

## Aplicación de técnicas de Inteligencia Artificial para evaluar la producción científico-académica de investigadores de Universidades públicas del Noroeste Argentino.

José Federico Medrano, Alejandro Vargas, Miguel Augusto Azar, Fernando Rubén Aramayo, Sergio Elias Rodriguez, Maria Antonia Tapia, Alfredo Ariel Diaz

jfedericomedrano@gmail.com, aleva98@yahoo.com, auazar@live.com, fernando.ruben.aramayo@gmail.com, egiors@gmail.com, maria.antonia.tapia@gmail.com, alfredodiazjy@gmail.com

VRAIn / Visualización y Recuperación Avanzada de Información / Facultad de Ingeniería  
Universidad Nacional de Jujuy - Ítalo Palanca 10, +54 (388) 4221587

### RESUMEN

La evaluación de la producción científica o específicamente la evaluación de la productividad de un científico, ha sido desde que se iniciaran las publicaciones de los resultados de la investigación, una tarea nada sencilla. Los indicadores bibliométricos se encargan de contabilizar y representar en cierta medida cuan productivo es un investigador, contabilizando la cantidad de publicaciones y la cantidad de citas recibidas por estas. Pero para tener una imagen completa, es necesario agregar otro aspecto relacionado con el impacto y popularidad de estas publicaciones, para ello el uso de las redes sociales académicas enriquecen los indicadores bibliométricos dándoles un sentido mucho más amplio, por ejemplo: cantidad de veces que un contenido es descargado, comentado, compartido y leído entre tantos otros, alternativas *altmetrics*. Este proyecto plantea un estudio bibliométrico transversal a todas las Universidad públicas del Noroeste Argentino empleando herramientas y técnicas de recuperación y visualización de

información, dos ramas de la inteligencia artificial que han tenido un gran desarrollo en la última década. Contar con un panorama de estos aspectos no solo reflejará la presencia y popularidad de las investigaciones realizadas por estos científicos, sino también sentará las bases para poder comparar el estado de cada una de las universidades.

**Palabras clave:** *Bibliometría, producción científico-académica, redes sociales académicas, indicadores, altmetrics*

### CONTEXTO

La línea de investigación aquí presentada se encuentra inserta en el proyecto: “*Aplicación de técnicas de Inteligencia Artificial para evaluar la producción científico-académica de investigadores de Universidades públicas del Noroeste Argentino*”, ejecutado a partir de enero de 2018 con una duración de 2 años. Dicho proyecto es llevado a cabo por el grupo de investigación VRAIn (Visualización y Recuperación Avanzada de Información) de la Facultad de Ingeniería de la Universidad

Nacional de Jujuy. El proyecto se encuentra acreditado y financiado parcialmente por la Secretaría de Ciencia y Técnica y Estudios Regionales de la Universidad Nacional de Jujuy (Resolución R. N° 2928/17)

## 1. INTRODUCCIÓN

El crecimiento de la producción científica en las últimas décadas así como su recopilación en bases de datos bibliográficas automatizadas ha potenciado el uso de la "Bibliometría" (Aström, 2007) (White & McCain, 1989) y la generación de indicadores para medir los resultados de la actividad científica y tecnológica. Los indicadores bibliométricos son datos estadísticos deducidos de las distintas características de las publicaciones científicas, en base al importante papel que desempeñan estas en la difusión y transmisión del conocimiento generado en la investigación. Son válidos cuando los resultados de la investigación se transmiten a través de publicaciones científicas y técnicas. Proporcionan información cuantitativa y objetiva sobre los resultados del proceso investigador, su volumen, evolución, visibilidad y estructura, pero no informan sobre los progresos del conocimiento (Mainardi & De Morán-Suárez, 2011).

Una investigación profunda y exhaustiva es llevada a cabo por (Waltman, 2016) sobre los indicadores de impacto de citas, tomando como fuente de datos *Web of Science (WoS)*, *Scopus* y *GS (Google Scholar)*. Este estudio analiza los indicadores más importantes relacionados con el recuento de citas de publicaciones, teniendo en cuenta las diferencias que existen al comparar estos indicadores entre campos distintos, indicando que para ello es necesario un proceso de normalización de datos, ya que reconoce que existen áreas donde la publicación de material es mucho más frecuente que en otras, por ejemplo, un bioquímico con 25 citas no puede considerarse que posee un impacto de citas mayor que un matemático con 10 citas. Lo mismo sucede con el año de publicación,

existen casos en que hay años más prolíficos que otros, y en el mismo sentido el tipo de publicación, sea que se trata de un artículo, una revisión o libro.

Hoy en día el análisis o recuento de citas bibliográficas es el mecanismo más efectivo o al menos el más utilizado para medir la productividad científico-académica, no solo por su sencillez de aplicación sino por su transparencia. Constituye un pilar a la hora de conceder subvenciones y financiamientos de proyectos y en la toma de decisiones, ya que son cada vez más utilizados como una medida de desempeño por el cual los científicos y profesores se clasifican, son promovidos y financiados, además el análisis de citas permite a los investigadores dar seguimiento al desarrollo e impacto de un artículo a través del tiempo (Medrano, 2017).

Según (Capaccioni & Spina, 2012), "La cuestión más debatida es, por supuesto, cuáles son los criterios de evaluación. ¿Criterios únicos o criterios diferentes para las diferentes áreas? ¿Criterios cualitativos o criterios cuantitativos? La importancia sobre la evaluación de la investigación científica se ha incrementado consistentemente en la primera década del nuevo siglo, sumado a la aplicación de algoritmos sofisticados que actualmente se encuentran disponibles para el análisis de citas permitiendo evaluar la "calidad" de estas, permitió la proliferación de un gran número de indicadores, entre ellos dos grupos bien diferenciados, los que se encargan de evaluar una revista científica y los que evalúan la productividad de un científico.

Además de las bases de datos tradicionales para evaluar la productividad científica y además de los distintos indicadores mencionados, existen otro tipo de iniciativas dirigidas a medir el impacto de los materiales publicados online, de forma más amplia que con el recuento de citas. Estas iniciativas son conocidas como *altmetrics* o métricas alternativas (Priem, Taraborelli, Groth, & Neylon, 2017). El término *altmetrics* fue introducido por primera vez en un Tweet en 2010 y más tarde (Priem, Altmetrics, 2014)

amplió la definición para incluir medidas de impacto académicas disponibles en cualquier plataforma en línea.

El conjunto de métricas comúnmente conocida como *altmetrics*, se basan generalmente en la medición de la actividad en línea relacionada con los académicos o los contenidos académicos derivados de las redes sociales y plataformas web 2.0 (Martín-Martín, Orduña-Malea, Ayllon, & López-Cózar, 2016). Sin embargo, la definición de lo que constituye un indicador *altmetrics* está en constante cambio, ya que está determinada en gran medida por las posibilidades técnicas y, más específicamente, la disponibilidad de Interfaces de Programación de Aplicaciones (API) (Haustein, Sugimoto, & Larivière, 2015).

La motivación de este proyecto resulta clara, ya que contar con mecanismo y herramientas para evaluar objetivamente los resultados de la investigación de un científico o grupo de investigación es importante.

Una muestra de la importancia de la evaluación científica son los rankings que se publican de forma anual para establecer el nivel de las universidades, y del mismo modo los rankings de los investigadores más prolíficos. Es más, estos indicadores tanto a nivel personal o institucional resultarían de utilidad al momento de conceder subvenciones y financiamientos de proyectos y en la toma de decisiones, ya que son cada vez más utilizados como una medida de desempeño por el cual los científicos y profesores se clasifican, son promovidos y financiados, ejemplo de ello es la evaluación a través de los sexenios de investigación de universidades europeas. Este instrumento, el sexenio investigador, es considerado en la mayor parte de los ámbitos científicos como una medida incuestionable de la calidad de la actividad investigadora, dado su elevado grado de aceptación por la comunidad universitaria (García-Berro, Amblàs, Sallarés, Bugada, & Roca, 2013).

Por esta razón, conocer el estado actual de cada universidad del NOA en cuanto a la cantidad de publicaciones y al impacto de estas, ofrecerá un

panorama adecuado además de marcar el camino para poder comparar cada institución con el resto de universidades del país.

En estos últimos años evaluar y cuantificar la productividad científica se ha convertido en tema de un profundo análisis ya que las cantidades de información aumentan constantemente y se vuelven cada vez más dinámicas (Miguel & Dimitri, 2013). En este momento existen variados métodos cuantitativos para medir esta producción, entre los más conocidos y usados destacan: el factor de impacto y *h-index* (Hirsch, 2005). Si bien estos números reflejan la relación entre artículos publicados y la popularidad de los mismos, entiéndase esto en términos de la cantidad de veces que un artículo es citado por otro o la importancia asignada a la revista en donde se publique, muy a menudo no resultan suficientes o precisos.

Los motores de búsqueda académicos de libre acceso como *Google Scholar* y *Microsoft Academic*, en contraposición a las bases de datos bibliográficas tradicionales como *Scopus* o *Web of Science*, presentan ciertas ventajas: El rango de cobertura es amplio, no solo artículos de revistas y congresos sino libros, tesis, informes y muchos otros tipos de documentos, son herramientas de libre acceso y gratuitas, no están circunscritos a determinadas áreas del conocimiento, y cada vez el alcance es mayor hacia áreas de menor relevancia, son muy populares y de fácil uso, son útiles para cualquier trabajo académico sin importar la envergadura de este.

Y para darle un mayor significado a estos indicadores, poder evaluar la presencia y la utilidad que se le da a estas publicaciones en redes sociales académicas, propone un enfoque desde otro ángulo. Hoy en día, ambos enfoques son distintos pero complementarios, valerse del uso de solo un tipo de indicadores no entregaría toda la información o solo se limitaría a evaluar un aspecto, por ello la necesidad de contar con los dos protagonistas de la historia, por un lado los indicadores bibliométricos tradicionales encargados de contabilizar la cantidad de

publicaciones y citas recibidas, y por otro lado los indicadores alométricos encargados de evaluar el impacto en redes sociales de dichas publicaciones.

## 2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Esta línea de investigación propone estudiar, evaluar y aplicar distintas técnicas de inteligencia artificial para evaluar la producción de los científicos de universidades públicas del NOA. Para ello se pretende utilizar métodos de recuperación de información y de visualización de información para conformar una base de datos tanto de los investigadores como de las publicaciones, realizar una limpieza de los mismos (empleando técnicas de detección de duplicados, fusión de citas) y poder obtener nuevo conocimiento a partir de ello, realizando comparaciones, inferencias y predicciones. En este sentido se emplearán algoritmos de estas áreas de la inteligencia artificial para lograr el propósito deseado.

## 3. RESULTADOS OBTENIDOS/ESPERADOS

El presente proyecto se centra en conocer el estado actual de la producción científico-académica de los investigadores de las universidades públicas del Noroeste Argentino

Particularmente se espera lograr:

- Analizar las técnicas de recuperación automática de información.
- Conformar una base de datos de publicaciones de investigadores libre de sesgos y normalizada.
- Conocer el estado actual y la presencia de investigadores en redes sociales académicas.
- Desarrollar y adaptar las técnicas de visualización de información para presentar los resultados obtenidos.
- Desarrollar una aplicación que permita procesar y evaluar los datos recolectados.

- Calcular indicadores bibliométricos y alométricos sobre los datos procesados.
- Realizar predicciones y obtener nuevo conocimiento a partir de los datos existentes.
- Evaluar el alcance de los resultados de la investigación, elaborar rankings y publicar las conclusiones alcanzadas.

## 4. FORMACIÓN DE RECURSOS HUMANOS

Durante la realización de este proyecto se espera lograr, como mínimo, la culminación de 1 tesis de maestría y 2 tesis de grado dirigidas y/o codirigidas por los integrantes del proyecto y relacionadas con la temática.

Una tesis doctoral, relacionada ampliamente con la línea de investigación presentada (Medrano, 2017), emplea varias técnicas y herramientas que se utilizarán en la implementación de este proyecto.

Finalmente, es constante la búsqueda hacia la consolidación como investigadores de los miembros más recientes del grupo.

## 5. BIBLIOGRAFÍA

- Aström, F. (2007). Changes in the LIS Research Front: Time-Sliced Cocitation Analyses of LIS Journal Articles, 1990-2004. *Journal of the American Society for Information Science and Technology*, 58(7), 947-957.
- Capaccioni, A., & Spina, G. (2012). Italian SSH journals in Journal Citation Reports (JCR) and in SCImago Journal Rank (SJR): data and first analysis. *Italian Journal of Library & Information Science*, 3(1), 4787-4807.
- García-Berro, E., Amblàs, G., Sallarés, J., Bugada, G., & Roca, S. (2013).

- Docencia e investigación: ¿un falso dilema? *Aula Abierta*, 41(2), 13-22.
- Haustein, S., Sugimoto, C. R., & Larivière, V. (2015). Guest editorial: Social media metrics in scholarly communication. *Aslib Journal of Information Management*, 67(3).
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 16569-16572.
- Mainardi, C. F., & De Morán-Suárez, M. A. (2011). La responsabilidad social corporativa (rsc) en las bases de datos scopus y wos (estudio bibliométrico). *EDICIC*, 1(4), 141-160.
- Martín-Martín, A., Orduña-Malea, E., Ayllon, J. M., & López-Cózar, E. D. (2016). The counting house: measuring those who count. presence of bibliometrics, scientometrics, informetrics, webometrics and altmetrics in the google scholar citations, researcherid, researchgate, mendeley & twitter. *CoRR*.
- Medrano, J. F. (2017). *Evaluación de la producción científica mediante motores de búsqueda académicos y de acceso libre*. Tesis doctoral, Universidad de Salamanca, Informática y Automática, Salamanca.
- Miguel, S., & Dimitri, P. J. (2013). La investigación en bibliometría en la Argentina: quiénes son y qué producen los autores argentinos que realizan estudios bibliométricos. (29), 117-138.
- Priem, J. (2014). Altmetrics. (B. Cronin, & C. R. Sugimoto, Edits.) *Beyond bibliometrics: harnessing multidimensional indicators of performance*, 263-287.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (20 de 08 de 2017). *Altmetrics: A manifesto*, (v.1.0). Obtenido de <http://altmetrics.org/manifesto>
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Informetrics*, 10(2), 365-391.
- White, H., & McCain, K. W. (1989). Bibliometrics. *Annual Review of Information Science and Technology*, 24, 119-186.