



EL ACADÉMICO DIGITAL: ¡LAS REVOLUCIONES QUE ESTÁN POR VENIR!



JÉRÉMY BARBAY

Profesor Asistente del Departamento de Ciencias de la Computación de la Universidad de Chile. Doctor en Computación por la Universidad de Paris XI, Orsay, Francia. Sus intereses de investigación incluyen el análisis adaptativo de algoritmos y estructuras de datos, sistemas de control de calidad colectivos y evaluación de técnicas pedagógicas. Ha liderado iniciativas dirigidas a la implementación de nuevas técnicas pedagógicas como el proyecto teachingislearning.cl
jeremy@barbay.cl



VANESSA P. ARAYA

Investigadora de Postdoctorado en el equipo ILDA del centro de Inria Saclay, Francia. Sus intereses de investigación están principalmente centrados en el área de visualización de la información y procesamiento de datos, con especial énfasis en datos geotemporales. Terminó su Doctorado en Computación el año 2018 en la Universidad de Chile.
vnaraya@protonmail.com

Cada área de investigación tiene sus necesidades y herramientas, pero cualquier resultado de investigación tiene que ser *comunicado* al resto de la comunidad, para ser *validado*, ojalá *reproducido*, y ulteriormente *mejorado* y *aplicado*. La comunicación entre personas partió desde tradiciones orales, pasando por manuscritos y libros impresos hasta llegar a los medios digitales de la actualidad (incluyendo textos, pero también videos y software interactivo). En este artículo describiremos cómo la evolución de los medios que almacenan y transmiten información ha permitido cambios que nos parecen interesantes sobre la *comunicación* y *reproducción* de resultados, además de su validación.

La docencia trata de comunicar conocimientos y permitir a los estudiantes *reproducir* los procesos enseñados. Tales tareas son similares a las descritas para la investigación pero a escala completamente distinta (e.g. la cantidad de estudiantes tomando cursos del mismo

docente, y la cantidad de profesores enseñando el mismo tema). En este artículo, describiremos cómo la digitalización de los medios docentes permite una *colaboración* entre éstos a escalas inimaginables con documentos impresos y, en particular, cómo esta colaboración permite mejorar la enseñanza a través de la comparación científica entre técnicas pedagógicas. Finalmente, nos referiremos a cómo la digitalización del material docente facilita la *personalización* de la docencia.

Aunque muchas veces ignoradas, las tareas administrativas de un profesor universitario conllevan una alta carga de trabajo cuyo potencial de automatización ha sido mayormente ignorado. Con tecnologías de base de datos y de firmas electrónicas ya disponibles, la digitalización de los medios de publicación permite una generación automatizada de los reportes de "productividad", permitiendo que sean más fiables y mejor distribuidos. A su vez, el análisis de tales reportes permitirá medir y *comparar*

el impacto de incentivos y otras medidas administrativas, permitiendo la detección de incentivos defectuosos, ya sea por mal diseño, corrupción o baja escalabilidad.

La aparición de una nueva tecnología raramente genera una revolución inmediata. En el mejor caso, las nuevas tecnologías se acumulan hasta que un conjunto de ellas permiten cambios mayores, mientras que en el peor caso sociedades enteras se quedan bloqueadas en una vía sin salida y son derrotadas por otras sociedades [1]. La evolución del libro es un buen ejemplo de esta demora: los primeros libros transcribieron relatos orales, y los primeros libros impresos replicaron biblias que habían sido copiadas a mano por siglos [2], inicialmente sin hacer cambios significativos en su contenido. Las tecnologías digitales no están exentas de tal retraso y bloqueo, pero con este artículo esperamos comunicar a profesores universitarios acerca de los cambios que de a poco van permeando la sociedad y así reducir la latencia

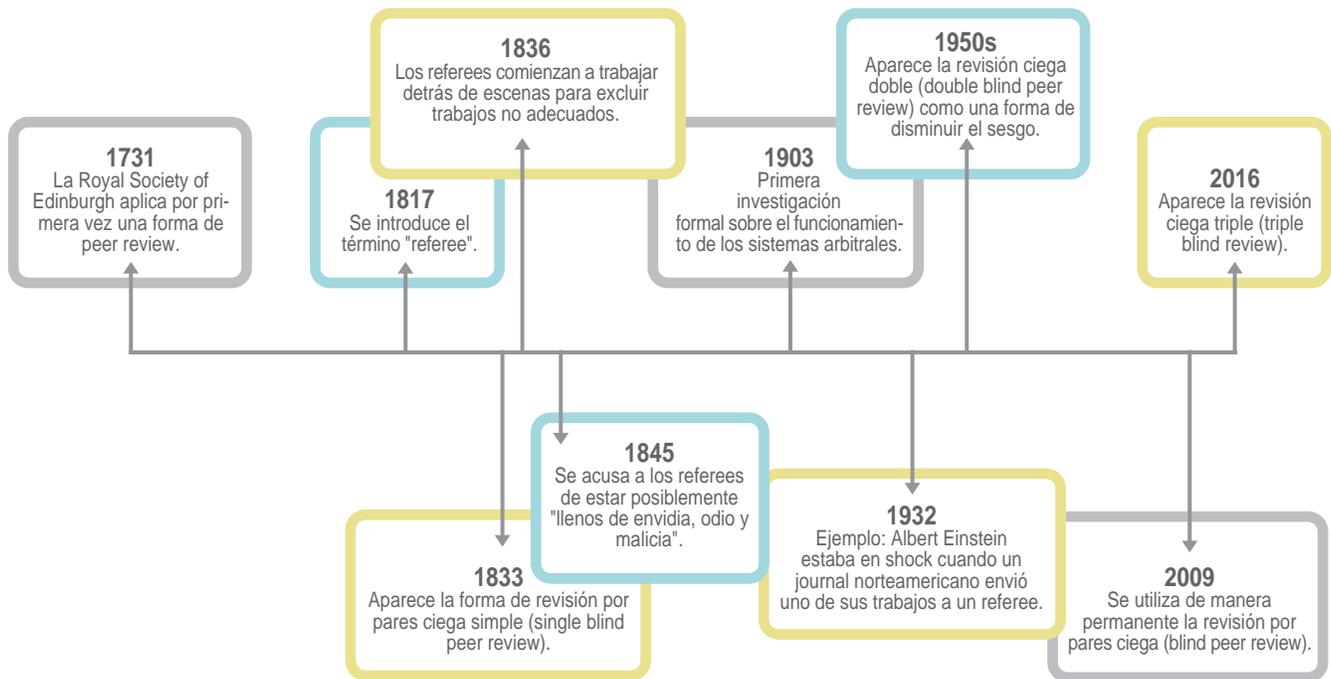


Figura 1. Algunos eventos relevantes en la historia de los sistemas de revisión por pares.

de su adopción, ya que procesos de investigación, docencia y administración más eficientes pueden tener un impacto importante en el trabajo diario de un profesor universitario y, aún más, sobre la sociedad.

INVESTIGACIÓN

Para aportar en algún área de la ciencia, primero que nada es necesario conocer el estado del arte, es decir, leer los reportes temáticamente relacionados, realizados por otros científicos. Un investigador que busca un artículo científico de otro académico, puede encontrarse con dos problemas principales. El primero, es que la cantidad de publicaciones está en crecimiento constante, por lo que encontrar y filtrar publicaciones relevantes puede ser difícil en un volumen tan grande de artículos. En esta misma línea, cada vez se hace más difícil para un investigador (¡o docente!) estar al día con los últimos resultados de su área. La era digital ha sido buena para reducir este problema: sitios como Research Gate [11] o Google Scholar [8] automatizan muchas tareas, como notificar sobre nuevas publicaciones, citando artículos de interés o sugiriendo publicaciones previas relacionadas con un tema de interés. El segundo problema, es que el monopolio de los editores de revistas y los altos costos económicos que exigen para

acceder a sus publicaciones, hacen que la ciencia se vea restringida a unos pocos. En respuesta a este problema han surgido varias iniciativas alternativas de publicaciones científicas, yendo desde Scientific Research Open Access [13] hasta el proyecto Sci-Hub [12], más controversial.

Una vez decidida la idea a seguir, la tarea de realización y comunicación de un trabajo científico puede dividirse en dos etapas: pre y post realización de una investigación. Para la primera etapa, se está volviendo común en muchas áreas de la ciencia el proceso de preregistro a través de sitios como Center for Open Science [3]. Tal proceso consiste en dejar registrado el diseño de un estudio o experimento, incluyendo el análisis que se pretende llevar a cabo, *antes* de iniciar el experimento. Además de ser más riguroso (e.g. prohíbe realizar análisis adicionales si los primeros no dieron los resultados esperados, una práctica de riesgos científicos), el proceso tiene la ventaja de motivar al investigador a pensar detalladamente su investigación. En la postrealización de una investigación, sus resultados, procesos y datos deben quedar accesibles a otros investigadores, de manera de facilitar su reproducción, confirmación (y mejora) por pares. Herramientas de *literate programming* como Jupyter Notebooks [10], que permiten mezclar código ejecutable (en varios lenguajes de programación o scripts analizando

datos y generando gráficos) en el mismo texto de un reporte científico, deberían facilitar tal reproducción en el futuro.

Finalmente, un trabajo científico debe ser evaluado por otros pares para ser publicado. Aunque los procesos de revisión por pares se consideran esenciales para una investigación científica de calidad, éstos no son tan antiguos como la investigación misma: mientras una forma de revisión por pares apareció por primera vez en 1731, el protocolo de revisión por pares con anonimato de los revisores llegó en 1833, y los mayores problemas asociados ya estaban identificados en 1845 (ver **Figura 1**), tales como el sesgo de género, de raza, falta de control de la calidad, y variaciones sobre la manera de hacer revisión por pares (ver **Figura 2**). Tecnologías existentes pueden apoyar estos procesos como, por ejemplo, el alto nivel de anonimato que permiten los nuevos métodos de certificados encriptados. Este anonimato podría expandirse a los reportes publicados a través de un sistema de certificados firmado, con lo que los investigadores podrían elegir con quién compartir (y certificar) su autoría de cada reporte, resolviendo el problema presente de sesgo (*bias*) de género y raza, y potencialmente problemas futuros de sesgo en las citas.

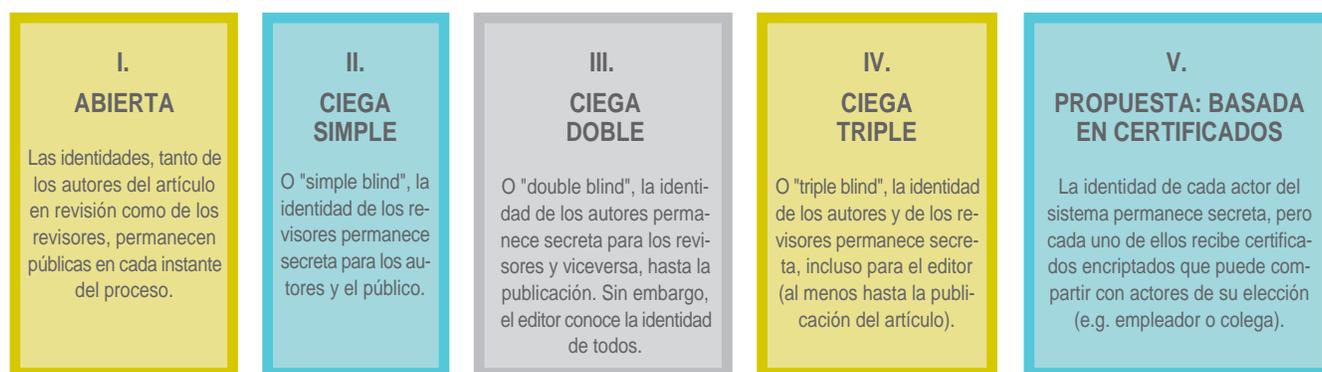


Figura 2. Tipos de revisión por pares.



EDUCACIÓN

Tradicionalmente, las tareas y exámenes son generados localmente en cada universidad, a veces con el apoyo de libros de referencia, pero en la mayoría de los casos con un "toque" personal del equipo docente. Mientras este enfoque se justificaba en la era de papel, en la era digital no tiene mucho sentido: las diferencias de currículum entre universidades no son muy grandes, y una buena tarea o un ejercicio de examen puede ser usado por varias instituciones. Herramientas como EasyExam [5], EasyProgramChecking [6] o Judge [9] inician un movimiento para hacer comunitarios los recursos de varias universidades para diseñar currículos, tareas y exámenes, y la colaboración docente entre instituciones del futuro.

Con la reducción de los costos de producción de material docente digital, queda la oportunidad de resolver un problema antiguo de la docencia: el alto costo de la docencia personalizada hace que quede accesible a una minoría de elite. Aunque los cursos masivos ("Massive Online Open Courses" en inglés, identificados por el acrónimo MOOCs) han intentado resolver este problema, los resultados muestran que no han sido realmente efectivos. Así, al tener una capacidad de producción (y de reuso) más grande, podemos esperar la aparición de sistemas de educación más especializados para diversas características de estudiantes como por ejemplo, edad, cultura o nivel educacional. En una era donde se espera que los trabajadores renueven constantemente sus conocimientos después de salir del sistema escolar, el rol de dichos sistemas personalizados puede solamente crecer: los conocimientos adquiridos mientras se trabaja en varias empresas ("hacer es aprender") después de salir del ámbito académico, pueden solamente hacer crecer la diversidad de "background" de alumnos-trabajadores comparado con "alumnos académicos".

Las iniciativas mencionadas anteriormente permitirán una diversificación del material docente que requerirán de metodologías do-

centes distintas a las que usamos actualmente, y, por lo tanto, de investigación que las diseñe y evalúe. La simplicidad inicial de los cursos masivos en línea (MOOCs) ofrecidos por empresas como Coursera [4] y Edx [7] permitió explorar (validando algunas, e invalidando otras) hipótesis básicas de docencia a través de "A/B testing" que posibilitaron estructuras y herramientas más complejas. Desde su creación, dichos proyectos han sido regularmente presentados como competidores de las universidades (y, por tanto, como una amenaza al trabajo del profesor digital). En los hechos, dichos cursos sufren de problemas importantes en términos de retención de alumnos y de calidad del material docente, por una mezcla de causas psicológicas (e.g. la experiencia universitaria es también una experiencia social) y económicas (e.g. problemas con los derechos de autores del material producido). El futuro de los MOOCs es aún incierto, pero parece previsible que la formalización de la investigación en docencia que iniciaron estos proyectos seguirá desarrollándose.

ADMINISTRACIÓN

La generación de los reportes, cualquiera sea su tipo, conlleva un alto costo, pero puede ser automatizada mucho más de lo que está en la actualidad. Es más costoso aún en el caso de las instituciones académicas, las cuales persiguen objetivos múltiples como investigación y educación, además de otras actividades como, por ejemplo, consultorías sobre políticas públicas. Anteriormente, en la era en que los reportes eran escritos y las instituciones eran más pequeñas, dichos reportes o no existían o eran muy limitados. Hasta ahora los reportes de la era digital son equivalentes a los reportes en papel, pero en volúmenes más grandes y más redundantes. La revolución en este aspecto será un acto político ya que no se necesita nueva tecnología. Establecer un estándar universal para la generación de tal información permitirá que cada académico pueda tener la información de su trabajo de manera centralizada, de tal manera que la institución a la que pertenece pueda acceder a ellos para generar

reportes de manera automática, reduciendo así el costo de trabajo de los académicos.

El costo de verificar la información entregada en tales reportes ya es un problema. Éste crecerá más aún con el desarrollo de la academia (en tamaño, en complejidad y en competitividad), mientras estándares digitales permitirán más y mejor comunicación, a menor costo. Técnicas para que un investigador certifique una publicación en una revista prestigiosa, o para que un docente certifique un premio docente, son similares en esencia a las técnicas ya usadas para que un vendedor de software certifique el origen legal de su producto, o para que un servidor DNS certifique sus respuestas a consultas DNSsec. Tales soluciones están basadas en técnicas criptográficas bien establecidas como pares de claves pública/privada (RSA) por lo que será "solamente" una decisión política por parte de las instituciones el usar tales certificados para entregar certificados automáticamente verificables a agentes (e.g. investigador, docente), quienes a su vez entregarán tales certificados a las instituciones solicitantes (e.g. empleador a la fecha, o empleador potencial). La evaluación humana seguirá siendo necesaria para evaluar la calidad de las contribuciones científicas de un candidato, pero revisar si tal candidato no ha puesto dos veces la misma publicación, si no ha "inventado" una publicación o ha declarado una publicación que no es suya, debería hacerlo un computador, no un humano.

Más allá de la rutina diaria de los reportes y de su verificación, una administración eficiente tiene que poder evaluar situaciones pasadas y presentes, a fin de poder planificar el futuro. El almacenamiento de grandes volúmenes de datos confiables de instituciones académicas con distintos incentivos y políticas permitirá comparar experimentalmente el impacto de tales incentivos y políticas con un nivel de detalle nunca realizado hasta ahora. Con tales herramientas se podrán identificar incentivos con efectos perversos, o comportamientos ilegales o poco éticos que se aprovechan de algunas políticas, y remediar tales situaciones con los cambios apropiados.

CONCLUSIONES

Las tecnologías digitales, aunque en gran parte desarrolladas por académicos, no se han aplicado a las tareas de éstos con la misma extensión que en la industria.

Hemos descrito varios aspectos (educativo, investigativo y administrativo) de la revolución digital por venir en la carrera de un profesor, basados tanto en el bajo costo y en el potencial para masificar las tecnologías de tratamiento digital de la información, como en las técnicas que impactan tal trabajo de manera más indirecta, como protocolos de control de calidad colectivos (e.g. ReCaptcha, Calibrated Peer Review), técnicas criptográficas (e.g. claves asimétricas, firmas electrónicas) y personalización (e.g. big data).

Deliberadamente, hemos dejado de lado técnicas con impacto que nos parecen menos específicas a la carrera académica, como *machine learning* e inteligencia artificial, enfocándonos en tecnologías digitales maduras, y dejando la descripción del impacto esperado de tecnologías más nuevas a sus expertos respectivos.

Sin pretender ser expertos en todos los temas descritos en este reporte, esperamos que tal descripción de técnicas bien establecidas para solucionar —de manera completa o parcial— problemáticas existentes en el mundo académico permita a profesores universitarios jóvenes orientar algunas de su decisiones al punto de organizar su carrera y su entorno de trabajo, y a instancias universitarias orientar algunas de sus inversiones tecnológicas tanto en la administración como en la investigación y la docencia. ■

REFERENCIAS

- [1] G. Berns, *Iconoclast*. Harvard Business Review Press (March 17, 2010), 2010. ISBN: 978-1422133309.
- [2] J. Diamond, *Guns, Germs, and Steel: The Fates of Human Societies*. W. W. Norton Company; 1st edition (April 1, 1999), 1999. ISBN-10: 0393317552, ISBN-13: 978-0393317558.
- [3] “Center for open science: Preregistration”. <https://cos.io/prereg/>
- [4] “Coursera”. <https://www.coursera.org/>
- [5] “Easy exam”. <https://easyexam.repositorium.cl>
- [6] “Easy program checking”. <https://easyprogramchecking.repositorium.cl>
- [7] “Edx”. <https://www.edx.org/>
- [8] “Google scholar”. <https://scholar.google.cl/>
- [9] “Jutge”. <https://jutge.org>
- [10] “Project Jupyter Notebook”. <https://jupyter.org/>
- [11] “Research gate”. <https://www.researchgate.net>
- [12] “Sci-hub”. <http://scihub.org>
- [13] “Scientific research open access”. <https://www.scirp.org/journal/OpenAccess.aspx>