

Caso de éxito | Caltech

El uso de API Web of Science para agilizar procesos de solicitud de financiación

"La creciente naturaleza

multidisciplinaria de colaboraciones, colaboradores y de la ciencia en general requiere una fuente confiable e integral que garantice amplitud y profundidad de cobertura de manera transversal para las ciencias" - Tom Morrell, Bibliotecario en Caltech

Solicitar formalmente financiaciones y otros tipos de apoyos financieros puede ser un proceso complicado y largo, con exigentes requisitos y protocolos. Este fenómeno, conocido por todos los involucrados en el quehacer de la investigación, ha llevado a investigadores y bibliotecarios a buscar nuevos enfoques y procesos para simplificar y hacerlo más eficaz.

Un grupo de bibliotecarios de Caltech (Instituto de Tecnología de California) ha tenido un notable éxito al reducir la carga de trabajo a lo largo de los procesos de requisición de apoyos financieros para la facultad. Este trabajo conllevó el enfrentarse a enormes conjuntos de datos, incluyendo búsquedas en plataformas como Web of Science. Afortunadamente, los recursos de biblioteca de Caltech estaban en expansión para satisfacer esta demanda, incrementando la capacidad para manejar y analizar archivos grandes. Uno de los pasos de esta evolución fue añadir una utilidad programática diseñada para acelerar consultas y mejorar el formato de los datos recuperados. Esta herramienta añadida es alimentada por la API de *Web of Science*.

Colaboración de Datos

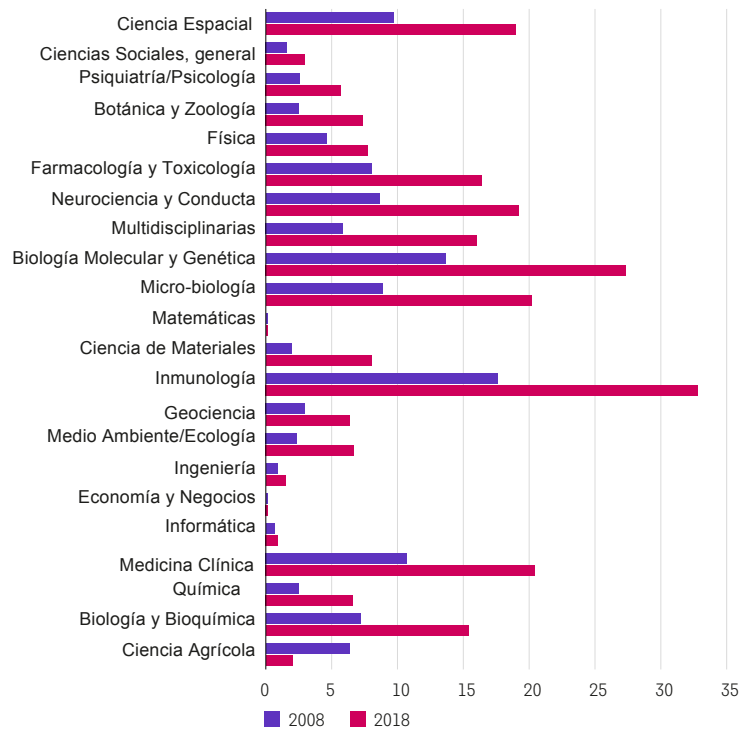
Los bibliotecarios de Caltech recopilaron múltiples consultas provenientes de investigadores principales (PIs) en disciplinas del campo de las ciencias físicas, pidiendo ayuda a través de solicitudes de apoyo financiero a la Fundación Nacional de Ciencias de los Estados Unidos (US National Science Foundation: NSF).¹ Las estipulaciones de la NSF incluyen información detallada de los colaboradores —o, el empleo del término oficial, datos de Colaboradores y otras Afiliaciones (COA por sus siglas en inglés).

Específicamente, al procesar solicitudes de financiamiento, la NSF utiliza información de COA durante el proceso de revisión de méritos, como la agencia lo indica en su sitio, “para ayudar a gestionar la selección de revisores”.² En la práctica, esto supone evitar cualquier sesgo de los revisores excluyendo de manera preventiva a cualquiera con el cual el investigador principal (PI) pudiese tener una relación de negocios o académica —incluyendo la coautoría de un trabajo de investigación, servicios prestados a un consejo editorial, o cualquier otra interacción oficial.

En teoría, esta operación parece muy sencilla. Sin embargo; en Caltech, los bibliotecarios se encontraban asistiendo a Investigadores Principales en astrofísica, física de alta energía y otros campos en los cuales el grado y frecuencia de los trabajos de autor múltiple han incrementado de manera notoria en los últimos años.

En 2012, datos de Web of Science apuntalaron un reporte de ScienceWatch⁴ sobre la proliferación de los trabajos académicos de autor múltiple. Cuatro años atrás, como lo señala la historia, los trabajos académicos del equipo multinacional en el Gran Colisionador de Hadrones (LHC por sus siglas en inglés) de la Organización Europea para la Investigación Nuclear (CERN por sus siglas en inglés) habían abatido el récord por trabajo académico con mayor cantidad de autores, con una lista de más de 3,000 colaboradores. Posteriormente, la búsqueda del bosón de Higgs, la cual se acercaba a una exitosa conclusión justo cuando apareció la historia de ScienceWatch, mantuvo la tendencia de la “híper-autoría”, así como hicieron otras colaboraciones.

Figura 1: Porcentaje de artículos publicados con 10 o más autores ordenados por categoría ESI



Los datos incluyen artículos publicados en Web of Science Citation Index Expanded (SCI-E) y Social Science Citation Index (SSCI). La totalidad de los datos no es del 100% debido a redondeo. Fecha de extracción: 5 de abril de 2019.

¹ <https://www.nsf.gov/pubs/2019/nsf19003/nsf19003.pdf>

² <https://www.nsf.gov/bfa/dias/policy/coa.jsp>

³ <https://clarivate.com/?p=35127&preview=1&ppp=0f3e27ff4c>

⁴ ScienceWatch, July 2012: http://archive.sciencewatch.com/newsletter/2012/201207/multi-author_papers/

La revista Nature reportó en 2018⁵ (con información de Web of Science) el constante incremento de trabajos académicos con más de 1000 autores. En ese momento, el récord de 3000 autores había sido fácilmente rebasado por un reporte de 2015 del diario Physical Review Letters⁶ que contaba con un listado de más de 5100 autores. (La publicación, de manera nada sorprendente, describía el trabajo de seguimiento llevado a cabo en el Gran Colisionador de Hadrones a propósito de la búsqueda del bosón de Higgs). En general, como reportó Nature, los campos de la física nuclear y la física de partículas representaban la mayoría de los trabajos publicados con 1000 o más autores.

Mientras tanto, como lo muestran las cifras de la plataforma Web of Science, el porcentaje de trabajos publicados con 10 o más autores ha aumentado de manera constante en todos los campos principales durante esta última década. En general, en artículos de diario contabilizado

454

publicaciones y cerca de 11,000 coautores únicos

Reportes Complejos

En Caltech, la tendencia ascendente vista en publicaciones con múltiples autores se hizo evidente en el momento en que personal de biblioteca decidió abordar la existencia del componente COA en las solicitudes ante la NSF.

El problema:

En ocasiones las solicitudes de una facultad conllevan una docena o más de Investigadores Principales (PIs), cada uno de los cuales es responsable de una proporción significativa de los resultados, lo cual involucra cientos de publicaciones y miles de autores de manera rutinaria, (en una ocasión, 454 publicaciones y casi 11000 coautores únicos). La recolección de datos en varios pasos tomó mucho tiempo, de la misma manera que el procesamiento y depuración de los resultados – todo ello mientras los bibliotecarios tenían trabajo a manos llenas atendiendo otras peticiones y actividades.

Complicaciones adicionales:

El personal de biblioteca comenzaba a recopilar solicitudes adicionales de la facultad que rebasaban la especialización de astrofísica de sus bases de datos. Aunque el archivo propio de Caltech ⁷ y el Sistema de Datos Astrofísicos de Harvard (ADS por sus siglas en inglés) permitiese acceso programático a los datos solicitados en astrofísica, las facultades de otros departamentos solicitaban asistencia de los bibliotecarios con la gestión de subsidios. El problema requería un recurso multidisciplinario.

⁵ www.natureindex.com/news-blog/paper-authorship-goes-hyper

⁶ <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.114.191803>

⁷ <https://authors.library.caltech.edu/>



"Muchas bibliotecas han comenzado a conformar equipos de especialistas de datos — personas que puedan interactuar con los datos de manera programática con el objetivo de generar reportes o llevar a cabo integraciones de sistemas"

Joy Painter, Bibliotecario, Caltech

Web of Science y REST API

La biblioteca de Caltech pudo utilizar la plataforma Web of Science como una fuente de datos mucho más amplia, ya que tiene una cobertura de más de 250 materias. Sin embargo, dada la escala de los datos, incluso usando la aplicación en línea por sí sola para soportar la recolección de datos y la curaduría de los mismos demostró ser problemático. Dicha capacidad experimentó un incremento a mediados del año 2018, cuando el personal de Caltech comenzó a utilizar una nueva versión de API REST de Web of Science como parte de una mayor colaboración entre ambas organizaciones. Esta nueva versión de API es una evolución del marco para la consulta programática de datos de Web of Science que facilita el aprovechamiento de datos de la misma para su uso en proyectos y sistemas internos.

Al reducir la complejidad técnica de la API (RESET vs. SOAP) y ampliar los formatos de datos soportados (XML/JSON), la funcionalidad y flexibilidad de la herramienta supera la enorme escala de la información involucrada, automatizando muchos de los pasos del proceso de recuperación de datos, dando así al personal de biblioteca más tiempo para concentrarse en otros aspectos de la asistencia que dan respecto de las solicitudes de subsidios y la facultad a la que sirven.

La experiencia de Caltech y de la API de Web of Science puede resultar didáctica para cualquier biblioteca que desee construir y desarrollar sus aptitudes de procesamiento de datos y así poder manejar solicitudes de datos de gran escala en manera interdisciplinaria, con la finalidad de servir a la facultad y cumplir su misión.

Joy Painter, bibliotecario en materias de Física, Matemáticas y Astronomía en Caltech, junto con un miembro clave del personal encargado de las solicitudes a la NSF comentó lo siguiente acerca de la incorporación de Web of Science y la nueva API al arsenal de la biblioteca:

"Muchas bibliotecas han comenzado a incorporar equipos de especialistas en datos — personas que puedan interactuar con los datos de manera programática con la finalidad de elaborar reportes o llevar a cabo integración de sistemas. La pregunta es, en tiempos en que las bibliotecas se adentran cada vez más en el mundo digital, ¿cómo deben posicionarse las bibliotecas para potenciar su competencia tanto en uso de dominios como en trabajo con datos para servir mejor a sus facultades y a las universidades como un todo?"

A través de este trabajo, creemos que se ha vuelto evidente que no sólo es posible sino también fundamental conjugar roles que anteriormente se habían desempeñado de manera independiente para así poder atacar de manera eficiente problemas reales que se dan en el ámbito de la investigación"

Caltech ha hecho que el código de su proyecto 'collaborator_reports' un código de fuente abierto, disponible en Github:

https://github.com/caltechlibrary/collaborator_reports

Conozca más en:

clarivate.com