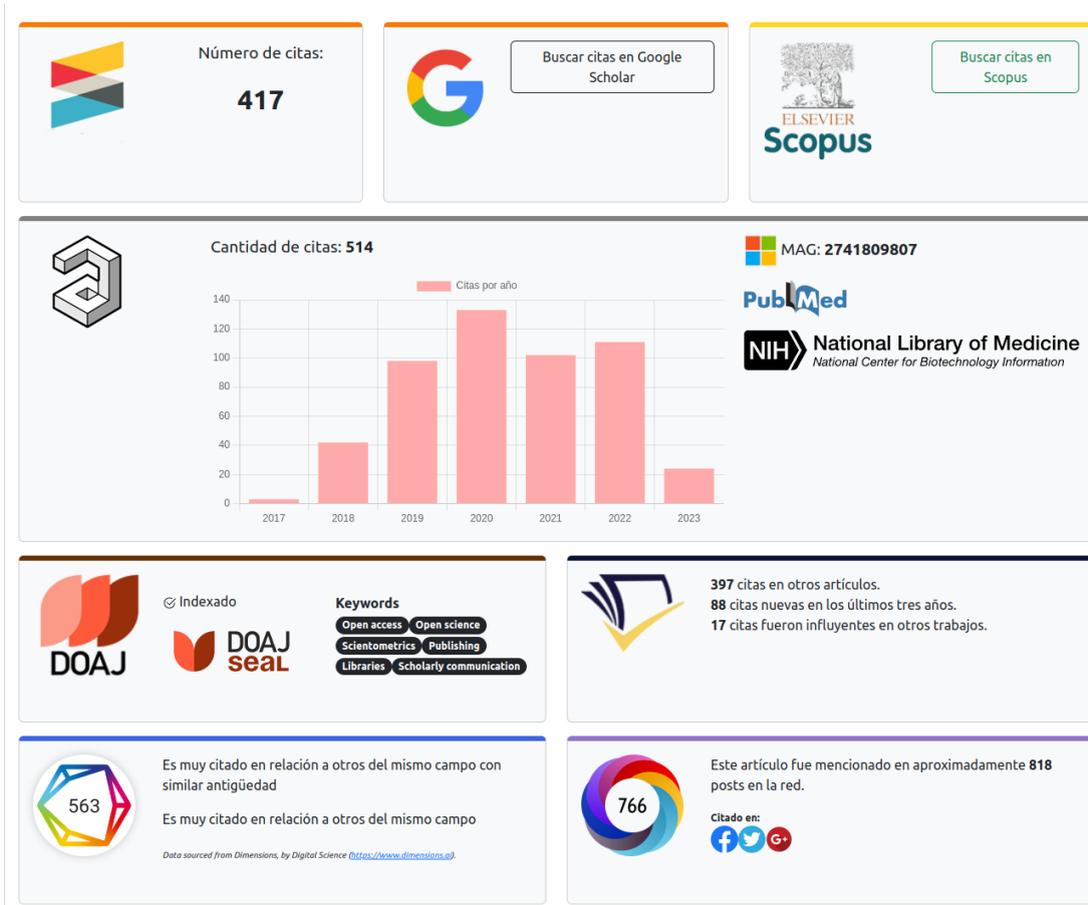


Figura 4.4. Vista expandida de métricas de un paper

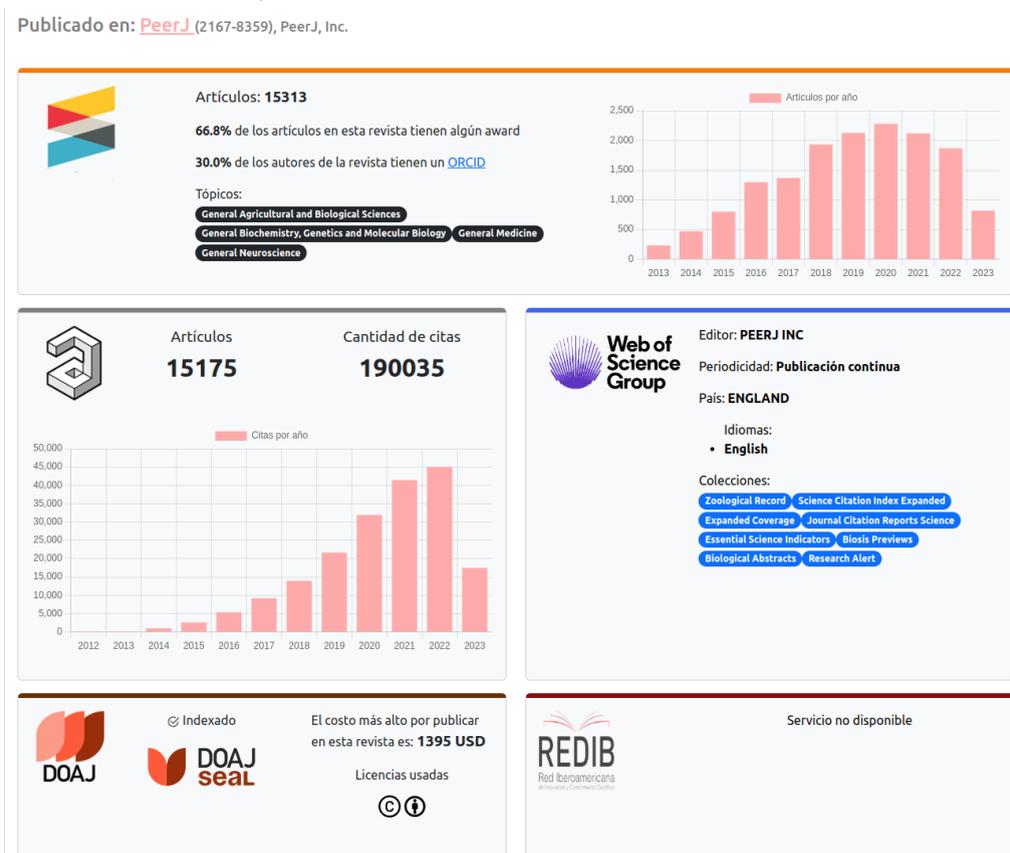


Figura 4.5 Vista expandida de métricas de un journal (parte 1)

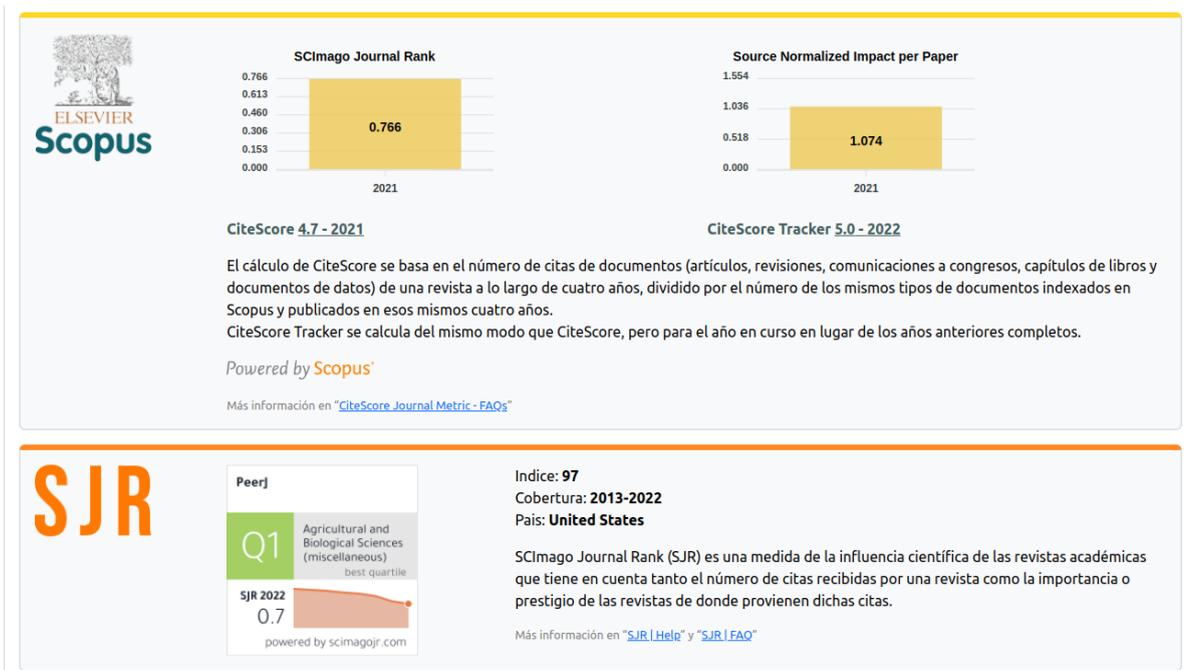


Figura 4.6. Vista expandida de métricas de un *journal* (parte 2)

Búsqueda de un *journal*

Similar a la búsqueda de un *paper*, la información del *journal* se presenta de manera organizada y accesible, como se puede ver en la siguiente figura 4.7.

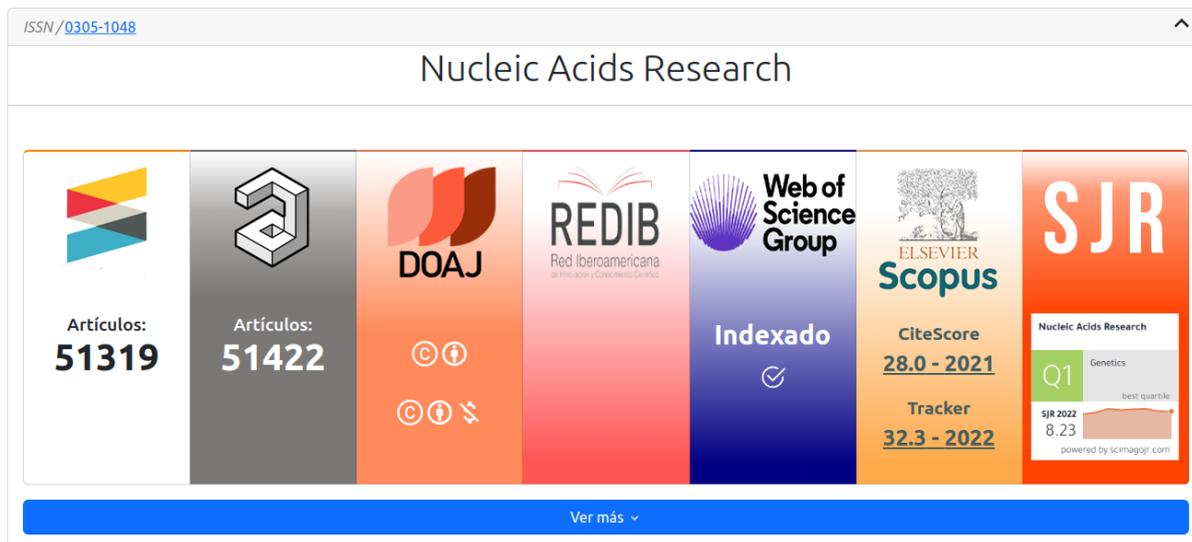


Figura 4.7 Resultados generales de la búsqueda de un *journal* (vista resumida)

Los resultados de la búsqueda de un *journal* proporcionan una visión detallada de las características y el rendimiento del *journal*, incluyendo detalles como el factor de impacto, cantidad de artículos, entre otros.

Cabe destacar que, como en la búsqueda de *papers*, la presentación de la información de los *journals* en HERA depende de la disponibilidad de los datos. No todos los *journals* pueden tener todos los detalles disponibles. Sin embargo, en

todos los casos, HERA se esfuerza por proporcionar la mayor cantidad de información posible.

Ajustes de interfaz

Un cambio en la interfaz ha sido la forma en que se muestran las estadísticas. En lugar de presentar las estadísticas en un formato de texto simple como se puede ver en la figura 4.8 (por ejemplo, 232: 2013, 471: 2014, 801:2015 ...), en la nueva versión se utilizan gráficos de barras para representar visualmente los datos (ver Figura 4.9). Considerando que dicho ajuste facilita la interpretación de las estadísticas y mejora la estética general de la aplicación.

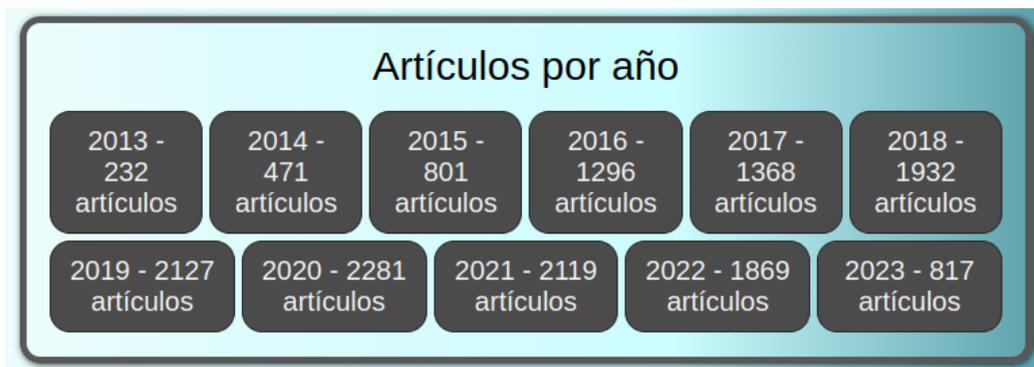


Figura 4.8 Artículos por año según Crossref en HERA 1.0



Figura 4.9 Artículos por año según Crossref en HERA 2.0

Como parte de los esfuerzos por mejorar la experiencia del usuario y la legibilidad de la información, la presentación de los datos en las tarjetas de HERA ha sido simplificada y compactada. Aunque la cantidad y el tipo de información permanecen iguales, la forma en que se muestra ha sido revisada y mejorada.

Cada tarjeta ahora presenta la información de manera más compacta y visualmente atractiva. Considerando que este diseño más reducido y estilizado no solo hace que las tarjetas sean más agradables a la vista, sino que también permite a los usuarios identificar y procesar la información clave más rápidamente.

La simplificación del diseño de las tarjetas también tiene en cuenta la funcionalidad de búsqueda múltiple de HERA, cuyos resultados se detallan en la sección 4.7. Al presentar la información de manera más compacta, se pueden mostrar más resultados en la misma cantidad de espacio, lo que facilita la comparación de múltiples *papers*, *journals*, DOI o ISSN.

En las Figuras 4.10 y 4.11 se puede ver una comparación en el cambio de las tarjetas.

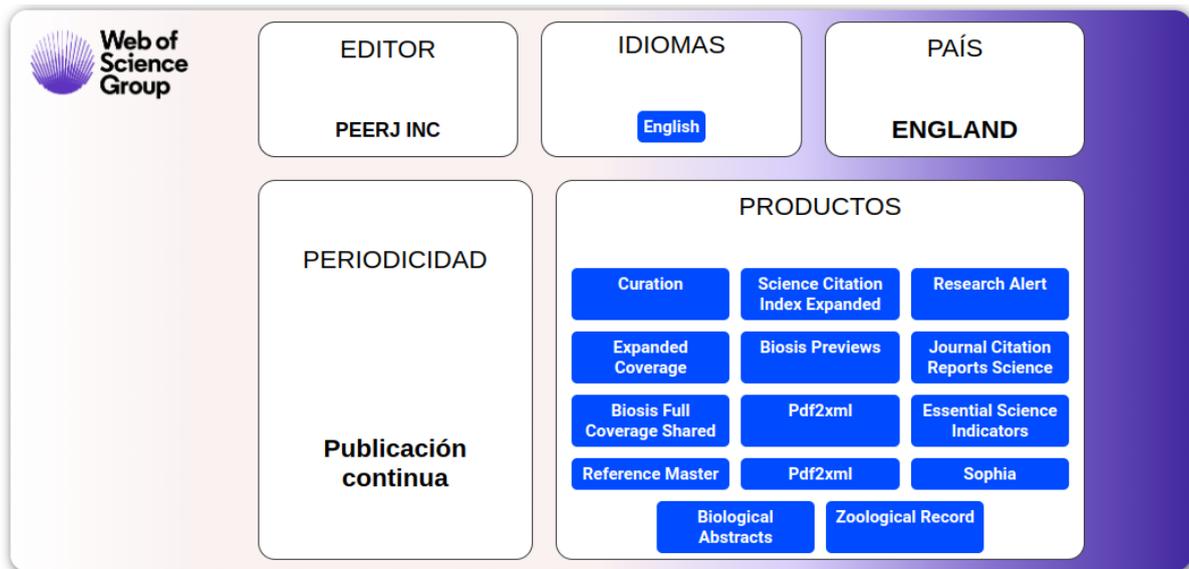


Figura 4.10 Visualización de WoS de la versión 1.0 de HERA

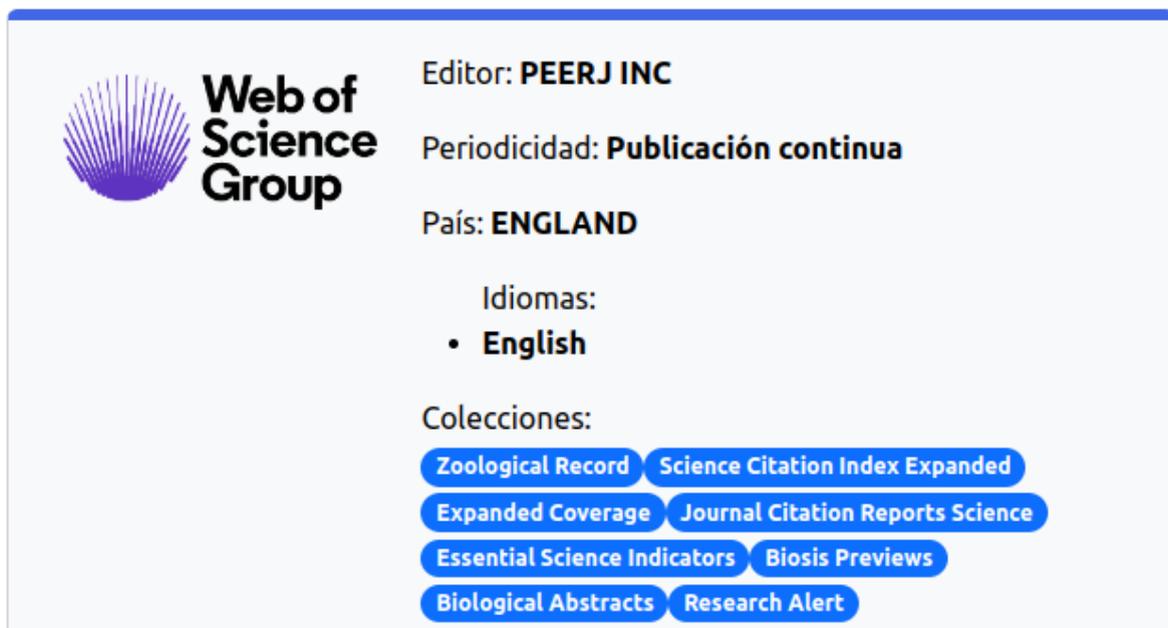


Figura 4.11 Visualización de WoS de HERA 2.0

Además, se han integrado iconos en la nueva interfaz para aumentar la claridad y la facilidad de navegación (ver en el ejemplo de la Figura 4.12). Los íconos, junto con

el sistema *responsivo* y la gama de colores heredada de la versión 1.0 de HERA, se consideran factores que podrían contribuir a una interfaz más intuitiva y atractiva.



Figura 4.12 Botones para exportar la información (csv o json)

Sólo exportar

En la versión 2.0 de HERA, debajo del campo de búsqueda, los usuarios pueden encontrar una casilla de verificación etiquetada como "Sólo exportar". Esta opción, como se indica en el mensaje de ayuda emergente que aparece al pasar el cursor sobre el icono de ayuda contiguo, permite a los usuarios obtener los datos de exportación sin tener que visualizarlos primero en la interfaz web. Al seleccionar esta opción, HERA genera una lista de los DOI o ISSN solicitados, y presenta los botones para exportar los datos en formato CSV o JSON, reduciendo así el tiempo de procesamiento general.

La Figura 4.13 ilustra cómo aparece esta nueva característica en la interfaz de usuario. Se ofrece una explicación más detallada sobre el proceso de exportación de datos en la sección [4.8](#) del presente documento.

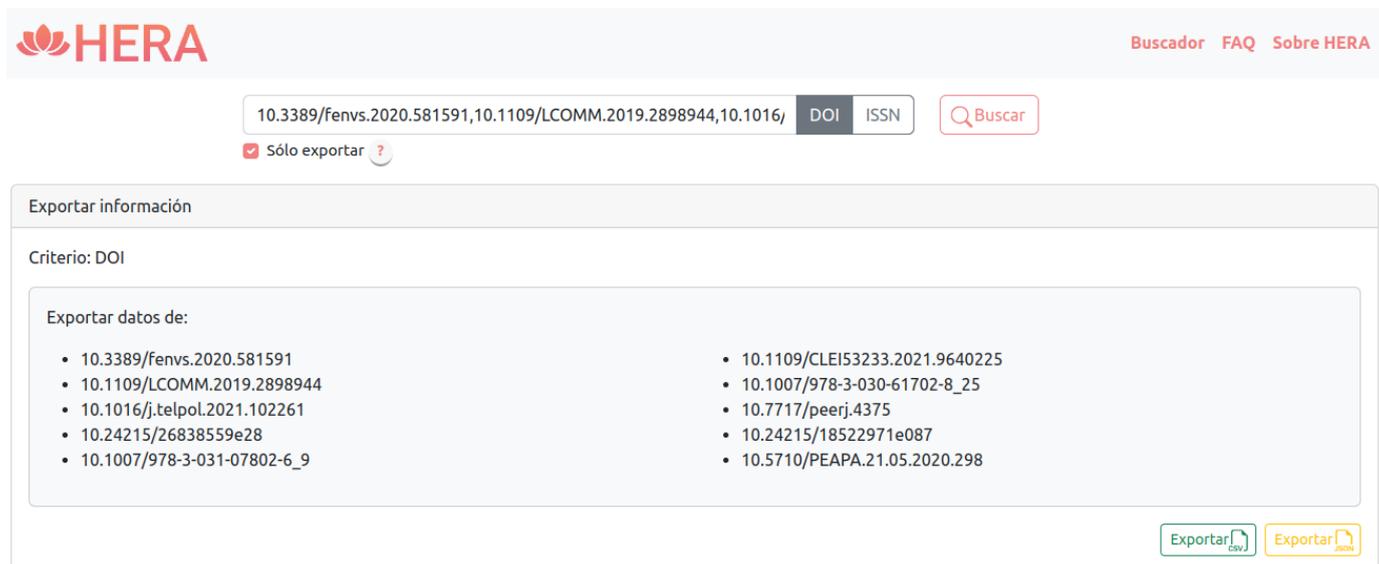


Figura 4.13 Función de "Sólo exportar" en la web de HERA 2.0

Por último, el logotipo de HERA fue rediseñado (ver Figura 4.14), manteniendo cohesión visual con la versión 1.0, con la intención de proporcionar una imagen fresca y moderna que se pretende sea representativa de las mejoras realizadas en la aplicación.



Figura 4.14 Logo HERA 2.0

4.2 Resultados de: 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: ROR y ORCID.

ROR (Research Organization Registry) [58] es un registro global y comunitario que proporciona identificadores únicos para las organizaciones que participan en la producción de investigación. 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.

ORCID (Open Researcher and Contributor ID) [59] por su parte, proporciona un identificador digital persistente que distingue a un investigador de otros. 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 carrera de cada autor, lo que mejora la utilidad y la riqueza de la información que HERA 2.0 puede proporcionar.

Como se observa en la Figura 4.15, dicha implementación brinda varios beneficios para los usuarios de HERA. Por un lado, permite conocer la institución de cada autor, dando así una perspectiva más amplia sobre el contexto en el que se realizó la investigación. Además, posibilita la identificación de colaboraciones entre diferentes instituciones o investigadores. Por otro lado, permite el acceso a otras publicaciones de los mismos autores, lo que ayuda a los usuarios a seguir de cerca la trayectoria de los investigadores de su interés.

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](#)

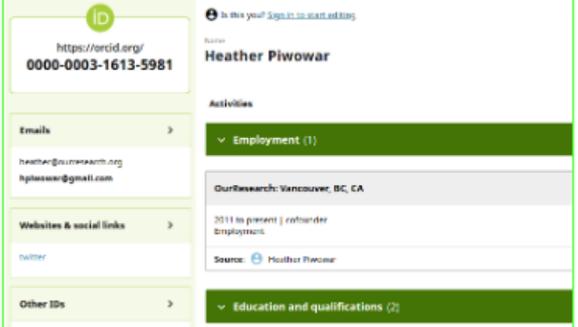


Figura 4.15: Ejemplo de vista de autores en HERA 2.0.

4.3 Resultados de: 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.

Evaluación de las Fuentes de Datos de HERA 1.0

Evaluar las fuentes de datos actuales es imprescindible para confirmar su idoneidad y compatibilidad continuada con HERA. En este contexto, 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 ya no está operativo, tal como se anunció en un comunicado oficial en mayo de 2021, indicando que su sitio web y las API subyacentes se retirarían el 31 de diciembre de 2021 (detalle ampliado en la sección [2.2.10](#)). 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 proyecto¹³. 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 [60] [61], la cual es una plataforma de datos académicos de acceso abierto desarrollada por la organización sin fines de lucro OurResearch [62]. 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.

Como se verá más adelante, la adición de OpenAlex mejora tanto la cantidad como la diversidad de los resultados de búsqueda en HERA, permitiendo a los usuarios acceder a un rango más amplio de información académica. La descripción detallada y la integración de OpenAlex en HERA se discutirán en la siguiente sección.

OpenAlex

OpenAlex recopila y organiza datos de diversas fuentes, como revistas, editoriales, repositorios y otras plataformas de investigación. Utiliza algoritmos y técnicas de aprendizaje automático para identificar, enlazar y estructurar la información de manera coherente y útil. Esto permite a los usuarios acceder y analizar datos académicos de manera eficiente.

Las dos fuentes de datos más importantes son MAG y Crossref. Otras fuentes clave incluyen:

- ORCID
- ROR
- DOAJ
- Unpaywall
- Pubmed
- Pubmed Central
- El Centro Internacional ISSN
- Internet Archive
- Rastros web

¹³ [Comunicado Oficial de REDIB](#)

- Repositorios de áreas temáticas e institucionales, desde arXiv hasta Zenodo y todos los demás en el medio.

Estas fuentes de datos contribuyen a la riqueza y diversidad de la información académica disponible en OpenAlex, permitiendo a los usuarios acceder a un amplio espectro de datos de investigación en un solo lugar.

OpenAlex es importante debido a su compromiso con el acceso abierto y la transparencia en el ámbito académico. Facilita la colaboración, el descubrimiento y la innovación en la investigación al ofrecer una fuente confiable y accesible de datos académicos. Además, al ser una plataforma de acceso abierto, ayuda a reducir las barreras económicas y de información en el ecosistema de investigación¹⁴.

Comparación OpenAlex con otras fuentes de datos académicos:

Fuente	Works	Open Access Works	Citas	Apertura de los datos
OpenAlex	250M	43.8M	1.9B	Totalmente abierto, licencia CC0
Scopus	87M	20.5M [61]	1.8B	Cerrado
WoS (core)	87M [63]	12M [64]	1.8B	Cerrado
Dimensions	135M	29M [65]	1.7B	Parcialmente abierto, uso personal
Google Scholar	389M (estimado) [66]	?	?	Cerrado
Crossref	145M	20M	1.45B	Totalmente abierto, licencia CC0

Tabla 4.1: Comparación de fuente de datos académicas ¹⁵

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. Esto se debe a que OpenAlex ya contiene la información de Crossref. Además, este cambio permitió ampliar las capacidades de HERA para incluir información y enlaces a PubMed¹⁶ y PubMed Central¹⁷.

¹⁴ [OpenAlex - About](#)

¹⁵ [OpenAlex - Comparison with other scholarly data sources](#)

¹⁶ [PubMed - Sitio Web](#)

¹⁷ [PubMed Central - Sitio Web](#)

Los principales datos obtenidos de OpenAlex para un DOI e ISSN abarcan tanto métricas, así como información adicional:

- Título y resumen
- Cantidad de citas en un artículo (total y anual).
- Número de trabajos publicados en una revista (total y anual).
- Autores en conjunto con sus respectivas instituciones, datos proporcionados por ORCID y ROR.
- Tipo de publicación (journal, repositorio, conferencia, ebook).
- Tipo de recurso (sección de libro, monografía, reporte, artículo, etc).
- Editorial.
- Lugares de publicación (host, venue).

Estos elementos se obtienen mediante el uso de la API REST proporcionada por esta base de datos, la cual cuenta con una amplia documentación (ver Figura 4.16) [67].

- Trabajos (*Works*): documentos académicos como artículos de revistas, libros, conjuntos de datos y tesis.
- Autores (*Authors*): personas que crean documentos académicos.
- Fuentes (*Sources*): lugares donde se alojan los documentos académicos (como revistas, conferencias y repositorios).
- Instituciones: universidades y otras organizaciones a las que los autores reclaman afiliaciones.
- Conceptos (*Concepts*): temas asignados a los documentos académicos.
- Editores (*Publishers*): empresas (compañías) y organizaciones que distribuyen obras.
- Financiadores (*Funders*): organizaciones que financian la investigación.
- Geo: dónde se encuentran en el mundo.

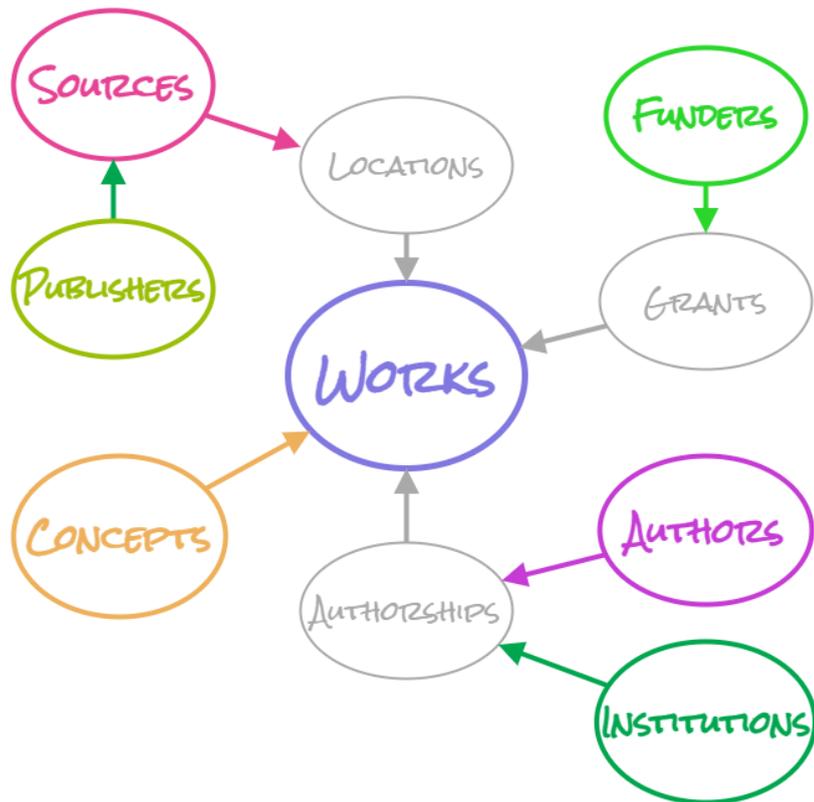


Figura 4.16: Entidades y relaciones entre las entidades de la API de OpenAlex.[68]

En la siguiente Figura 4.17 se ilustra la página web en desarrollo de OpenAlex visualizando el artículo: *The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles*.

explore.OpenAlex

The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles

2018 journal article published by [PeerJ](#).

Gold Open Access

[Citation](#) [License](#) [Scholarly communication](#) [Open science](#) [Web of science](#) [Bibliometrics](#) [World Wide Web](#) [Computer science](#)
[Citation analysis](#) [Gold standard \(test\)](#) [Medicine](#) [Library science](#) [Political science](#) [Meta-analysis](#) [Internal medicine](#) [Statistics](#)
[Mathematics](#) [Publishing](#) [Law](#) [Operating system](#)

[Heather A. Plowar](#) ([Our Research](#))
[Jason Priem](#) ([Our Research](#))
[Vincent Larivière](#) ([University of Montreal](#))
[Juan Pablo Alperin](#) ([Public Knowledge Project](#))
[Lisa Matthias](#) ([Simon Fraser University](#))
[Bree Norlander](#) ([University of Washington](#))
[Ashley Farley](#) ([University of Washington](#))
[Jevin D. West](#) ([University of Washington](#))
[Stefanie Haustein](#) ([University of Quebec at Montreal](#), [University of Ottawa](#))

[VIEW IN API](#) [VIEW WORK](#)

Identifiers

- **openalex:** <https://openalex.org/W2741809807>
- **doi:** <https://doi.org/10.7717/peerj.4375>
- **mag:** 2741809807
- **pmid:** <https://pubmed.ncbi.nlm.nih.gov/29456894>
- **pmcid:** <https://www.ncbi.nlm.nih.gov/pmc/articles/5815332>

Figura 4.17: Captura de búsqueda de un DOI en OpenAlex por medio de su web en desarrollo.

Incorporación de Google Scholar en la búsqueda de artículos:

Google Scholar (GS) (sección [2.2.9](#)) 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 versión 2.0 de la herramienta.

A continuación, se describe en detalle cómo se abordan estos desafíos:

Dificultades en el Web Scraping de Google Scholar

Obstáculos técnicos

Google Scholar, como muchos otros productos de Google, implementa una serie de sofisticados mecanismos anti-rastreo para proteger sus datos y servicios. Estos mecanismos, diseñados para detectar y bloquear comportamientos de scraping automatizados, pueden identificar y restringir las direcciones IP que muestran patrones de acceso anormales o excesivamente frecuentes.

Además, GS no proporciona una API pública, lo que significa que cualquier intento de extracción de datos debe hacerse a través del análisis de las páginas web directamente. Esto presenta desafíos adicionales debido a la naturaleza dinámica y cambiante de la estructura del sitio web de GS. Cualquier cambio en esta estructura puede resultar en la necesidad de realizar ajustes significativos en el código de scraping, lo que aumenta el costo y el tiempo requeridos para mantener una operación de scraping efectiva.

Consideraciones legales y éticas

Más allá de los desafíos técnicos, la extracción de datos de GS también plantea serias preocupaciones legales y éticas. Los términos de servicio de Google prohíben explícitamente el scraping de sus servicios¹⁸. El incumplimiento de estos términos puede resultar en acciones legales y la terminación de los servicios de Google.

Además, Google tiene medidas en su lugar para detectar y prevenir el scraping. Por ejemplo, utiliza reCAPTCHA Enterprise, una herramienta que puede identificar bots que están trabajando en silencio en el fondo y dar herramientas y visibilidad para prevenir que esos bots accedan a contenido web valioso¹⁹.

Desde una perspectiva ética, el scraping indiscriminado de GS puede ser visto como una violación del principio de uso justo, especialmente si los datos recopilados se utilizan para fines comerciales o se redistribuyen sin permiso.

¹⁸ [Términos de servicio - Google](#)

¹⁹ [Mitigación del Web Scraping con reCAPTCHA Enterprise](#)

Viabilidad de la extracción de datos

Dado el alcance y la profundidad de la información disponible en GS, la extracción completa y efectiva de datos es una tarea formidable. Los millones de documentos académicos, cada uno con múltiples autores, citas, referencias y metadatos asociados, hacen que la tarea de extracción y normalización de datos sea una operación altamente compleja y que consume mucho tiempo.

Integración a HERA 2.0

A pesar de estos desafíos, se reconoció la importancia de GS para HERA 2.0. Por lo tanto, se buscó una alternativa de solución que pudiera integrar GS en la experiencia del usuario de HERA pero sin infringir sus términos de servicio. En consecuencia, 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 4.18).



Figura 4.18: Botón de Google Scholar en HERA 2.0

Esta solución permite a los usuarios de HERA beneficiarse de los datos disponibles en GS, sin infringir sus términos de servicio. Además, al proporcionar un enlace directo a GS, se mejora la usabilidad de HERA 2.0, ofreciendo a los usuarios un acceso directo y conveniente a la información académica relevante.

La Figura 4.19 muestra como el botón de GS redirige a los resultados en dicho sitio cuando se busca el DOI:10.7717/peerj.4375 en HERA 2.0, el cual corresponde al artículo: *The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles*.

The screenshot shows a Google Scholar search interface. The search bar contains the query "allintitle:The state of OA: a large-scale analysis of the prevalence and impact". The results page displays several entries for the article "The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles".

On the left side, there are filters for "Cualquier momento" (Any time), "Ordenar por relevancia" (Sort by relevance), "Cualquier idioma" (Any language), and "Cualquier tipo" (Any type). Under "Cualquier tipo", there are checkboxes for "incluir patentes" (include patents) and "incluir citas" (include citations), with "incluir citas" checked. There is also a "Crear alerta" (Create alert) button.

The search results list the following entries:

- [HTML] The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles** [HTML] peerj.com
 ..., JP Alperin, L Matthias, B Norlander, A Farley... - PeerJ, 2018 - peerj.com
 ... publisher website, without an explicit **Open** license. We also ... **the citation impact of OA articles**, corroborating **the so-called open-access** citation advantage: accounting for age and ...
 ☆ Guardar 🔄 Citar Citado por 1018 Artículos relacionados Las 26 versiones 🔄
- The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles**
 ..., L Matthias, B Norlander, A Farley... - ... Research and ..., 2020 - neiconjournal.com
 ... publisher website, without an explicit **Open** license. We also ... **the citation impact of OA articles**, corroborating **the so-called open-access** citation advantage: accounting for age and ...
 ☆ Guardar 🔄 Citar Citado por 3 Artículos relacionados 🔄
- [CITAS] The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles**
 B Grace - 2019 - peerj.com
 ... RT @CochraneUK: **'The State of OA'** estimates that 45% of papers in 2015 were **open access**. Pre-print via @thePeerJ https://t.co/Lc1OpJqL5W ... RT @oaDOI_org: Lots of #openaccess ...
 ☆ Guardar 🔄 Citar Citado por 10 Artículos relacionados Las 2 versiones 🔄
- [CITAS] The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles**
 V Marie-Pierre, KB Boughida - 2018 - peerj.com
 ... **Large-scale** study shows 28% of scholarly literature is already #OA, proportion of **OA articles** amongst accessed >45% https://t.co/ExhPmb57L9 ... Definitely worth a read ...
 ☆ Guardar 🔄 Citar Citado por 5 Artículos relacionados 🔄
- [CITAS] The state of OA: a large-scale analysis of the prevalence and impact of open access articles.** PeerJ
 ..., JP Alperin, L Matthias, B Norlander, A Farley... - 2018
 ☆ Guardar 🔄 Citar Citado por 45 Artículos relacionados
- [CITAS] Jevin West and Stefanie Haustein. The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles [Internet]. 2017**
 ..., JP Alperin, L Matthias, B Norlander, A Farley
 ☆ Guardar 🔄 Citar Citado por 9 Artículos relacionados

Figura 4.19: Citas en google scholar de *The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles*

Scopus:

Además, en el proceso de mejora de la búsqueda de DOI, se decidió incorporar Scopus (sección 2.2.1).

La API de Scopus suministra una imagen y un enlace, que redirecciona al usuario a la página de Scopus correspondiente al artículo buscado, permitiendo la visualización de sus citas. Pese a que la documentación de Scopus sugiere la inclusión de una clave API como un encabezado de solicitud *HTTP*, este método no resultó ser operativo en la práctica [69].

En su documentación, Scopus sugiere un enfoque alternativo que implica la incrustación (*embed*) del conteo de citas en el HTML. Sin embargo, este método requiere la exposición de la clave API en el cliente, lo cual se consideró inadecuado debido a las implicancias en materia de seguridad.

En base a lo anterior, se implementó una solución similar a la que se empleó con GS. Se añadió un botón que redirige al usuario a la página de Scopus (ver Figura 4.20), donde se visualizan las citas del artículo correspondiente (Figura 4.21).

Este ajuste facilita a los usuarios de HERA 2.0 el acceso a la información pertinente de citas en Scopus, agilizando aún más el proceso de búsqueda y evitando la necesidad de que el usuario tenga que realizar esta tarea de forma manual.



Figura 4.20: Botón de Scopus en HERA 2.0

513 documents have cited:

The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles
 Piwowar H., Priem J., Larivière V., Alperin J.P., Matthias L., Norlander B., Farley A., (...) Haustein S.
 (2018) *PeerJ*, 2018 (2), art. no. e4375

You are in Preview mode, only the first 20 documents are visible.

Search within results... Analyze search results Sort on: Date (newest)

Refine results: Limit to Exclude

Open Access

- All Open Access (348) >
- Gold (208) >
- Hybrid Gold (60) >
- Bronze (30) >
- Green (259) >
- View more

Document title	Authors	Year	Source	Cited by
1 CORE: A Global Aggregation Service for Open Access Papers	Kneth, P., Herrmannova, D., Cancellieri, M., (...), Gyawali, B., Pride, D.	2023	Scientific Data	0
2 OpCitation: Citation contexts identified from the PubMed Central open access articles	Hsiao, T.-K., Torvik, V.I.	2023	Scientific Data	0

Figura 4.21: Citas en Scopus de *The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles*

Actualización de la tabla de bases de datos para HERA 2.0:

La siguiente tabla muestra las bases de datos consultadas por HERA 2.0 y las métricas correspondientes que se obtienen en cada caso:

Base de Datos	Método de selección de contenido	Principales métricas e indicadores	Recurso
Crossref	Metadatos estandarizados por miembros de la plataforma.	<ul style="list-style-type: none"> • Número de citas • Estadísticas de papers por año y autores con ORCID • Información bibliográfica 	Artículo; Revista
Scimago Journal & Country Rank	Recursos contenidos en Scopus	<ul style="list-style-type: none"> • Posición en ranking global • Cuartil • H-index del journal • Información bibliográfica 	Revista
Scopus	Proceso de evaluación por junta asesora	<ul style="list-style-type: none"> • Número de citas • Predicción de citas • Factores de impacto 	Revista
Web of Science	Proceso de evaluación por editores expertos	<ul style="list-style-type: none"> • Índice de WoS • Información bibliográfica 	Revista
DOAJ	Proceso de evaluación de acceso abierto y calidad	<ul style="list-style-type: none"> • Licencias • Sello de DOAJ 	Artículo; Revista
Semantic Scholar	Indexación mediante algoritmos de machine learning ejecutados sobre bases de datos afiliadas	<ul style="list-style-type: none"> • Número de citas • Frecuencia de citas • Número de citas influyentes 	Artículo
Altmetric	Realiza cálculos reconociendo menciones en redes sociales, foros, etc. Mediante inteligencia artificial	<ul style="list-style-type: none"> • Puntuación de atención basada en menciones ponderadas según el sitio • Presencia de menciones en redes sociales 	Artículo
Google Scholar	Indexación por Google mediante algoritmos de machine learning	<ul style="list-style-type: none"> • Citas 	Artículo; Revista
Dimensiones	Recopilación de contenidos de bases abiertas (Crossref y PubMed)	<ul style="list-style-type: none"> • Número de citas • Proporción de citas por campo y relativa 	Artículo
OpenAlex*	Metadatos estandarizados por miembros de la plataforma. Recopilando información de múltiples bases, entre ellas: Crossref, MAG, ORCID, Unpaywall, etc	<ul style="list-style-type: none"> • Número de citas • Estadísticas de <i>papers</i> por año y autores con ORCID • Información bibliográfica • Evento/Venue • Sello de DOAJ 	Artículo; Revista
Red Iberoamericana (REDIB)*	Proceso de evaluación de calidad	<ul style="list-style-type: none"> • Posición de ranking • Información bibliográfica • Indicadores de calidad editorial y acreditaciones 	Revista

*Open Alex: Nueva fuente de información.

*REDIB: Temporalmente sin servicio.

Tabla 4.2: Resumen de bases de datos consultadas en HERA 2.0.

4.4 Resultados de: Refactorización de código

Introducción

El código de 1.0 HERA presenta varios desafíos que requieren atención. Entre los problemas más notables se encuentran la falta de modularización del código, dificultando su lectura, entendimiento, mantenimiento, haciendo tediosa la tarea de incorporación de nuevas funcionalidades de manera eficiente. Otro inconveniente se halla en el manejo de múltiples fuentes de datos y en la complejidad inherente a la recopilación y selección de información proveniente de diversas API y de técnicas de web scraping. Frente a estas dificultades, se decide emprender un proceso de refactorización del código con el objetivo de mejorar su estructura, eficiencia y mantenibilidad.

Problemas con el diseño original

Falta de modularización: la versión 1.0 de HERA carece de una estructura modular en su código, lo que dificulta su comprensión y mantenimiento. La mayoría del código está concentrado en pocos archivos con un elevado número de líneas de código. Esta falta de modularización afecta negativamente la escalabilidad de la aplicación, complicando la incorporación de nuevas funcionalidades y adaptaciones. Además, un código poco modularizado tiende a ser más propenso a errores, ya que las modificaciones en una parte del código pueden tener efectos imprevistos en otras partes de la aplicación.

Uno de los principales inconvenientes en HERA 1.0 se ilustra en la Figura 4.22 una secuencia de condicionales anidados *if* empleado para la recuperación del título durante la búsqueda de un recurso académico. Este código representa una lógica en cascada, donde se consulta a varias bases de datos, una tras otra, hasta obtener el título requerido. Si, por ejemplo, la base de datos crossref proporcionaba el título, el sistema lo recogería y cesaría en su búsqueda. Si no, pasaría a la base de datos DOAJ, y así sucesivamente a través de cada fuente de información externa consultada.

Es importante enfatizar el tamaño y la complejidad de este bloque condicional, que es solo uno de los muchos necesarios en HERA 1.0. Cada vez que una fuente de datos externa dejaba de funcionar como Microsoft Academic (detallado en la sección [2.2.10](#)), era necesario modificar el código fuente para eliminarla. Del mismo modo, la incorporación de nuevas fuentes de información requería la expansión de este ya voluminoso bloque condicional.

Además, esta estrategia de búsqueda condicional no se limitaba a la obtención del título. Se aplicaba también a la recuperación de otros elementos de datos como el resumen, el tipo de artículo, los autores, la fecha de publicación y la editorial, entre otros. Esta repetición de estructuras condicionales grandes y complejas añadía considerablemente a la dificultad de mantener y evolucionar el código.

Diagrama de flujo de la búsqueda del título con 'if' anidados en HERA 1.0

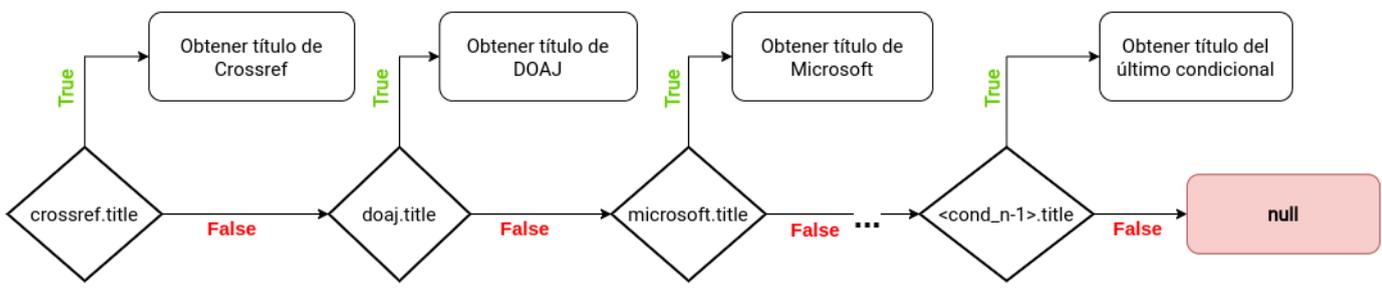


Figura 4.22: diagrama de flujo de la búsqueda condicional de if anidados en HERA 1.0

Manejo de múltiples fuentes de datos: HERA se apoya en la recolección de información de un gran número de API y técnicas de web scraping para construir un resultado. Sin embargo, la arquitectura original del código no es óptima para gestionar múltiples fuentes de datos. La información recolectada de cada fuente se guarda indistintamente, y para construir el contexto general de un DOI o ISSN, el código debe seleccionar los mejores datos de todas las fuentes disponibles. El código consulta por la existencia de ciertos datos (por ejemplo, el título de un artículo) en todas las fuentes, siguiendo un proceso iterativo que complica su mantenimiento y escalabilidad. Si una fuente de datos falla, la construcción del contexto se ve afectada, y agregar nuevas fuentes de datos implica un aumento en la complejidad y longitud del código. Este diseño no solo es ineficiente, sino que también compromete la calidad y fiabilidad de los resultados proporcionados por la aplicación.

En la Figura 4.23 se ilustra un escenario común que ocurre cuando se manejan múltiples fuentes de datos en la lógica de back-end de una aplicación como HERA. En el diagrama, las flechas del lado izquierdo representan solicitudes HTTP fetch realizadas desde la lógica de *back-end* hacia varias fuentes de datos externas. La figura también ilustra que el cliente ejecuta siete solicitudes, cada una correspondiente a un recurso distinto, en lugar de efectuar una única petición al back-end para obtener los siete recursos necesarios en el ejemplo dado. Esta desviación en la lógica de negocio se abordará en la sección subsiguiente.

En la representación gráfica, una de las flechas que apuntan a la fuente de datos Microsoft Academic se muestra en rojo, indicando que ha ocurrido un error. Esto puede ser el resultado de múltiples circunstancias, por ejemplo, la fuente de datos puede estar temporalmente inaccesible debido a problemas de red, o puede haber un cambio en la estructura de los datos proporcionados por Microsoft Academic que no fue previsto en el diseño del código de back-end.

A raíz de este error, se observa que la respuesta del back-end al cliente también falla, lo que se representa en el diagrama con una flecha en rojo direccionada hacia

el cliente. Esto resalta que, bajo el diseño original de HERA 1.0, la falla en el acceso a una sola fuente de datos puede provocar una falla completa del sistema, lo cual pone en peligro la calidad y la fiabilidad de los resultados que la aplicación puede proporcionar.

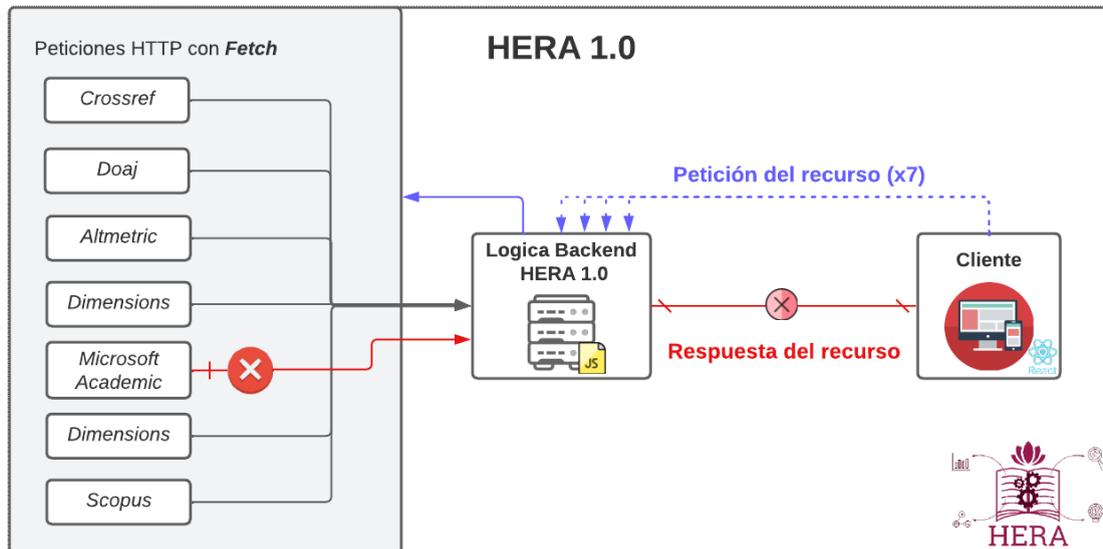


Figura 4.23: Manejo de múltiples fuentes de datos en HERA 1.0

Lógica de negocio: una problemática adicional con la arquitectura de la herramienta HERA en su versión 1.0 es la distribución inadecuada de las responsabilidades entre el front-end y el back-end. El front-end, desarrollado con React, maneja una cantidad considerable de lógica de negocio, incluyendo múltiples solicitudes fetch a API externas y tareas de web scraping. Este diseño violaba los principios fundamentales de la arquitectura cliente-servidor y presentaba varias dificultades.

En primer lugar, realizar tareas intensivas de red y CPU en el front-end, como por ejemplo las solicitudes *fetch* y el web scraping, puede tener un impacto significativo en el rendimiento global. Dichas operaciones pueden ralentizar la respuesta de la interfaz, afectando negativamente la experiencia del usuario.

En segundo lugar, el front-end se encarga no solo de la recuperación de los datos, sino también de la construcción y formato de la respuesta a ser mostrada. Esta responsabilidad incluye tareas como reorganizar los datos, descartar información no utilizada, formatear la información para su presentación y seleccionar los datos más relevantes para la visualización.

Además, este diseño compromete la escalabilidad de la aplicación. A medida que la aplicación crece y la cantidad de solicitudes fetch y operaciones de web scraping se incrementa, manejar estas tareas en el front-end se vuelve cada vez menos factible. Este problema de escalabilidad, junto con la complejidad de mantenimiento debido a la mezcla de lógica de presentación y lógica de negocio en el front-end, subrayaba la necesidad de una reestructuración de las responsabilidades del código.

Por otro lado, el back-end, implementado con Node.js y Express se subutiliza, ya que su única responsabilidad termina siendo realizar solicitudes fetch y enviar las respuestas al front-end. Este enfoque desaprovecha las capacidades de Node.js y Express, que son muy adecuados para manejar las operaciones intensivas de red y CPU como las solicitudes fetch y el web scraping.

La necesidad de una refactorización del código se hace más evidente al tener en cuenta estas dificultades. Para mejorar el rendimiento y la escalabilidad de la aplicación, así como facilitar su mantenimiento, es crucial reasignar las responsabilidades de manera que el back-end maneje la lógica de negocio y la gestión de los datos, y el front-end se centre en la presentación de los datos y la interacción con el usuario.

La Figura 4.24 ilustra la distribución desequilibrada de la carga de trabajo entre el front-end y el back-end en el diseño de HERA 1.0. El front-end, indicado por el círculo de la izquierda, muestra una cantidad considerable de flechas que representan las solicitudes fetch y las tareas de web scraping. Este gráfico sugiere que el front-end asume una cantidad excesiva de la lógica de negocio y las operaciones de red y CPU, representado por el termómetro con alta carga.

Por otro lado, el back-end, representado por el círculo de la derecha, tiene pocas flechas que indican su única responsabilidad: realizar solicitudes fetch y transmitir las respuestas al front-end. Un termómetro con poca carga visualiza la subutilización de esta sección, destacando un problema en la distribución de la lógica de negocio.

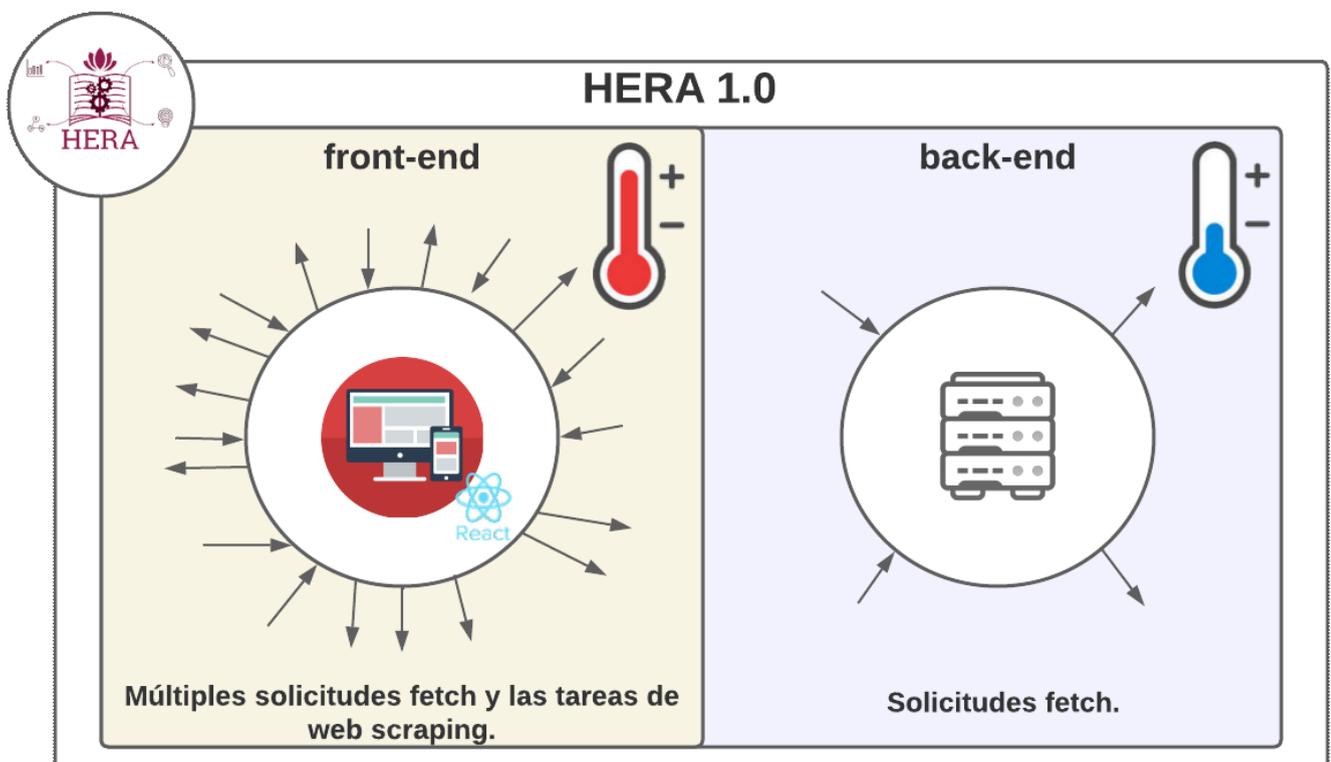


Figura 4.24: Representación de la lógica de negocio en HERA 1.0

A continuación, se detallan las principales mejoras implementadas.

Solución implementada:

Para abordar los desafíos identificados en la arquitectura inicial de la herramienta HERA, se implementó una reestructuración significativa del código. El primer paso fue trasladar la lógica de negocio del front-end al back-end, siguiendo los principios fundamentales de la arquitectura cliente-servidor [70]. Este cambio permitió mejorar el rendimiento, la seguridad y la escalabilidad de la aplicación, y también simplificó su mantenimiento al separar claramente la lógica de presentación y la lógica de negocio.

En términos de seguridad, la reestructuración mencionada tuvo implicaciones positivas ya que al negociar con API keys u otros datos sensibles desde el servidor, se evita la exposición de estos detalles en los clientes. Esto refuerza la seguridad y la privacidad de los datos, mitigando los riesgos asociados a las filtraciones de información.

Adicionalmente, el traslado de la lógica de negocio al servidor facilitó la centralización de las consultas, permitiendo un seguimiento más preciso del uso o carga en los diferentes servicios. Este control más detallado posibilita la regulación de las consultas, lo cual es crucial para cumplir con las cuotas establecidas y evitar el sobrecargo de los servicios.

A largo plazo, esta centralización también facilita la recopilación de datos útiles para la monitorización del sistema, como los tiempos de respuesta de cada servicio, la medición del porcentaje de tiempo de inactividad y la identificación de servicios que presentan interrupciones permanentes. Este tipo de análisis puede ser valioso para identificar oportunidades de mejora y optimizar el rendimiento general de la aplicación.

En la Figura 4.25, se ilustra la reconfiguración de la lógica de negocios en la cual se delega al backend la responsabilidad de gestionar todas las peticiones a las fuentes de datos externas. Esta modificación permite que el cliente realice una única petición al backend, optimizando así la interacción y mejorando la eficiencia de la comunicación. Esta estructura proporciona una mayor resistencia y adaptabilidad al sistema, ya que si una fuente de datos externa falla, la funcionalidad general no se ve afectada.

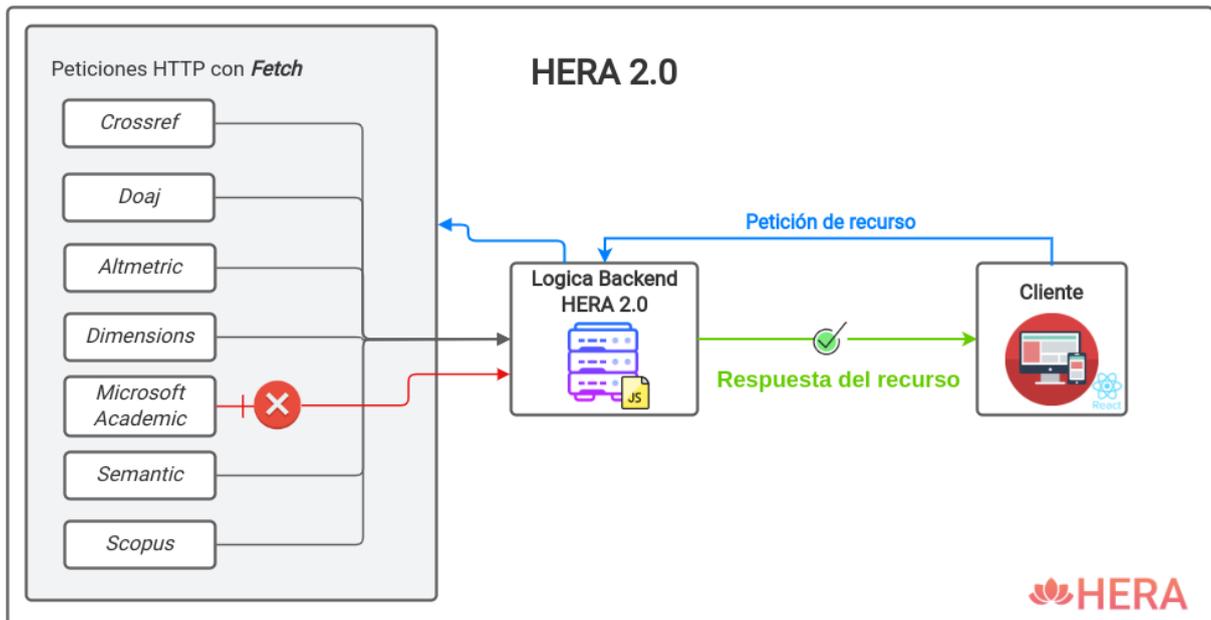


Figura 4.25: Manejo de múltiples fuentes de datos en HERA 2.0

En la Figura 4.26, se presenta el núcleo de la lógica de backend de HERA 2.0. Cuando se recibe una petición del cliente, el primer paso es la ejecución del gestor de fuentes de información, como se explica en la sección 4.5. Este gestor es responsable de determinar y transmitir al gestor de endpoints las fuentes de datos externas que serán consultadas (ver sección 4.7). Tras obtener las respuestas de manera concurrente, estas se envían al contexto de datos (Ver figura 4.27), cuyo diseño y funcionalidad se detallan en la siguiente sección de [estructura de datos](#). Cabe destacar que cada conjunto de datos cuenta con un manejador de errores, que se activará en caso de ser necesario. Finalmente, se genera la respuesta que será enviada al cliente. Este proceso de contextualización se explica en detalle en la sección: [creación del contexto](#).

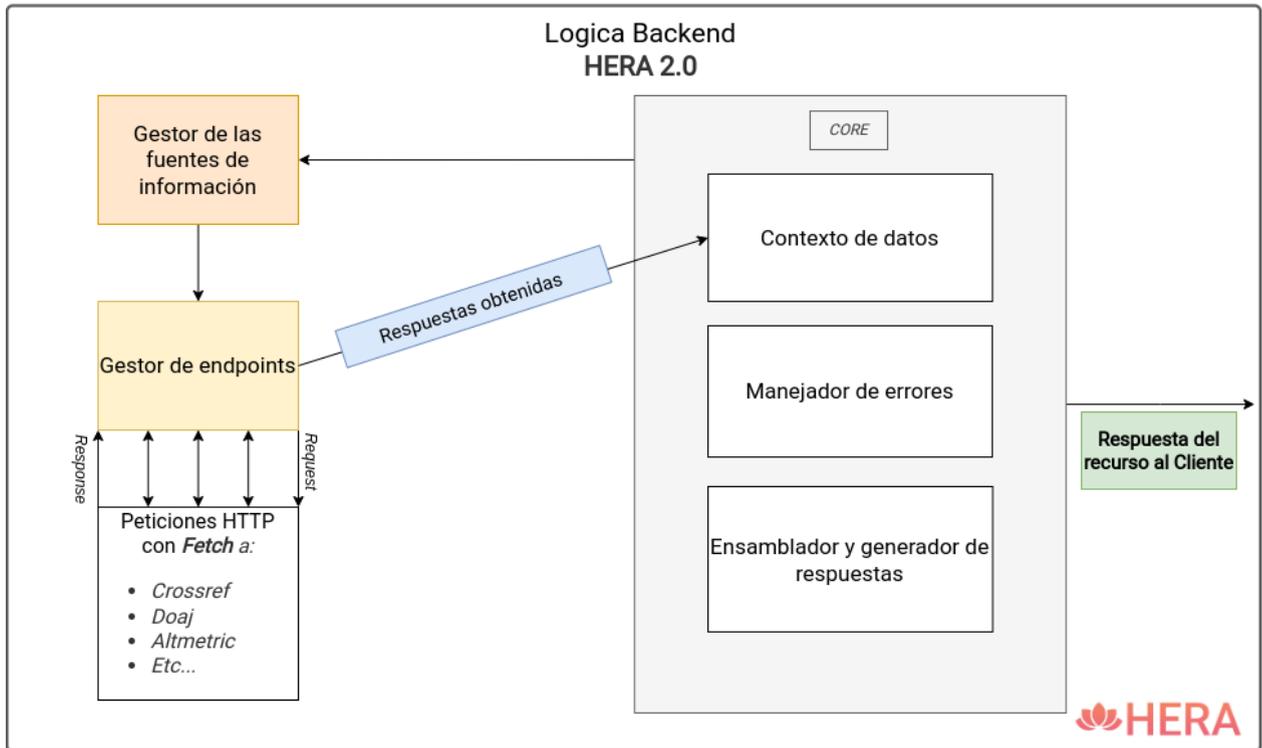


Figura 4.26: Lógica backend de Hera 2.0

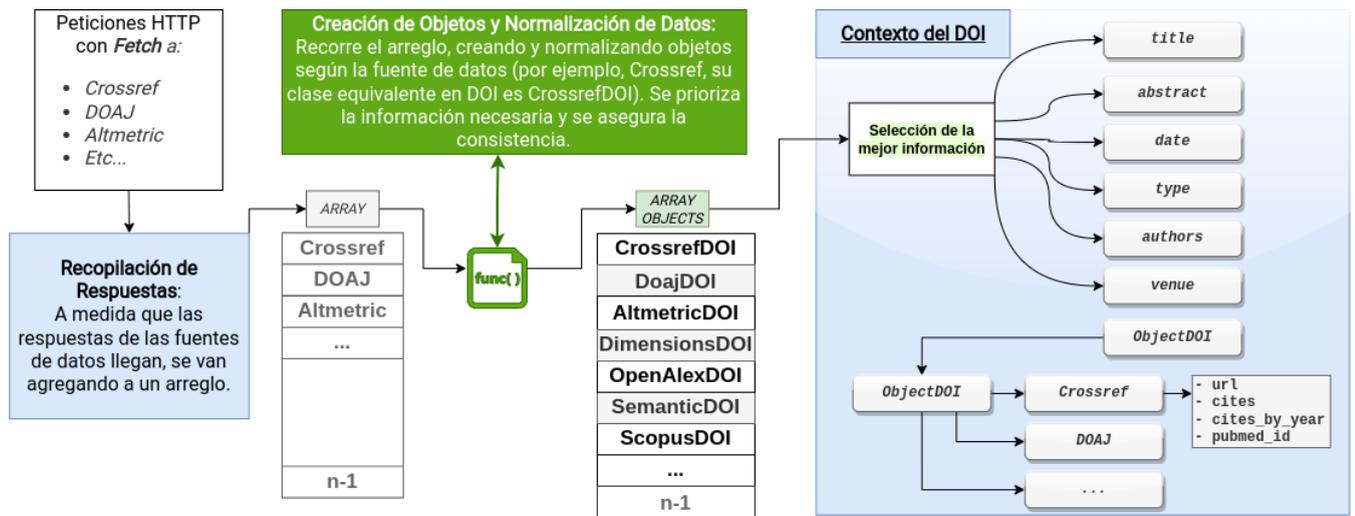


Figura 4.27: Lógica backend de Hera 2.0 en detalle

Por último, se decidió migrar el código de JavaScript a TypeScript. TypeScript es un superconjunto de JavaScript, se diferencia principalmente por su tipado estático que permite detectar errores de tipo en tiempo de compilación y por su enriquecida programación orientada a objetos con características como interfaces y tipos genéricos [71]. Aunque JavaScript es flexible y poderoso, carece de ciertas características que podrían ayudar a escalar proyectos más grandes de manera eficiente, algo que TypeScript sí proporciona.

Algunas de sus principales características son:

- **Tipado Estático:** la introducción de tipado estático significa que se asignan tipos de datos a todas las variables y objetos, lo que permite detectar y prevenir muchos errores en tiempo de compilación en lugar de hacerlo en ejecución. Este cambio facilitó la tarea de depuración y ayudó a prevenir la introducción de nuevos bugs.
- **Legibilidad y Estructura del Código:** TypeScript añade características de programación orientada a objetos, como clases e interfaces, que facilitan la organización del código. Esto fue de particular importancia para este proyecto, donde la complejidad y la necesidad de interactuar con varias bases de datos requerían una buena estructura de código.
- **Desarrollo más Ágil:** la transición a TypeScript también ayudó a agilizar el proceso de desarrollo. Gracias a las características de tipado estático y la detección anticipada de errores, se redujo la cantidad de tiempo dedicada a la depuración y prueba del código.

Estructura de datos

Una vez migrado el back-end a TypeScript, el siguiente paso fue descentralizar el código responsable del armado de la respuesta. Para ello, se creó una clase para cada fuente de información (OpenAlex, DOAJ, Crossref, WoS, Scimago, Altmetric, Dimensions, Elsevier, etc. descritas en el [Capítulo 2](#) o en la [tabla de bases de datos para HERA 2.0](#)), en cuanto al recurso DOI, en la Figura 4.28, se puede apreciar que la interfaz *ObjectDOI* interactúa con casi todas las clases, con la excepción de Dimensions y Scopus. Esto se debe a que estas dos clases poseen características distintivas, ya que implementan widgets personalizados propios de cada fuente de datos. De esta manera, su estructura y naturaleza de datos divergen de las demás clases, lo que justifica su aislamiento de la interconexión con *ObjectDOI*.

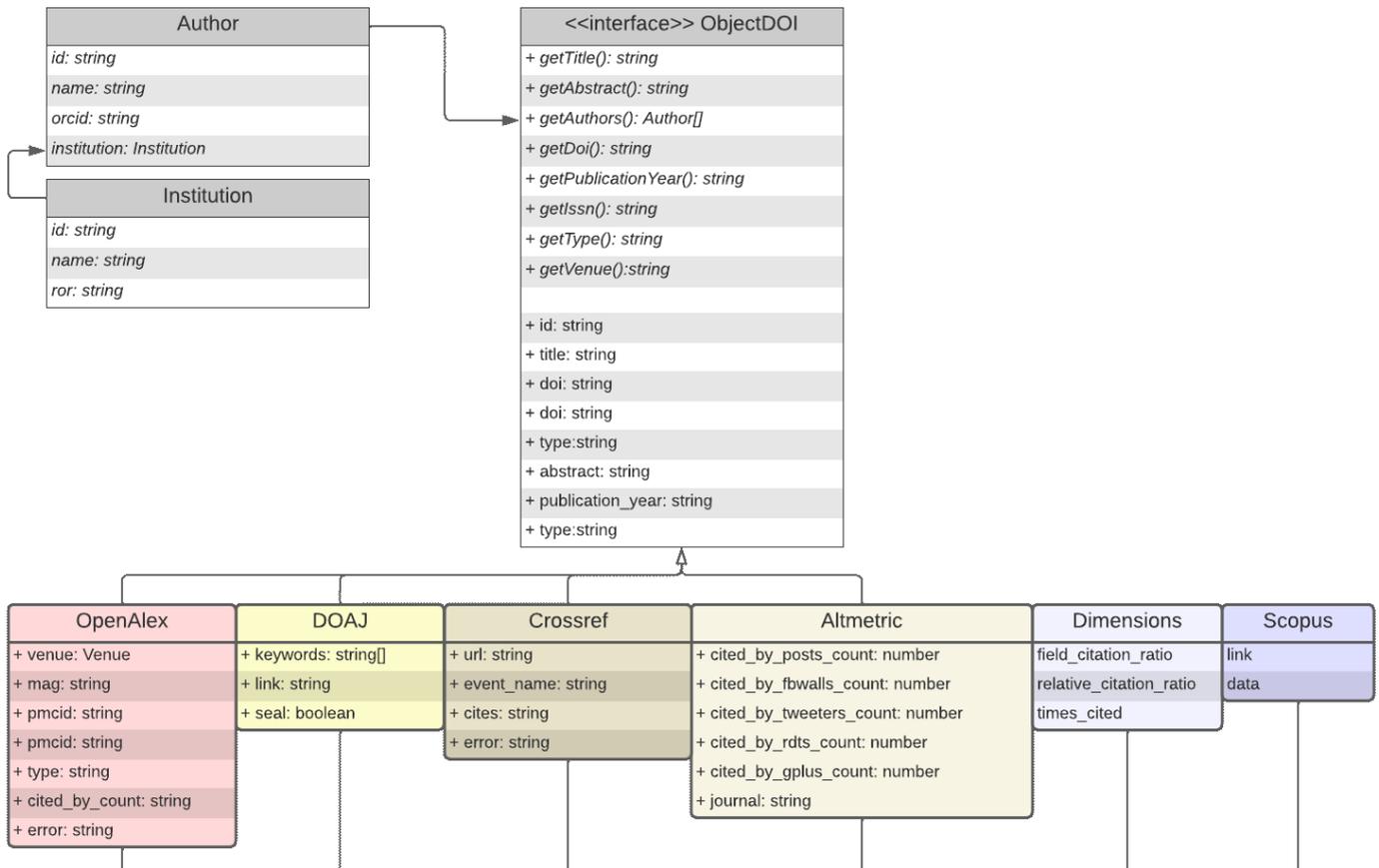


Figura 4.28: Interfaz y las clases para el recurso DOI

Por otro lado, en el caso del recurso ISSN, como lo ilustra la Figura 4.29 todas las fuentes de datos emplean la interfaz ObjectISSN.

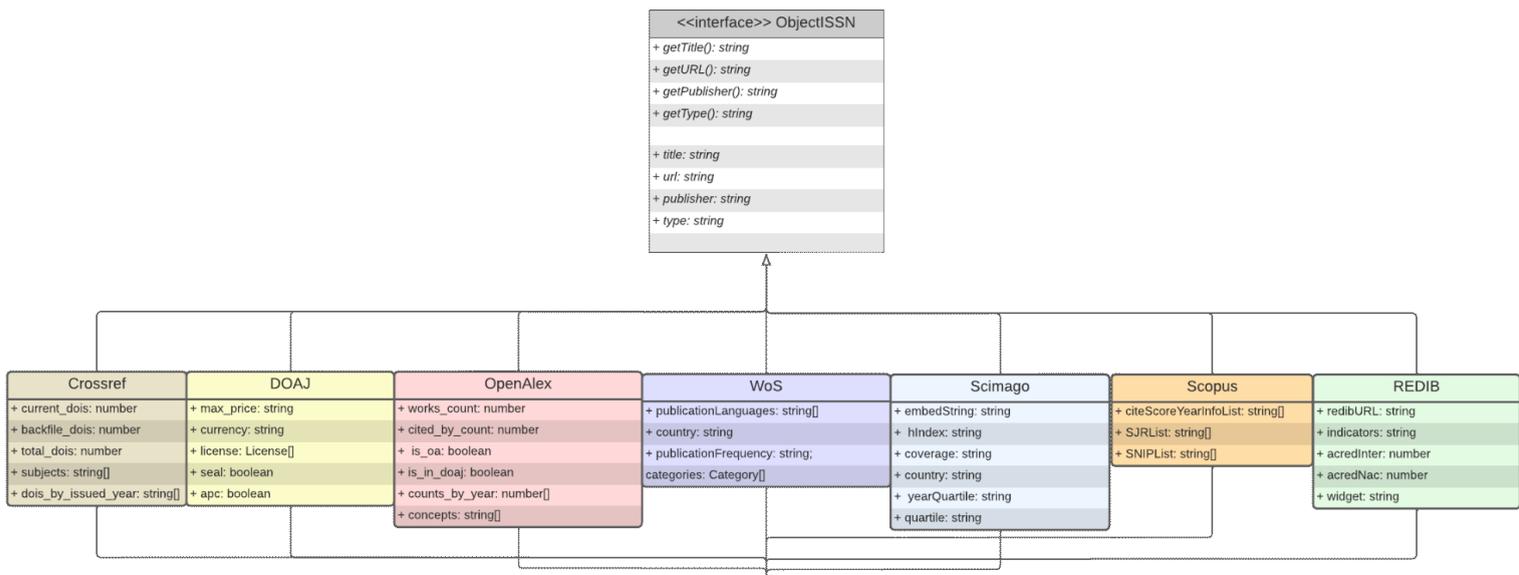


Figura 4.29: Interfaz y las clases para el recurso ISSN

Este enfoque ha mejorado la interacción con las API externas al proporcionar datos tipificados, lo que acelera el proceso de codificación. Con esta solución, los valores

se encuentran en cada clase y se pueden formatear según sea necesario. Además, se descarta cualquier información que no sea útil para HERA, aligerando así la carga de las solicitudes. Se ha estandarizado también la gestión de errores y la gestión de casos en los que el recurso solicitado no se encuentra, mejorando así la robustez y la fiabilidad del sistema.

Dado que las diferentes fuentes de información proporcionan datos de los autores en diferentes formatos, se decidió crear una clase de Autor común para normalizar la forma en que se manejan estos datos. Esta decisión sigue el principio de la programación orientada a objetos de la "composición sobre la herencia", permitiendo un mayor grado de flexibilidad y reutilización del código [72].

Estas decisiones de diseño se basan en el patrón de diseño *Strategy* [72] (ver la Figura 4.30), que promueve la definición de una familia de algoritmos, encapsulando cada uno de ellos y haciéndolos intercambiables. En este caso, cada fuente de información puede ser vista como una estrategia diferente para obtener los datos requeridos, y la aplicación puede seleccionar y utilizar la estrategia que sea más adecuada en cada caso.

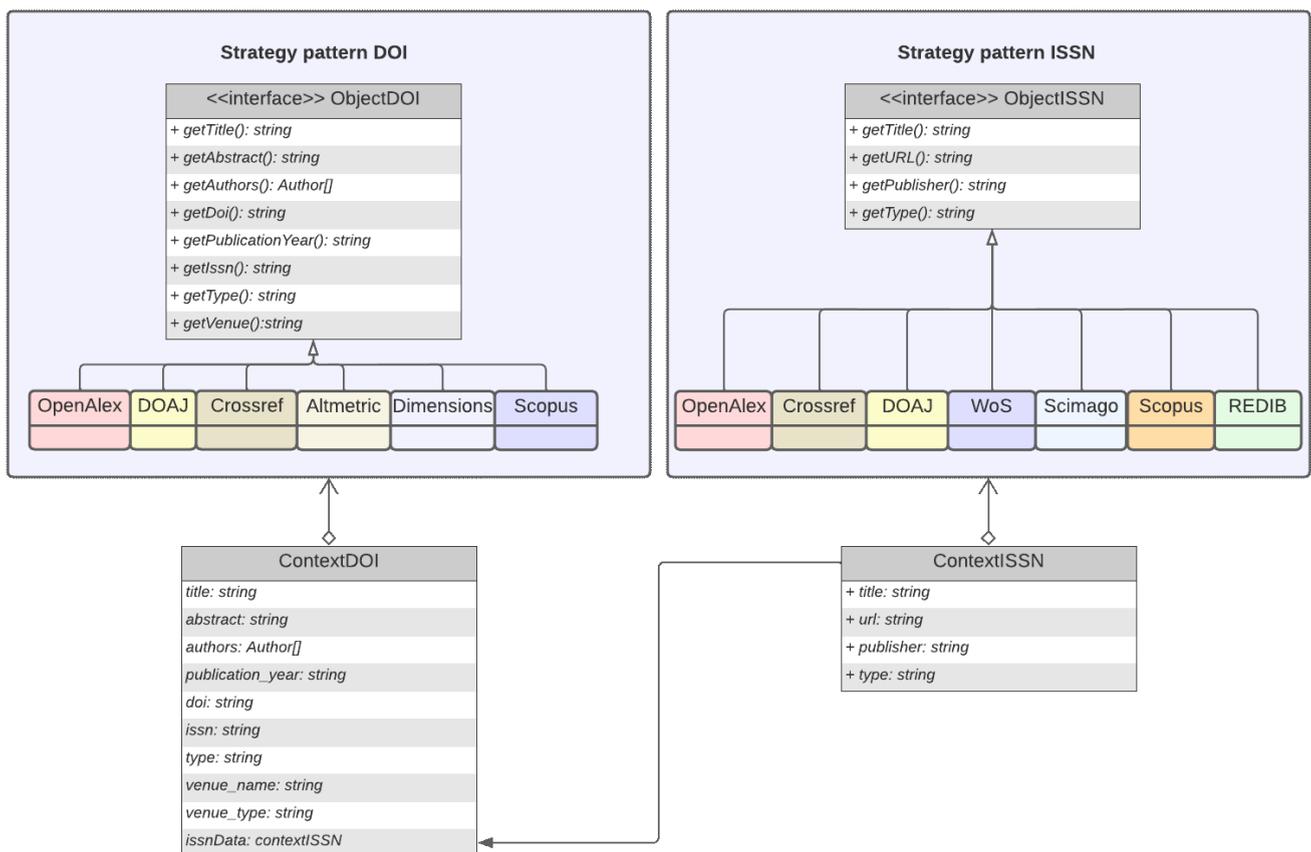


Figura 4.30: Diseño de patrón *strategy* en HERA 2.0

Creación del Contexto

Para mejorar la gestión de datos obtenidos de varias fuentes de datos y optimizar su manipulación, se introdujo el concepto de *contexto*. Este nuevo enfoque se basó en la creación de una estructura de objeto que reúne toda la información recopilada sobre cada DOI e ISSN.

1. **Estructura del contexto:** cada *contexto* es una estructura de objeto que incluye campos para almacenar información general del *paper* o *journal*, como el título, el resumen, los autores, y demás datos de interés. A su vez, este objeto conserva en su interior la información recolectada de cada fuente de datos en particular. Esto incluye los datos “crudos” de la consulta a cada fuente de datos, lo que facilita la comparación y selección de la mejor información disponible.
2. **Simplificación del manejo de datos:** en HERA 1.0 era necesario consultar individualmente cada fuente de datos y verificar la existencia de información relevante, lo que resultaba en una manipulación de datos compleja y propensa a errores. Con la introducción del *contexto*, esta tarea se simplifica significativamente. Ahora, una vez creado el *contexto* para un determinado DOI o ISSN, este objeto centraliza y maneja toda la información relevante.
3. **Mejora en la eficiencia y fiabilidad del código:** al centralizar el manejo de los datos en el *contexto*, se logró una mayor eficiencia en la gestión de la información. Esto redujo la complejidad del código y mejoró su robustez, ya que se minimizó el riesgo de errores causados por la manipulación incorrecta de los datos.

Refactorización de código y optimización de Web Scraping

Se ha llevado a cabo una importante refactorización de las funciones de web scraping en HERA 2.0.

El web *scraping* es una técnica que permite extraer información de sitios web. En HERA, estas funciones se utilizan para recoger datos y métricas de las diversas fuentes académicas en las que se basa la plataforma. La refactorización de estas funciones ha implicado varias mejoras significativas [73].

Además, se ha hecho uso de las características de Node.js y jsdom para mejorar el rendimiento de las funciones de web scraping. Node.js es un entorno de ejecución de JavaScript que permite ejecutar el código en el servidor, lo cual es particularmente útil para tareas como el web scraping que requieren una comunicación intensiva con el servidor. Por su parte, jsdom [74] es una implementación de JavaScript del modelo de objetos de documento (DOM) que permite manipular y navegar por el contenido de las páginas web como si se estuviera trabajando en un navegador. Con jsdom, es posible extraer datos de las páginas web de forma más eficiente y precisa.

Web Scraping con REDIB y SJR:

La refactorización del código para las operaciones de web *scraping* en HERA involucró una consideración cuidadosa de las especificidades de cada fuente de datos. En el caso de REDIB y SJR, se encontraron desafíos particulares que requerían soluciones a medida.

Para REDIB, la ausencia de una API accesible hizo necesario el uso de web scraping para acceder a sus datos. Se descubrió que el scraping directo del Ranking REDIB podría causar pérdida de información, dado que el ranking se basa en métricas de Clarivate y sólo incluye revistas previamente indexadas por esta entidad. En su lugar, se decidió realizar búsquedas del recurso y su página individual a través de solicitudes *HTTP GET*.

En el proceso de recopilación de datos de REDIB, se extrajo información sobre indicadores de calidad editorial, acreditaciones y, cuando fue posible, el widget exportable del Ranking REDIB. Se optó por no extraer información adicional como gráficos estadísticos, con el objetivo de mostrar la indexación en REDIB como una "métrica" adicional para el investigador, más que replicar la funcionalidad de la web de REDIB. Las pocas métricas extraídas tienen un carácter simbólico y ejemplificador, con el fin de captar la atención del investigador y ofrecer el enlace para acceder a la información completa en el sitio original.

Por otro lado, el caso de SJR también presentó desafíos debido a la falta de una API REST u otro método conveniente para extraer su información. Sin embargo, la organización simple del código HTML presentado en el sitio web del SJR permitió el uso de web scraping para obtener algunas métricas, como el índice H, la cobertura, el país y el widget exportable. Estas métricas se incorporaron a HERA junto con la información extraída de Scopus para facilitar su análisis.

En el caso de SJR, se decidió no incluir la información de los gráficos de citas y documentos proporcionados por SCImago, ya que se consideró que su valor pertenece a dicha plataforma. En su lugar, se optó por presentar información suficiente para captar la atención de un investigador, permitiéndole decidir si desea visitar el sitio web original para analizar los datos de manera más detallada.

4.5 Resultados de: Gestión de las fuentes de información

Configuración en 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 en formato JSON que almacena información detallada de cada fuente de datos externa. Este archivo incluye elementos como el identificador único de la 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 4.3, se presenta un ejemplo conciso 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. Es importante destacar que esta tabla representa una versión simplificada del archivo de configuración y, por lo tanto, puede no incluir todos los elementos presentes en el archivo de configuración completo.

Fuente de datos	URL	<i>enabled</i>
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 4.3: 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.

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.

Para administrar esta estructura de configuración, se implementó un gestor de configuración. Este componente lee la configuración, identifica las fuentes de datos habilitadas y crea instancias de las clases correspondientes. Posteriormente, realiza solicitudes fetch a cada fuente de datos como lo ilustra la Figura 4.31 y utiliza la respuesta para inicializar los objetos de las clases respectivas utilizando sus constructores. Los objetos resultantes se almacenan en una cola y luego se pasan

al contexto, que es responsable de formatear y combinar los datos de manera coherente y eficiente.

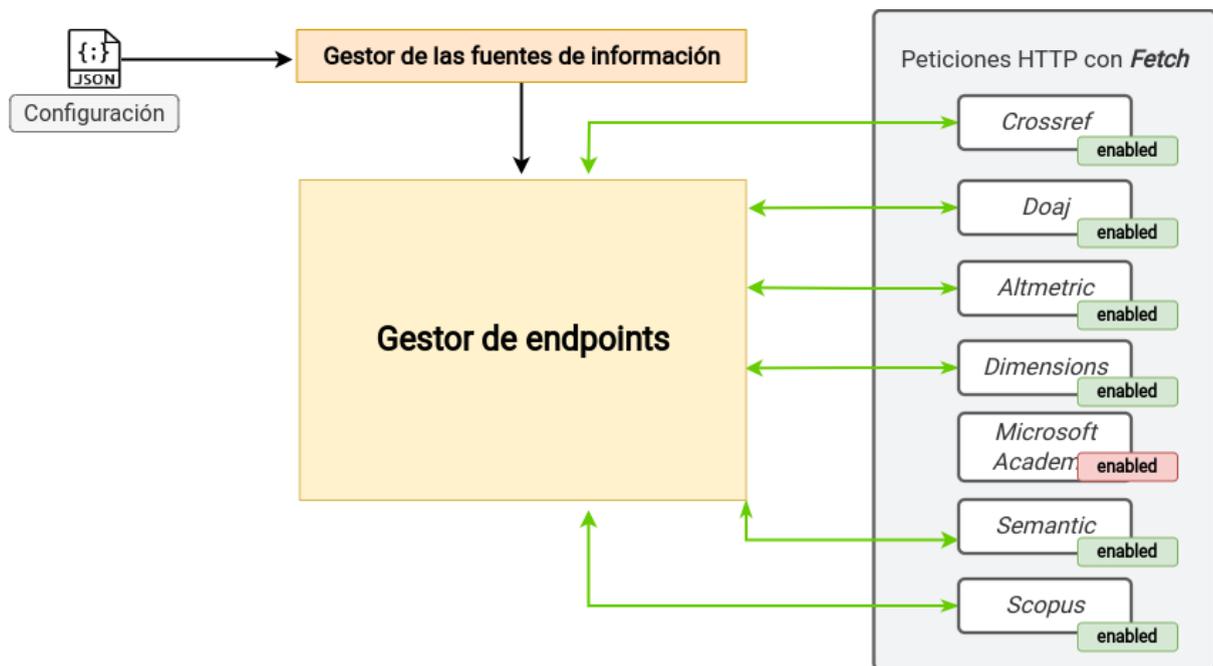


Figura 4.31: Diagrama del ciclo de ejecución de la configuración de las fuentes de información externas. Se muestra la interacción entre el archivo de configuración y el gestor de endpoints, ilustrando el flujo de procesamiento y gestión de las peticiones de datos.

Selección de la mejor información

La selección de la "mejor" información entre las múltiples fuentes de datos es una tarea crítica que se realiza en el proceso de completar el contexto. El término "mejor" se refiere a la información que es más completa y precisa entre todas las fuentes disponibles y su selección está dirigida por un conjunto de prioridades definidas en un archivo de configuración.

El archivo de configuración se diseña de manera que las fuentes de datos están ordenadas según la calidad y la completitud de la información que proporcionan. De esta manera, las fuentes con información más completa y precisa se sitúan en la parte superior de la lista de prioridades. Durante el proceso de rellenado del contexto, las fuentes de datos se consultan en el orden definido por esta lista de prioridades.

Este orden de prioridad no se estableció arbitrariamente; en cambio, se realizó a través de una serie de pruebas empíricas para determinar la confiabilidad y la integridad (tanto en términos de calidad como de cantidad de datos) de cada fuente de datos. Este enfoque permite que HERA 2.0 sea más adaptable y proporcione resultados más precisos y fiables.

Como lo ilustra la Figura 4.32, la lógica detrás de este enfoque es que, al consultar primero las fuentes de datos de mayor prioridad, se maximiza la posibilidad de obtener la mejor información disponible desde el principio. Cuando se consulta una

fuente de datos, la información obtenida se utiliza para rellenar el *contexto*. Si la misma información está disponible en una fuente de datos de menor prioridad consultada posteriormente, no se reemplaza la información existente en el contexto, ya que se asume que la información de la fuente de mayor prioridad es más precisa o completa.

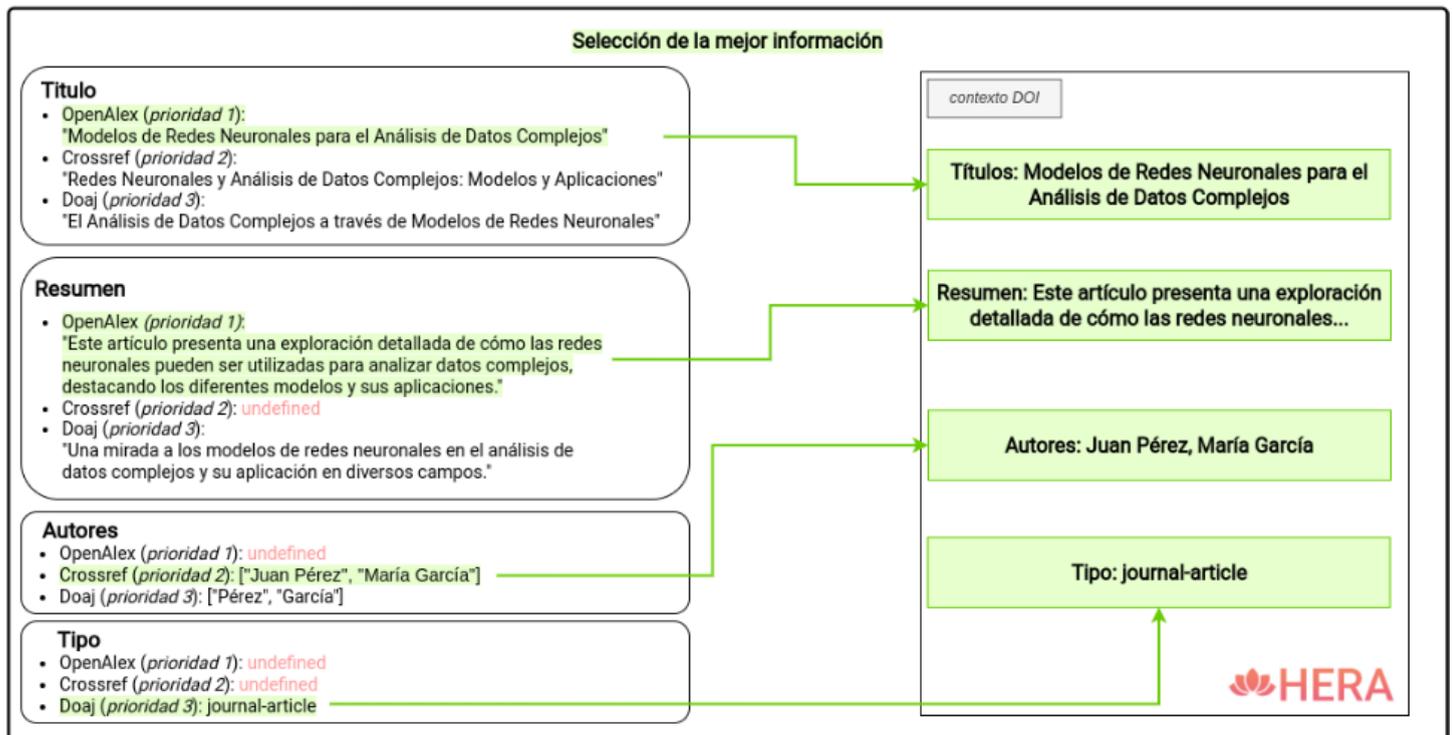


Figura 4.32: Imagen ilustrativa de cómo se elige la información según la prioridad de cada fuente de datos.

Esta estrategia resulta en la selección y almacenamiento de la información más completa y precisa posible dentro del *contexto*. Además, permite una cierta flexibilidad, ya que el orden de prioridad se puede ajustar fácilmente modificando el archivo de configuración en función de la calidad y la completitud de la información proporcionada por las fuentes de datos a lo largo del tiempo. Esta capacidad de adaptación es esencial en el contexto dinámico y en constante evolución de las bases de datos bibliográficas.

4.6 Resultados de: 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 4.33), 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 *stateless*²⁰, 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.

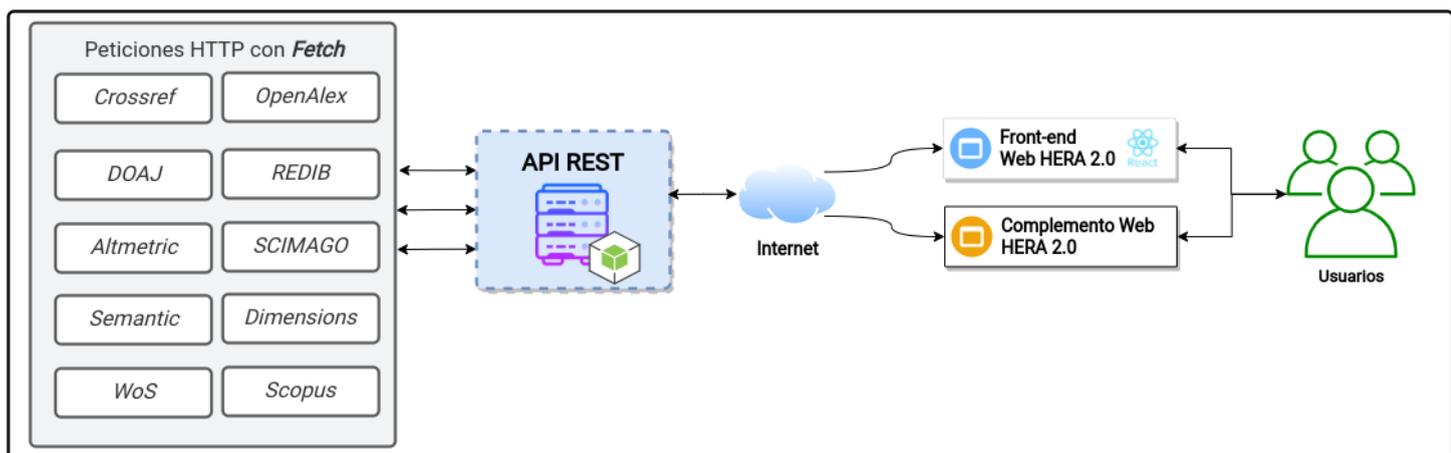


Figura 4.33: Esquema de comunicación de la API REST

²⁰ Protocolo sin estado.

4.7 Resultados de: Búsqueda simultánea de múltiples recursos

El traslado de la lógica de negocio del *front-end* al *back-end* en HERA se considera un avance en términos de rendimiento y tiempo de respuesta. 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, al efectuarse en el front-end, se realizaban de manera secuencial, lo que implicaba un consumo considerable de tiempo.

En HERA 2.0, estas peticiones se gestionan mediante *Node Express* en el back-end, permitiendo su ejecución concurrente (o 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. 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.

Análisis:

Antes de los resultados obtenidos, es esencial detallar las condiciones y parámetros bajo los cuales se realizó la prueba comparativa.

- La máquina de prueba cuenta con un procesador Intel Core i7-6700K de 4 núcleos, una memoria RAM de 16 GB, y sistema operativo Ubuntu Linux LTS 22.04.
- Como parte de los parámetros de la prueba, se emplearon el DOI 10.7717/peerj.4375 y el ISSN 2167-8359. Estos recursos se seleccionaron por su representatividad y la disponibilidad de información relevante a través de varias fuentes de datos.

Con búsqueda secuencial

Tiempo de consulta para (DOI)	Tiempo en ms
OpenAlexDOI	570
DoajDOI	324
Dimensions	58
Altmetric	291
SemanticScholar	1955
Total	3203

Tabla 4.4: Datos de búsqueda simultánea secuencial de un DOI

Tiempo de consulta para (ISSN)	Tiempo en ms
OpenAlexISSN	565
DoajISSN	324
ScopusISSN	321
ScimagoISSN	4070
RedibISSN	9560
WosISSN	1097
Total	15937

Tabla 4.5: Datos de búsqueda simultánea secuencial de un ISSN

Con búsqueda simultánea de múltiples recursos

Tiempo de consulta para (DOI)	Tiempo en ms
Dimensions	157
DoajDOI	305
Altmetric	328
OpenAlexDOI	588
SemanticScholar	2182
Total	2187

Tabla 4.6: Datos de búsqueda simultánea concurrente de un DOI

Tiempo de consulta para (ISSN)	Tiempo en ms
DoajISSN	315
ScopusISSN	335
OpenAlexISSN	595
WosISSN	893
ScimagoISSN	4549
RedibISSN	10013
Total	10016

Tabla 4.7: Datos de búsqueda simultánea concurrente de un ISSN

El análisis comparativo de los tiempos de consulta para el DOI 10.7717/peerj.4375 y el ISSN 2167-8359 bajo dos enfoques, secuencial y concurrente, demuestra la superioridad de este último en términos de eficiencia.

En el método secuencial, el tiempo total de consulta para el DOI fue de 3203 ms, mientras que para el ISSN fue considerablemente más alto, 15937 ms. Estos tiempos son la suma de las consultas individuales realizadas a cada API externa: OpenAlex, DOAJ, Dimensions, Altmetric, Semantic Scholar para el DOI; y OpenAlex, DOAJ, Scopus, Scimago, Redib y WoS para el ISSN.

Sin embargo, al utilizar el método paralelo, se observa una reducción significativa en los tiempos totales de consulta. Para el DOI, el tiempo total se reduce a 2187 ms, mientras que para el ISSN disminuye a 10016 ms. Esto evidencia la eficacia del método concurrente para optimizar las consultas a las API externas, ya que las peticiones no tienen que esperar a que se complete la anterior para iniciar.

Es importante destacar que en la consulta concurrente, el tiempo total de consulta no es la suma de los tiempos de cada consulta individual, sino que se corresponde con el tiempo de la consulta más larga. Esto se debe a que todas las consultas se ejecutan simultáneamente, y el tiempo total es el necesario para que todas finalicen.

Además de la implementación de consultas concurrentes, se adoptó una estrategia de optimización adicional al reordenar las consultas según el tiempo promedio de respuesta. Las consultas que, en promedio, tardan más tiempo en responder son las primeras en ser solicitadas. Esto permite maximizar la eficiencia del método concurrente, ya que las consultas más largas se inician primero y se ejecutan simultáneamente con las más rápidas.

Este ajuste optimiza aún más el tiempo total de consulta, ya que las consultas más cortas pueden completarse mientras las más largas aún están en proceso, lo que

reduce aún más el tiempo de espera general. En esencia, esto significa que el tiempo total de consulta se acerca aún más al tiempo de la consulta más larga, en lugar de la suma total de todas las consultas.

Con búsqueda simultánea de múltiples recursos (ajustada)

Tiempo de consulta para (DOI/ISSN)	Tiempo en ms
REDIBISSN	9955
ScimagoISSN	4498
SemanticScholar (DOI)	2150
OpenAlexISSN	580
OpenAlexDOI	578
Altmetric (DOI)	320
ScopusISSN	330
DoajISSN	310
DoajDOI	300
Dimensions (DOI)	150
Total (DOI)	2153
Total (ISSN)	9958

Tabla 4.8: Datos de búsqueda simultánea concurrente c/ajuste de un DOI e ISSN

Comparando los resultados anteriores con los originales, se observa una disminución notable en los tiempos de consulta totales para DOI e ISSN, lo que demuestra la eficacia de este método de optimización. El tiempo total de consulta para DOI disminuyó de 2187 ms a 2153 ms, y el tiempo total de consulta para ISSN disminuyó de 10016 ms a 9958 ms.

Teniendo en cuenta esta nueva condición, la consulta a REDIB fue configurada con un tiempo límite (*timeout*) de 3 segundos debido a su inactividad temporal. Con la limitación de tiempo aplicada a REDIB, la mejora en el tiempo de consulta total para ISSN es drástica. El tiempo total de consulta para ISSN se redujo de 9958 ms en el caso anterior a **4498 ms** en el actual, mostrando una mejora significativa en la eficiencia global del sistema. Esto demuestra la eficacia de la limitación de tiempo como herramienta para manejar servicios externos con tiempos de respuesta prolongados o indisponibilidad.

Consulta múltiple:

La mejora obtenida por la implementación de la estrategia anterior se ve maximizada en escenarios donde se deben buscar múltiples recursos. Por ejemplo, en un escenario donde se requieren consultar 30 DOI diferentes, aplicando un enfoque de búfer con 5 DOI consultados en concurrente, podemos inferir los siguientes datos para el tiempo total de consulta.

- En cada lote de consultas en concurrente (5 DOI a la vez), el tiempo de consulta será el del DOI más largo dentro de ese grupo.
- Dado que los tiempos de consulta varían, en un promedio de tiempo de consulta de 3155 ms por DOI (basado en los datos previos).
- Para 30 DOI consultados en grupos de 5 en concurrente, tendríamos 6 lotes de consultas.
- Por lo tanto, el tiempo total de consulta estimado para 30 DOI = 6 lotes * 3155 ms = 18930 ms.

Este tiempo es significativamente menor que el tiempo que tomaría si se realizaran las consultas de manera secuencial (30 DOI * 3155 ms = 94650 ms), lo que demuestra la eficiencia de usar un enfoque de búfer y consultas en concurrente. Esta estrategia de manejo de consultas resulta en un uso más eficiente de los recursos y reduce drásticamente el tiempo de espera para los usuarios.

4.8 Resultados de: Exportación a formatos descargables

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.

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.

Cada uno de estos campos representa una métrica o dato específico del ISSN en cuestión:

- **issn**: el identificador de la revista/publicación seriada.
- **journal_title**: el título de la revista/publicación seriada.
- **publisher**: la editorial la revista/publicación seriada.

- **openalex_issn_cites**: el número de citas que tiene la revista/publicación seriada según OpenAlex.
- **crossref_DOI**: el número total de DOI según Crossref.
- **doaj_issn_presence**: indicador de presencia en DOAJ.
- **redib_presence**: indicador de presencia en REDIB.
- **wos_collections**: colecciones de Web of Science en las que la revista/publicación seriada está presente.
- **scopus_citescore**: el CiteScore de la revista/publicación seriada según Scopus.
- **scopus_citescoretracker**: el seguimiento del CiteScore de la revista / publicación seriada según Scopus.
- **sjr_h-index**: el índice h de la revista / publicación seriada según SJR.
- **sjr_all-time_best_quartile**: el mejor cuartil de la revista / publicación seriada según SJR.

En la Figura 4.34, se representa la descarga de un CSV correspondiente a los siguientes nueve ISSN:

1. 2296-665X
2. 1089-7798
3. 0308-5961
4. 2683-8559
5. 0302-9743
6. 1865-0929
7. 2167-8359
8. 1852-2971
9. 2469-0228

	issn	type	title	publisher	openalex_issn_cites	crossref_dois	
1	2296-665X	Journal	Frontiers in Environmental Sc	Frontiers Media	34263	5214	
2	1089-7798	Journal	IEEE Communications Letters	IEEE Communications Society	238443	11736	
3	0308-5961	Journal	Telecommunications Policy	Elsevier BV	54508	3724	
4	2683-8559	Journal	Innovación y Desarrollo Tecnd	Universidad Nacional de La Plata	10	32	
5	0302-9743	Journal	Lecture Notes in Computer Sc	Sin editorial	5052724		
6	1865-0929	Journal	Communications in computer	Sin editorial	118837		
7	2167-8359	Journal	PeerJ	PeerJ, Inc.	198031	15556	
8	1852-2971	Journal	Derecho y Ciencias Sociales	National University of La Plata	77	113	
9	2469-0228	Journal	Publicación Electrónica de la	Asociación Paleontológica Argen	411	145	
	doaj_issn_presence	redib_presence	wos_collections	scopus_citescore	scopus_citescoretracker	sjr_h-index	sjr_all-time_best_quartile
1	No	No	Science Citation Index Expan	3.1 (2022)	3.2 (2023)	61	Q1 (2022)
2	No	No	Journal Citation Reports Scier	7.3 (2022)	6.3 (2023)	156	Q1 (2022)
3	No	No	Social Sciences Citation Index	8.1 (2022)	8.3 (2023)	81	Q1 (2022)
4	No	No					
5	No	No		2.2 (2022)	2.0 (2023)	446	Q3 (2022)
6	No	No		1.0 (2022)	0.8 (2023)	62	Q4 (2022)
7	No	No	Zoological Record, Science Ci	5.1 (2022)	4.1 (2023)	97	Q1 (2022)
8	No	No	Research Alert, Non-Essential				
9	No	No	Zoological Record	3.3 (2022)	1.8 (2023)	8	Q1 (2022)

Figura 4.34: Datos en formato CSV (ajustados para una visualización más sencilla en una única imagen)

o DOI:

- **doi:** el DOI de la publicación.
- **type:** el tipo de publicación.
- **title:** el título de la publicación.
- **authors:** los nombres de los autores de la publicación.
- **abstract:** el resumen de la publicación.
- **publication_year:** el año de publicación.
- **crossref_cites:** el número de citas de la publicación según Crossref.
- **openalex_doi_cites:** el número de citas de la publicación según OpenAlex.
- **doaj_doi_presence:** indicador de presencia de la publicación en DOAJ.
- **semanticscholar_cites:** el número de citas de la publicación según Semantic Scholar.
- **dimensions_cites:** el número de citas de la publicación según Dimensions.
- **altmetric_cites:** el número de citas según de la publicación Altmetric.
- Además, se adjuntan los datos del ISSN de la publicación (si es que los posee).

En la Figura 4.35, se representa la descarga de un CSV correspondiente a los siguientes diez DOI:

- | | |
|---------------------------------|-----------------------------------|
| 1. 10.3389/fenvs.2020.581591 | 6. 10.1109/CLEI53233.2021.9640225 |
| 2. 10.1109/LCOMM.2019.2898944 | 7. 10.1007/978-3-030-61702-8_25 |
| 3. 10.1016/j.telpol.2021.102261 | 8. 10.7717/peerj.4375 |
| 4. 10.24215/26838559e28 | 9. 10.24215/18522971e087 |
| 5. 10.1007/978-3-031-07802-6_9 | 10. 10.5710/PEAPA.21.05.2020 |

	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 Pourahma	In this letter, we present a	2019
3	10.1016/j.telpol.2021.102261	journal-article	Internet of things and the eco	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. Piwowar,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	semanticsscholar_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 4.35: Datos DOI en formato CSV sin la información de sus correspondientes ISSN (modificados para facilitar su visualización en una sola imagen)

4.9 Resultados de: Complemento web

Considerando la diversidad de navegadores y la ausencia de un estándar para identificar las publicaciones presentes en un buscador o base de datos académica, se desarrolló un prototipo de extensión para el navegador Chrome.

[El complemento web de HERA](#) ofrece a los usuarios la capacidad de acceder rápidamente a métricas e información relevante sobre documentos académicos a través de su navegador. Su principal función es reconocer los identificadores DOI presentes en la página web en la que el usuario se encuentra y, mediante una consulta a la API de HERA 2.0, recopila y presenta la información asociada a estos documentos.

La información proporcionada por el complemento se adapta al contexto de la página web en la que se encuentra el usuario. En particular, el complemento de HERA 2.0 está orientado a reconocer dos contextos diferentes, los cuales se describen a continuación.

Página específica de una publicación

Si la página es específica de un documento (es decir, se refiere a un único DOI), el complemento ofrece información detallada de dicho artículo. En particular, se presentan métricas e indicadores tanto del artículo como de la revista en la que se publicó (ver Figura 4.36, 4.37 y 4.38), y se incluye un enlace a los resultados de buscar el artículo en cuestión en HERA 2.0.

The image shows a screenshot of a PeerJ article page. The article title is "The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles". The authors listed are Heather Piwowar, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Matthias, Bree Norlander, Ashley Farley, Jevin West, and Stefanie Haustein. The article was published on February 13, 2018. The abstract discusses the need for large-scale, up-to-date, and reproducible studies assessing the prevalence and characteristics of OA, and mentions the use of oaDOI, an open online service that determines OA status for 67 million articles. The abstract also mentions the use of three samples, each of 100,000 articles, to investigate OA in three populations: (1) all journal articles assigned a Crossref DOI, (2) recent journal articles indexed in Web of Science, and (3) articles viewed by users of Unpaywall, an open-source browser extension that lets users find OA articles using oaDOI. The abstract estimates that at least 28% of the scholarly literature is OA (19M in total) and that this proportion is growing, driven particularly by growth in Gold and Hybrid. The most recent year analyzed (2015) also has the highest percentage of OA (45%). Because of this growth, and the fact that readers disproportionately access newer articles, we find that Unpaywall users encounter OA quite frequently: 47% of articles they view are OA. Notably, the most common mechanism for OA is not Gold, Green, or Hybrid OA, but rather an under-recognized category we dub Bronze: articles made available on a preprint server.

The HERA sidebar on the right provides the following metrics and links:

- DOI: 10.7717/peerj.4375
- Total citations: 588
- Recent citations: 197
- Field Citation Ratio: 486
- Relative Citation Ratio: 9.87
- Citas: 541
- Citas: 434
- Citas: 411
- Citas: 588
- Menciones: 817
- Publicado en: PeerJ (2167-8359), PeerJ, Inc.
- Ver en HERA

Figura 4.36: Complemento de HERA en funcionamiento en la web de PeerJ: The state of OA

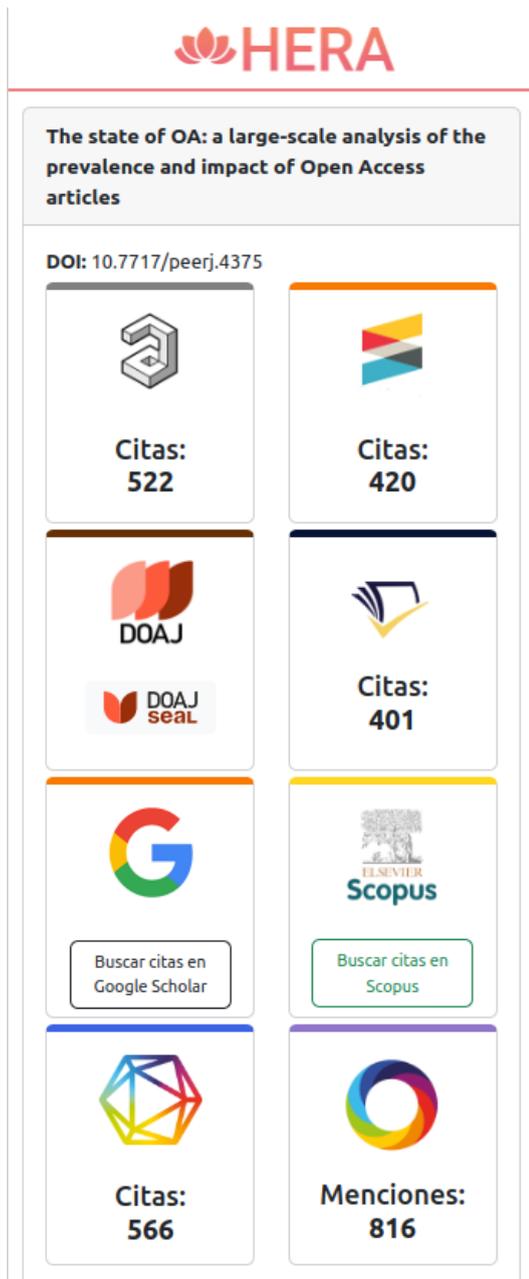


Figura 4.37: Zoom en la barra del complemento, ejemplo The state of OA (parte 1)

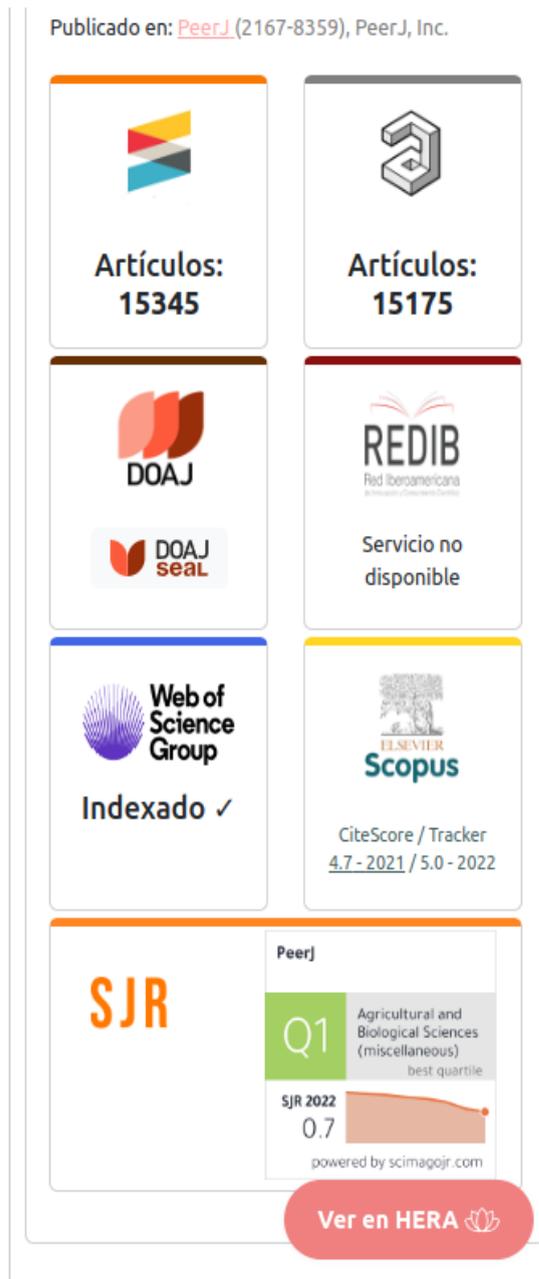


Figura 4.38: Zoom en la barra del complemento, ejemplo The state of OA (parte 2)

Página que contiene varias publicaciones

Por otro lado, si el usuario se encuentra en una página que presenta varios documentos, como podría ser una página de resultados en un motor de búsqueda académico (por ejemplo, Google Scholar) o un agregador de artículos académicos (por ejemplo, DOAJ), el complemento identifica todos los DOI presentes y proporciona una visión simplificada para cada uno de ellos. En este escenario, la información proporcionada se centra en el conteo de citas. Esta decisión se justifica principalmente por dos razones: por un lado, desde un punto de vista visual, mostrar una cantidad limitada de información permite una lectura rápida y clara para el

usuario; por otro lado, existen limitaciones inherentes a los servicios que proveen los datos, lo que restringe la cantidad de información que puede ser obtenida y mostrada en un contexto de múltiples documentos.

Este enfoque ofrece una visión rápida del impacto de cada documento sin abrumar al usuario con una cantidad excesiva de datos.

Ejemplo de dicho funcionamiento en las figuras 4.39 y 4.40:

The image shows a search results page on Semantic Scholar for the query 'clustalw'. The search results are sorted by relevance and show several entries. On the right side, there is a sidebar for the HERA (HERA Research and Analysis) tool. The sidebar contains a list of search results with their respective citation counts (Citas (OpenAlex)) and a 'Ver en HERA' button for each entry. The entries in the sidebar are: 'Multiple Sequence Alignment Using ClustalW and ClustalX' (1424 citations), 'CLUSTALW software' (1 citation), 'Computational Management of Alignment of Multiple Protein Sequences Using ClustalW' (4 citations), 'Primer design and in silico analysis using CLUSTALW and MUSCLE for L-arabinose isomerase (araA) gene detection in thermophilic bacteria' (4 citations), and 'Hybrid Parallel Implementation of Multiple Sequence Alignment Software ClustalW on Intel Xeon Phi' (4 citations).

Figura 4.39: Búsqueda de 'clustalw' en Semantic Scholar y funcionamiento del complemento de HERA

This image is a zoomed-in view of the HERA sidebar from the previous figure. It shows two search results. The first result is 'Multiple Sequence Alignment Using ClustalW and ClustalX' with 1409 citations (Citas (OpenAlex): 1409) and a 'Ver en HERA' button. The second result is 'GPU-ClustalW: Using Graphics Hardware to Accelerate Multiple Sequence Alignment' with 39 citations (Citas (OpenAlex): 39) and a 'Ver en HERA' button.

Figura 4.40: Zoom a la barra del complemento en búsqueda de 'clustalw' en google scholar

Además, el complemento incluye un botón (ver Figura 4.41) que redirige al usuario a la página web de HERA 2.0 para obtener una visión detallada del documento seleccionado. Esta función no solo ofrece al usuario la posibilidad de profundizar en la información del documento, sino que también promueve la visibilidad de la plataforma Hera.



Figura 4.41: Botón de redirección a la web de HERA

Este diseño de funcionamiento dual permite al complemento ser útil en una amplia gama de situaciones, al tiempo que se gestiona de manera eficiente las solicitudes a la API de HERA 2.0 y se minimiza el tiempo de respuesta.

En resumen, el objetivo principal del complemento web de HERA 2.0 es proporcionar a los usuarios un acceso rápido y sencillo a la información relevante sobre documentos académicos, directamente desde el sitio que están navegando, evitando el proceso de búsqueda manual en la página web de HERA.

4.10 Resultados de: 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:

1. **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 siguiente Figura 4.42, la fuente de información DOAJ responde con el mensaje de recurso no encontrado:



Figura 4.42: Métricas de DOAJ no encontradas

2. **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 4.43).



Figura 4.43: Tarjeta de REDIB informando: Servidor no disponible.

3. **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 4.44).



Figura 4.44: Tarjeta de OpenAlex informando: 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.

4.11 Resultados en: Base de datos local

En el análisis llevado a cabo para implementar una base de datos local, inicialmente se exploró la opción de utilizar el snapshot proporcionado por Unpaywall.

Unpaywall [75] es un recurso valioso que indexa artículos de revistas revisadas por pares, así como pre-publicaciones de repositorios como arXiv, proporcionando identificadores DOI. La fuente de estos datos proviene de más de 50000 revistas y

repositorios de acceso abierto de todo el mundo, así como datos abiertos de PubMed Central, DOAJ, Crossref y DataCite.²¹

Para la actualización constante de estos datos, se consideró la posibilidad de utilizar su data feed²². Se diseñó un prototipo utilizando Node, Elasticsearch²³, Redis²⁴ y GraphQL²⁵ para cargar el snapshot en Elastic/Kibana²⁶. Sin embargo, durante este proceso, se encontró y examinó a OpenAlex.

OpenAlex es una base de datos más amplia que Unpaywall y proporciona una mayor cantidad de recursos, posicionándose como una fuente de datos vital para HERA 2.0. Además, ofrece una API REST, así como la posibilidad de descargar e instalar una copia completa de la base de datos de OpenAlex en su propio servidor utilizando el correspondiente snapshot.

No obstante, a pesar de las fortalezas de OpenAlex, finalmente se optó por no implementar esta opción por una serie de razones. Aunque el snapshot se actualiza aproximadamente una vez al mes y la documentación de OpenAlex proporciona una guía detallada sobre cómo descargar el snapshot y cargarlo en su propia base de datos, la guía también advierte que este proceso es complicado.²⁷

OpenAlex reconoce que trabajar con un conjunto de datos tan grande y complejo rara vez se realiza según lo planeado y sugiere que los usuarios intenten primero con la API REST, que puede responder a la mayoría de las preguntas y tiene una barrera de entrada mucho más baja. Esta advertencia, combinada con las consideraciones de los recursos y el esfuerzo necesarios para implementar y mantener una base de datos local, llevó a la decisión de no avanzar con la implementación de OpenAlex en esta etapa.

Aunque el intento de implementar una base de datos local utilizando los snapshots de Unpaywall y OpenAlex no tuvo éxito, el proceso permitió un mejor entendimiento de las oportunidades y desafíos asociados con este enfoque. Estas experiencias y conocimientos serán valiosos en la exploración de futuras oportunidades para mejorar la gestión de datos en HERA.

²¹ [Unpaywall - OurResearch](#)

²² [Unpaywall - Data Feed](#)

²³ Es un servidor de búsqueda basado en Lucene, diseñado para realizar análisis y búsquedas en tiempo real en grandes volúmenes de datos.

²⁴ Es un almacén de estructura de datos en memoria, usado como base de datos, caché y corredor de mensajes, conocido por su flexibilidad y rendimiento.

²⁵ Es un lenguaje de consultas para APIs.

²⁶ Kibana es una interfaz de usuario para Elasticsearch que proporciona visualización de datos en tiempo real.

²⁷ [OpenAlex - Snapshot](#)

4.12 Resumen

En esta sección, se proporciona un resumen de las mejoras implementadas en la versión 2.0 del sistema HERA y los resultados obtenidos. Las propuestas se han centrado en extender la funcionalidad y el alcance, y en mejorar el rendimiento y la usabilidad de la aplicación.

- Mejora #1: Nueva interfaz: la interfaz de usuario se rediseñó completamente para mejorar su usabilidad y accesibilidad. Los resultados ahora se muestran en un formato más claro y fácil de entender, con secciones y etiquetas mejor definidas.
- Mejora #2: Integración de información de autores e instituciones: se integró información adicional de autores a los resultados de búsqueda, la cual permite a los usuarios obtener una visión más amplia y contextualizada de las publicaciones académicas que están buscando.
- Mejora #3: Actualización del banco de bases de datos: se realizó un análisis exhaustivo de las fuentes de datos existentes y se identificaron nuevas fuentes de datos potenciales para mejorar la relevancia y utilidad de los datos proporcionados por HERA 2.0.
- Mejora #4: Refactorización de código: el código del sistema se refactorizó para mejorar su eficiencia y mantenibilidad, lo cual implicó el uso de técnicas modernas de desarrollo de software y la adopción de buenas prácticas de programación.
- Mejora #5: Gestión de las fuentes de información: se implementó un gestor para administrar las diferentes fuentes de datos y la información que estas proveen. Por un lado, esto permite proveer al usuario de información de mejor calidad. Por otro lado, otorga flexibilidad a HERA 2.0 para adaptarse al dinamismo existente en el dominio.
- Mejora #6: Implementación de API REST: se incorporó una API REST para permitir la integración de HERA 2.0 con otras aplicaciones y sistemas, lo que amplía el alcance y sus posibilidades de uso.
- Mejora #7: Búsqueda simultánea de múltiples recursos: se implementó la capacidad de buscar simultáneamente en múltiples recursos, lo que aumenta la eficacia de las búsquedas y proporciona resultados más completos al mismo tiempo que mejora la experiencia del usuario.
- Mejora #8: Exportación a formatos descargables: se agregó la capacidad de exportar los resultados de búsqueda a varios formatos descargables populares (CSV y JSON), lo que facilita la recopilación y el post-procesamiento de los datos de forma separada para los usuarios.

- Mejora #9: Extensión para navegador Chrome: se desarrolló una extensión para el navegador Chrome que permite a los usuarios obtener información sobre métricas e indicadores de diferentes publicaciones directamente desde el sitio que están navegando. Si bien consiste de un prototipo, su desarrollo sirve a modo de prueba de concepto de la propuesta.
- Mejora #10: Manejo de errores y recursos no encontrados: se robusteció la gestión de errores en combinación con una comunicación más clara hacia los usuarios cuando ocurren situaciones excepcionales referidas a la no disponibilidad de recursos o servicios.
- Mejora #11: Base de datos local: se exploró la posibilidad de implementar una base de datos local. Aunque finalmente no se implementó esta mejora, el análisis realizado proporcionó una valiosa comprensión de las oportunidades y desafíos asociados con esta opción.

Capítulo 5

Conclusiones y Trabajo Futuro

En este capítulo, se resume el trabajo realizado y se presentan las conclusiones de esta tesina. Luego, se describen algunas posibles líneas de trabajo a futuro.

Conclusiones

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, esta tesina tuvo como objetivo principal el desarrollo de una segunda versión de HERA, aspirando a optimizar su rendimiento, escalabilidad, alcance y soporte.

Inicialmente, se realizó un análisis exhaustivo de diversas bases de datos académicas. Este análisis permitió evaluar el contenido disponible, los métodos de evaluación utilizados, las métricas exhibidas y los medios de acceso a los datos en cada una de ellas. El objetivo de este estudio fue seleccionar las bases de datos que proporcionaron la información más relevante y abarcativa y determinar los medios de acceso más viables para la implementación.

Luego, se emprendió un estudio minucioso de la herramienta HERA 1.0, buscando identificar sus puntos fuertes y las áreas de mejora, con el fin de proponer y realizar modificaciones que pudieran extender su funcionamiento y mejorar su rendimiento y usabilidad.

El trabajo anterior proporcionó un entendimiento detallado de diversos aspectos de la herramienta y oportunidades para su mejora. En ese sentido, se definieron 11 propuestas diferentes que apuntan a mejorar la funcionalidad, el rendimiento y la usabilidad de la aplicación HERA 1.0. A continuación, se resumen los resultados obtenidos:

- **Mejora #1: Nueva interfaz:** HERA 1.0 presentaba una interfaz funcional, no obstante algunos elementos como las estadísticas de citas y ciertas etiquetas resultaban difíciles de interpretar. En HERA 2.0, se realizó una completa reestructuración del diseño de la interfaz de usuario los resultados se muestran ahora en un formato más claro y comprensible, con secciones y etiquetas mejor definidas, lo que facilita la interacción del usuario con la plataforma.

- **Mejora #2: Integración de información de autores:** HERA 1.0 sólo lista los nombres y apellidos de los autores de un artículo. A partir de la incorporación de fuentes adicionales de datos como ORCID y ROR, HERA 2.0 integra información adicional de autores a los resultados de búsqueda, permitiendo a los usuarios obtener una visión más amplia y contextualizada de las publicaciones académicas que están buscando.
- **Mejora #3: Actualización del banco de bases de datos:** en HERA 1.0, algunas fuentes de datos, como Microsoft Academic, se volvieron inaccesibles. En HERA 2.0, se realizó un análisis exhaustivo para identificar y agregar nuevas fuentes de datos como OpenAlex, al tiempo que se verificó que las fuentes de datos existentes siguieran brindando información relevante.
- **Mejora #4: Refactorización de código:** el código de la versión 1.0 de HERA presentaba ciertas debilidades en cuanto a eficiencia y mantenibilidad, incluyendo una falta de modularización y una gestión inadecuada de múltiples fuentes de datos. En HERA 2.0, se realizó una completa refactorización del código utilizando técnicas modernas de desarrollo de software y buenas prácticas de programación, incluyendo la migración de JavaScript a TypeScript.
- **Mejora #5: Gestión de las fuentes de información:** En HERA 1.0, la gestión de las fuentes de datos no era óptima, debido a que la adición o remoción de bancos de datos requería cambios significativos en el código fuente. En HERA 2.0, se implementó un gestor para administrar las diferentes fuentes de datos y la información que estas proveen. Esta mejora permite proveer al usuario de información de mejor calidad y otorga flexibilidad a HERA 2.0 para adaptarse al dinamismo existente en el dominio.
- **Mejora #6: Implementación de API REST:** HERA 1.0 operaba de manera aislada, limitando su capacidad de integración con otras aplicaciones y sistemas. En HERA 2.0, se incorporó una API REST que permite la integración de HERA con otras aplicaciones y sistemas, ampliando así su alcance y sus posibilidades de uso.
- **Mejora #7: Búsqueda simultánea de múltiples recursos:** la versión 1.0 de HERA solo permitía realizar búsquedas de manera secuencial, limitando la eficiencia de las búsquedas. En HERA 2.0, se implementó la capacidad de buscar simultáneamente en múltiples recursos. Esta mejora no solo proporciona resultados más completos sino que también mejora la experiencia del usuario al acelerar la recuperación de datos.
- **Mejora #8: Exportación a formatos descargables:** HERA 1.0 no ofrecía la opción de exportar los resultados de las búsquedas. En HERA 2.0, se agregó la capacidad de exportar los resultados de búsqueda a varios formatos

descargables populares (CSV y JSON), facilitando la recopilación y el post-procesamiento de los datos para los usuarios.

- **Mejora #9: Extensión para navegador Chrome:** A pesar de la funcionalidad de HERA 1.0, la obtención de métricas e indicadores de publicaciones requería el acceso directo a la plataforma. Con HERA 2.0, se desarrolló una extensión para el navegador Chrome que permite a los usuarios obtener esta información directamente desde el sitio que están navegando, facilitando aún más el acceso a datos recolectados por HERA. Además, la extensión incluye un botón que redirige al usuario al sitio web para obtener una visión más detallada del documento seleccionado, lo que no sólo permite un mayor análisis del documento, sino que también incrementa la visibilidad de la plataforma web.
- **Mejora #10: Manejo de errores y recursos no encontrados:** la versión 1.0 de HERA no proporcionaba retroalimentación clara al usuario cuando ocurrían situaciones excepcionales. En HERA 2.0, se robusteció la gestión de errores y se proporciona una comunicación más clara a los usuarios cuando ocurren situaciones como la no disponibilidad de recursos o servicios.
- **Oportunidad de mejora futura #11: Base de datos local:** Aunque finalmente no se implementó esta mejora en HERA 2.0, se exploró la posibilidad de implementar una base de datos local. El análisis realizado proporcionó una valiosa comprensión de las oportunidades y desafíos asociados con esta opción y se mantiene como una oportunidad para mejoras futuras.

A partir de los resultados obtenidos, se considera que el objetivo planteado inicialmente se ha cumplido. 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.

Trabajos futuros

1. Implementación de una base de datos local: la implementación de una base de datos local que contenga la información podría proporcionar un mejor control y acceso a los datos. Lo cual podría estar basada en OpenAlex debido a su amplio conjunto de datos y su continuo desarrollo y mantenimiento. Este enfoque permitiría un acceso más rápido y controlado a los datos, reduciría la dependencia de las solicitudes fetch y mejoraría el rendimiento global del sistema. Al utilizar OpenAlex como base de datos local, se podrían explotar

las funcionalidades que ofrece esta plataforma, al tiempo que se mantendría la actualización y la relevancia de los datos.

2. Expansión de las Funcionalidades de Búsqueda en OpenAlex: el hecho de que OpenAlex ofrezca la posibilidad de buscar por título, autores y citas representa una oportunidad para expandir las funcionalidades de búsqueda de HERA. Por ejemplo, se podría implementar una función que permita buscar por título, recuperar el DOI del artículo y luego usar ese DOI para buscar en las demás fuentes de datos. Además, aprovechar la información sobre las citas podría permitir la creación de nuevas características, como la identificación de artículos científicos relacionados con el tema de búsqueda o la presentación de otros trabajos de los mismos autores.
3. Integración de un Sistema de Caché: incorporar un sistema de caché podría mejorar la eficiencia de las operaciones de búsqueda y recuperación de datos. Al almacenar temporalmente los resultados de las solicitudes más recientes o comunes, se podría reducir el número de solicitudes necesarias y acelerar el tiempo de respuesta.
4. Monitoreo de la Disponibilidad de API: desarrollar un sistema de monitoreo para las API podría ayudar a identificar y gestionar problemas de disponibilidad de manera más eficiente. Si una API no responde o deja de existir, el sistema podría suspender automáticamente las solicitudes a esa API durante un período de tiempo determinado o ajustar la configuración de manera dinámica.
5. Configuración Dinámica de API: en lugar de depender de un archivo JSON estático para la configuración de las API, podría ser útil desarrollar un sistema más flexible y adaptable. Este sistema permitiría ajustar la configuración de las API en tiempo real, en función de la disponibilidad, el rendimiento y otros factores.
6. Ampliación de la funcionalidad de la extensión web a otros navegadores: se ha centrado en la integración con Google Chrome debido a su simplicidad y prevalencia, es factible la adaptación de la extensión para su compatibilidad con una variedad de navegadores. Además, es importante destacar que siempre existen oportunidades para la mejora y optimización del código, así como para la implementación de nuevas funcionalidades.
7. Integración con Open Journal Systems (OJS): en un intento por aumentar la transparencia y facilitar el acceso a métricas relevantes, se podría considerar la integración con el sistema de gestión de revistas OJS a través del desarrollo de un plugin. Esto no solo permitiría visualizar las métricas directamente desde OJS, sino que también proporcionaría una forma eficaz de recopilar datos a través de fuentes de terceros.

Referencias

- [1] “How Does Technology Help In Doing Academic Research? | Scoop Byte”, el 15 de julio de 2021.
<https://www.scoopbyte.com/how-does-technology-help-in-doing-academic-research/> (consultado el 4 de julio de 2023).
- [2] R. S. Qualtrics, “How Technology is Changing Academic Research”, *Wired*, el 3 de julio de 2013. Consultado: el 4 de julio de 2023. [En línea]. Disponible en: <https://www.wired.com/insights/2013/07/how-technology-is-changing-academic-research/>
- [3] “2018_10_04_STM_Report_2018.pdf”. Consultado: el 4 de julio de 2023. [En línea]. Disponible en: https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf
- [4] R. Repiso, “Cómo identificar una revista de calidad”, *Cardiocyte*, vol. 50, ene. 2015, doi: 10.1016/j.carcor.2014.12.002.
- [5] J. F. Porto, E. Rucci, y G. L. Villarreal, “HERA - Herramienta para Enriquecimiento de Recursos Académicos”, *Conf. Int. BIREDIAL-ISTEC*, sep. 2022, Consultado: el 4 de julio de 2023. [En línea]. Disponible en: <https://submissions.istec.org/index.php/biredial-istec/article/view/174>
- [6] L. Team, “What Is Code Refactoring? Definition, Benefits and Why It’s Important”, *Lvivity*, el 3 de diciembre de 2021.
<https://lvivity.com/what-is-code-refactoring> (consultado el 4 de julio de 2023).
- [7] V. Durieux y P. A. Gevenois, “Bibliometric Indicators: Quality Measurements of Scientific Publication”, *Radiology*, vol. 255, núm. 2, pp. 342–351, may 2010, doi: 10.1148/radiol.09090626.
- [8] A. Martín-Martín, E. Orduna-Malea, M. Thelwall, y E. Delgado López-Cózar, “Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories”, *J. Informetr.*, vol. 12, núm. 4, pp. 1160–1177, nov. 2018, doi: 10.1016/j.joi.2018.09.002.
- [9] D. Hicks, P. Wouters, L. Waltman, S. de Rijcke, y I. Rafols, “Bibliometrics: The Leiden Manifesto for research metrics”, *Nature*, vol. 520, núm. 7548, Art. núm. 7548, abr. 2015, doi: 10.1038/520429a.
- [10] “El factor de impacto”.
http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0034-74502010000100014 (consultado el 5 de julio de 2023).
- [11] E. Garfield, “The History and Meaning of the Journal Impact Factor”, *JAMA*, vol. 295, núm. 1, pp. 90–93, ene. 2006, doi: 10.1001/jama.295.1.90.
- [12] R. Caldwell, “Research Guides: Calculating Journal Impact Factor: Home”.
<https://libguides.utk.edu/c.php?g=964647&p=6968521> (consultado el 7 de junio de 2023).
- [13] M. S. Di Bitetti y J. A. Ferreras, “Publish (in English) or perish: The effect on citation rate of using languages other than English in scientific publications”, *Ambio*, vol. 46, núm. 1, pp. 121–127, feb. 2017, doi: 10.1007/s13280-016-0820-7.
- [14] G. González-Alcaide, J. C. Valderrama-Zurián, y R. Aleixandre-Benavent, “The Impact Factor in non-English-speaking countries”, *Scientometrics*, vol. 92, núm. 2, pp. 297–311, ago. 2012, doi: 10.1007/s11192-012-0692-y.
- [15] J. E. Hirsch, “An index to quantify an individual’s scientific research output”,

- Proc. Natl. Acad. Sci.*, vol. 102, núm. 46, pp. 16569–16572, nov. 2005, doi: 10.1073/pnas.0507655102.
- [16] T. Ireland, K. MacDonald, y P. Stirling, “The H-Index: What Is It, How Do We Determine It, and How Can We Keep Up With It?”.
- [17] L. Bredahl, “Research guides: Calculate Your Academic Footprint: Your H-Index”.
<https://subjectguides.uwaterloo.ca/calculate-academic-footprint/YourHIndex> (consultado el 22 de junio de 2023).
- [18] “Track journal performance and make decisions with CiteScore”, *beta.elsevier.com*.
<https://beta.elsevier.com/connect/citescore-a-new-metric-to-help-you-choose-the-right-journal> (consultado el 29 de junio de 2023).
- [19] “Discover the attention surrounding your research”, *Altmetric*.
<https://www.altmetric.com/> (consultado el 4 de julio de 2023).
- [20] J. Priem, D. Taraborelli, P. Groth, y N. Cameron, “altmetrics: a manifiesto”, 2010.
<https://altmetrics.org/manifiesto/> (consultado el 29 de junio de 2023).
- [21] T. Gruber, “Academic sell-out: how an obsession with metrics and rankings is damaging academia”, *J. Mark. High. Educ.*, vol. 24, núm. 2, pp. 165–177, jul. 2014, doi: 10.1080/08841241.2014.970248.
- [22] “Reviewers | What is peer review? | Elsevier”, *beta.elsevier.com*.
<https://beta.elsevier.com/reviewer/what-is-peer-review> (consultado el 29 de junio de 2023).
- [23] “Scopus-fact-sheet-2022_WEB.pdf”. Consultado: el 5 de julio de 2023. [En línea]. Disponible en:
https://assets.ctfassets.net/o78em1y1w4i4/27oXomTnnXzUXDBXmWUbkd/d2f45dff31a77234f78f9e71abeb1d18/Scopus-fact-sheet-2022_WEB.pdf
- [24] “Content Policy and Selection - Content - Scopus - Solutions | Elsevier”.
<https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection> (consultado el 22 de junio de 2023).
- [25] “History of citation indexing”, *Web of Science Group*.
<https://clarivate.com/webofsciencegroup/essays/history-of-citation-indexing/> (consultado el 29 de junio de 2023).
- [26] “Web of Science editorial selection process”, *Clarivate*.
<https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/web-of-science/core-collection/editorial-selection-process/> (consultado el 29 de junio de 2023).
- [27] P. Mongeon y A. Paul-Hus, “The journal coverage of Web of Science and Scopus: a comparative analysis”, *Scientometrics*, vol. 106, núm. 1, pp. 213–228, ene. 2016, doi: 10.1007/s11192-015-1765-5.
- [28] M. Visser, N. J. van Eck, y L. Waltman, “Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic”, *Quant. Sci. Stud.*, vol. 2, núm. 1, pp. 20–41, abr. 2021, doi: 10.1162/qss_a_00112.
- [29] “Directory of Open Access Journals – DOAJ”, el 27 de junio de 2023.
<https://doaj.org/> (consultado el 4 de julio de 2023).
- [30] “Directory of Open Access Journals – The DOAJ Seal”.
<https://doaj.org/apply/seal/> (consultado el 4 de julio de 2023).
- [31] “Directory of Open Access Journals – DOAJ OAI-PMH”.
<https://doaj.org/docs/oai-pmh/> (consultado el 4 de julio de 2023).
- [32] “Directory of Open Access Journals – DOAJ API”. <https://doaj.org/api/v3/docs>

- (consultado el 4 de julio de 2023).
- [33] A. P. Llamazares, “Guías de la BUMA: Herramientas y recursos para localizar indicios de calidad: REDIB”. <https://biblioguias.uma.es/HerramientasIndiciosCalidad/REDIB> (consultado el 4 de julio de 2023).
- [34] “Consejo Superior de Investigaciones Científicas”. <https://www.csic.es/es> (consultado el 4 de julio de 2023).
- [35] Rosa-Clark, “History”, *Crossref*. <https://www.crossref.org/community/about/history/> (consultado el 4 de julio de 2023).
- [36] Rosa-Clark, “Maintaining your metadata”, *Crossref*. <https://www.crossref.org/documentation/register-maintain-records/maintaining-our-metadata/> (consultado el 4 de julio de 2023).
- [37] Rosa-Clark, “Membership terms”, *Crossref*. <https://www.crossref.org/membership/terms/> (consultado el 29 de junio de 2023).
- [38] E. Orduna-Malea y E. Delgado López-Cózar, “Taller: Research Analytics: la monitorización del impacto académico de un investigador”, presentado en VII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales de América Latina (BIREDIAL-ISTEC’17) y XII Simposio Internacional de Biblioteca Digitales (SIBD’17) (La Plata, 2017), 2017. Consultado: el 29 de junio de 2023. [En línea]. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/63599>
- [39] “Scimago Journal & Country Rank”. <https://www.scimagojr.com/> (consultado el 4 de julio de 2023).
- [40] “SCImago - SCImago”. <https://www.scimago.com/> (consultado el 4 de julio de 2023).
- [41] “Allen Institute for AI”. <https://allenai.org/> (consultado el 4 de julio de 2023).
- [42] A. E. Cha, “Paul Allen’s AI research group unveils program that aims to shake up how we search scientific knowledge. Give it a try.”, *Washington Post*, el 26 de octubre de 2021. Consultado: el 8 de junio de 2023. [En línea]. Disponible en: <https://www.washingtonpost.com/news/to-your-health/wp/2015/11/02/paul-allens-ai-research-group-unveils-program-that-aims-to-shake-up-how-we-search-scientific-knowledge-give-it-a-try/>
- [43] “Semantic Scholar Academic Graph API | Semantic Scholar”. <https://www.semanticscholar.org/product/api#Fetch-Paper> (consultado el 4 de julio de 2023).
- [44] “Digital Science: tools for open, collaborative and inclusive research”, *Digital Science*. <https://www.digital-science.com/> (consultado el 4 de julio de 2023).
- [45] L. Broadberry, “Dimensions Badges: A new way to see citations”, *Altmetric*, el 26 de enero de 2018. <https://www.altmetric.com/blog/dimensions-badges-a-new-way-to-see-citations/> (consultado el 4 de julio de 2023).
- [46] S. Ebrahimi, F. Setareh, y M. HosseinChari, “Assessing the relationship between the alternative metrics of visibility and social bookmarking with citation index in PLOS Altmetrics”, *Iran. J. Inf. Process. Manag.*, vol. 31, núm. 3, pp. 693–693, dic. 2022, doi: 10.35050/JIPM010.2016.011.
- [47] G. Lockwood y Neurobiology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, 6525XD, the Netherlands, “Academic clickbait: articles with positively-framed titles, interesting phrasing, and no wordplay get

- more attention online.” el 29 de junio de 2016. doi: 10.15200/winn.146723.36330.
- [48] J. L. Ortega, “Exploratory analysis of Publons metrics and their relationship with bibliometric and altmetric impact”, *Aslib J. Inf. Manag.*, vol. 71, núm. 1, pp. 124–136, ene. 2018, doi: 10.1108/AJIM-06-2018-0153.
- [49] E. Çelik, M. Dokur, B. Börkür Uysal, N. Şengül Samancı, y F. H. Demirelli, “Comparison of attention for cancer research on social media versus academia: An altmetric score analysis”, *Kanser arařtırmalarına sosyal medya ve akademik dünya ilgilerinin karşılaştırılması: Altmetrik skor analizi*, 2020, doi: 10.4999/uhod.204020.
- [50] J. G. Y. Luc *et al.*, “Does Tweeting Improve Citations? One-Year Results From the TSSMN Prospective Randomized Trial”, *Ann. Thorac. Surg.*, vol. 111, núm. 1, pp. 296–300, ene. 2021, doi: 10.1016/j.athoracsur.2020.04.065.
- [51] D. Godoy, A. Zunino, y C. Mateos, “Publication practices in the Argentinian Computer Science community: a bibliometric perspective”, *Scientometrics*, vol. 102, núm. 2, pp. 1795–1814, feb. 2015, doi: 10.1007/s11192-014-1450-0.
- [52] J. F. Porto, E. Rucci, y G. L. Villarreal, “HERA - Herramienta para Enriquecimiento de Recursos Académicos”, presentado en XI Conferencia Internacional de Bibliotecas y Repositorios Digitales (BIREDIAL-ISTEC) (Modalidad virtual, 3 al 7 de octubre de 2022), 2022. Consultado: el 29 de junio de 2023. [En línea]. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/148922>
- [53] M. Méndez, “Refactoring de código estructurado”, Tesis, Universidad Nacional de La Plata, 2010. Consultado: el 29 de junio de 2023. [En línea]. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/4184>
- [54] “¿Qué es una API de REST?” <https://www.redhat.com/es/topics/api/what-is-a-rest-api> (consultado el 29 de junio de 2023).
- [55] “React Bootstrap | React Bootstrap”. <https://react-bootstrap.netlify.app/> (consultado el 5 de julio de 2023).
- [56] M. O. contributors Jacob Thornton, and Bootstrap, “Bootstrap”. <https://getbootstrap.com/> (consultado el 5 de julio de 2023).
- [57] “React”. <https://es.react.dev/> (consultado el 5 de julio de 2023).
- [58] “Research Organization Registry (ROR)”, *Research Organization Registry (ROR)*. <https://ror.org/> (consultado el 4 de julio de 2023).
- [59] “ORCID”. <https://orcid.org/> (consultado el 4 de julio de 2023).
- [60] J. Priem, H. Piwowar, y R. Orr, “OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts”. arXiv, el 16 de junio de 2022. doi: 10.48550/arXiv.2205.01833.
- [61] “Scopus now includes 90 million + content records! | Elsevier Scopus Blog”. <https://blog.scopus.com/posts/scopus-now-includes-90-million-content-records> (consultado el 11 de julio de 2023).
- [62] “OurResearch: Tools to make research more open”. <https://ourresearch.org/> (consultado el 4 de julio de 2023).
- [63] T. Matthews, “LibGuides: Resources for Librarians: Web of Science Coverage Details”. <https://clarivate.libguides.com/librarianresources/coverage> (consultado el 11 de julio de 2023).
- [64] “Open Access”, *Web of Science Group*. <https://clarivate.com/webofsciencegroup/solutions/open-access/> (consultado el 11 de julio de 2023).

- [65] “Evaluate your university’s OA status”, *Dimensions*, el 10 de junio de 2022. <https://www.dimensions.ai/resources/evaluate-your-universitys-oa-status/> (consultado el 11 de julio de 2023).
- [66] M. Gusenbauer, “Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases”, *Scientometrics*, vol. 118, núm. 1, pp. 177–214, ene. 2019, doi: 10.1007/s11192-018-2958-5.
- [67] “OpenAlex API documentation - Overview”. <https://docs.openalex.org/> (consultado el 4 de julio de 2023).
- [68] “Entities overview - OpenAlex”. <https://docs.openalex.org/api-entities/entities-overview> (consultado el 4 de julio de 2023).
- [69] “Cited by in Scopus - Elsevier Developer Portal”. https://dev.elsevier.com/cited_by_scopus.html (consultado el 11 de julio de 2023).
- [70] “Architectural Styles and the Design of Network-based Software Architectures”. <https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm> (consultado el 11 de julio de 2023).
- [71] “JavaScript With Syntax For Types.” <https://www.typescriptlang.org/> (consultado el 6 de julio de 2023).
- [72] E. Gamma, *Patrones de diseño: elementos de software orientado a objetos reutilizable*. Pearson Educación, 2002.
- [73] C. Kenny, “What Is Web Scraping?”, *Zyte (formerly Scrapinghub) #1 Web Scraping Service*, el 1 de enero de 2021. <https://www.zyte.com/learn/what-is-web-scraping/> (consultado el 11 de julio de 2023).
- [74] “jsdom”. jsdom, el 11 de julio de 2023. Consultado: el 11 de julio de 2023. [En línea]. Disponible en: <https://github.com/jsdom/jsdom>
- [75] H. Else, “How Unpaywall is transforming open science”, *Nature*, vol. 560, núm. 7718, pp. 290–291, ago. 2018, doi: 10.1038/d41586-018-05968-3.