

# Declaración de La Haya sobre la extracción de conocimiento útil (*knowledge discovery*) en la Era Digital

Traducción al español de Eva Méndez ([emendez@bib.uc3m.es](mailto:emendez@bib.uc3m.es))

Universidad Carlos III de Madrid.

Eva Méndez es uno de los [participantes originales de la Declaración de la Haya](#)

## Visión

Las nuevas tecnologías están revolucionando la forma en que los humanos aprendemos sobre el mundo y sobre nosotros mismos. Estas tecnologías no son sólo un medio para hacer frente a grandes volúmenes de datos (*Big Data*<sup>1</sup>), también son la clave para la extracción de conocimiento útil [*knowledge discovery*] en la era digital; y su potencial se basa en la disponibilidad, cada vez mayor, de los propios datos. Factores como el aumento progresivo de la capacidad de procesamiento informático, el crecimiento de la Web, así como el compromiso gubernamental con el acceso abierto<sup>2</sup> a la investigación financiada con fondos públicos, están sirviendo para que la disponibilidad de hechos, datos e ideas sea cada vez mayor.

Sin embargo, los marcos legislativos actuales, en diferentes jurisdicciones, no siempre contribuyen a apoyar la utilización de nuevos enfoques para realizar la investigación, en particular en lo que se refiere a metodologías de minería de contenido. La minería de contenido [*content mine*] es el proceso de obtener o inferir información a partir de material legible por máquina. Funciona copiando grandes cantidades de material, extrayendo los datos, y recombinándolos para identificar patrones y tendencias y, en definitiva, descubrir nuevo conocimiento [*knowledge discovery*]

Al mismo tiempo, las leyes de propiedad intelectual, previas y más antiguas que la web, limitan el potencial de las técnicas de análisis de contenido digital como son la minería de datos y textos (para textos y datos) o minería de contenido (para análisis automatizado de contenido en todo tipo de formatos)<sup>3</sup>. Estos factores están produciendo al mismo tiempo, desigualdades a la hora de extraer conocimiento útil (*knowledge discovery*) en la era digital. La legislación en cuestión, puede ser la ley de derechos de autor, el derecho que rige las patentes o las leyes de bases de datos - todo lo que puede limitar la capacidad del usuario para realizar un análisis de contenido detallado.

Los investigadores deben tener la libertad de analizar y de dejarse llevar por su curiosidad intelectual sin miedo a ser vigilados o a las repercusiones que eso pueda tener. Estas libertades no deben erosionarse o reducirse en el entorno digital. Asimismo, la ética en torno al uso de los datos y la minería de contenido debe seguir evolucionando a tenor de los cambios tecnológicos.

El análisis informatizado del contenido en cualquier formato, que es lo que implica la minería de contenido, facilita el acceso a un conocimiento público desconocido, que aún no se ha extraído, y proporciona ideas importantes para todos y cada uno de los aspectos de nuestra vida económica, social y cultural (por ejemplo, la predicción de levantamientos políticos, el análisis de los cambios

---

<sup>1</sup> Para el Estatuto del G8 sobre Open data, ver (contenido del enlace en inglés)

<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>. Ver también el Informe sobre la extracción de datos de la RDA (contenido del enlace en inglés) [http://europe.rd-alliance.org/sites/default/files/report/TheDataHarvestReport\\_%20%20Final.pdf](http://europe.rd-alliance.org/sites/default/files/report/TheDataHarvestReport_%20%20Final.pdf).

<sup>2</sup> En relación a la declaración de Berlín, ver <http://openaccess.mpg.de/Berlin-Declaration>; véase también la Iniciativa de acceso abierto de Budapest 10 años después (texto del enlace en inglés) <http://www.budapestopenaccessinitiative.org/boai-10-recommendations>.

<sup>3</sup> Siempre que utilizamos en esta declaración "minería de textos y datos" (TDM por sus siglas en inglés), lo utilizamos para referirnos a la extracción de todas las formas de datos independientemente de que sean en texto, imágenes, grabaciones o películas sonoras.

demográficos). El uso de dichas técnicas tiene el potencial de revolucionar la forma en que hacemos investigación, tanto a nivel académico, como comercial.

### **Beneficios de la minería de contenido**

Los beneficios potenciales de la minería de contenido son inmensos. Algunos de ellos son:

- Abordar grandes retos como el cambio climático y las epidemias globales
- Mejorar la salud, la riqueza y el desarrollo de la población
- Crear nuevos puestos de trabajos y empleo
- Aumentar exponencialmente la velocidad y el progreso de la ciencia a través de nuevas ideas y de una mayor eficiencia de la investigación
- Aumentar la transparencia de los Gobiernos y sus acciones
- Fortalecer la innovación y la colaboración y aumentar el impacto de la ciencia abierta
- Crear herramientas para la educación y la investigación
- Proporcionar nuevas percepciones de la cultura más ricas
- Acelerar el desarrollo económico y social en todas las partes del mundo

Investigadores, PYMES (Pequeñas y Medianas Empresas) y grandes compañías tecnológicas han estado haciendo minería de contenidos, por lo menos en los últimos 10 años. Sin embargo, en esta era de oportunidades, es importante que todos los miembros de la sociedad puedan beneficiarse igualmente de los avances que ha experimentado la tecnología digital y la disponibilidad del contenido. Para que esto sea posible, se requiere la creación de nuevos principios en torno al acceso a hechos, datos e ideas.

Así, los firmantes de esta declaración, en reconocimiento de los enormes beneficios económicos y sociales de la extracción de conocimiento útil (*Knowledge Discovery*) en la era digital, establecemos estos principios:

1. ***La propiedad intelectual no fue diseñada para regular el libre flujo de hechos, datos e ideas, sino que tiene como objetivo fundamental la promoción de la actividad investigadora.***

El libre flujo de información e ideas es un derecho fundamental<sup>4</sup>. Es un catalizador para la producción del conocimiento humano, que afianza el bienestar y la prosperidad. Las sociedades de todo el mundo han optado por proteger ciertos derechos limitados de propiedad intelectual como incentivos tanto de la innovación, como de la difusión del conocimiento.

Las leyes de propiedad intelectual nunca han estado destinadas a cubrir hechos, ideas y datos puros. Sin embargo la interpretación moderna de las leyes de propiedad intelectual se está convirtiendo cada vez más en un obstáculo para la creación y la diseminación del conocimiento, aplicándose incluso a los elementos constitutivos más sencillos del conocimiento (hechos, ideas, datos).

En algunos países, el derecho de autor<sup>5</sup> en particular, se ha interpretado de tal manera que restringe la posibilidad de aplicar la lectura y el análisis informático al contenido disponible legalmente de otra forma.

Otros contextos legislativos como el de la ley de patentes y la ley de bases de datos pueden tener un impacto similar. Si la ley de propiedad intelectual permite que el contenido se lea y analice “manualmente” por los seres humanos, pero no por sus máquinas, algo falla, no ha cumplido sus objetivos originales.

2. ***La libertad de analizar y de dejarse llevar por la curiosidad intelectual sin miedo a ser vigilado o a las repercusiones.*** Los proveedores de contenidos deberían respetar la privacidad intelectual

---

<sup>4</sup> (contenido del enlace en inglés) <http://www.un.org/en/documents/udhr/index.shtml#a19>

<sup>5</sup> Y en la ley de la Unión Europea de bases de datos también en el formulario de la Directiva de bases de datos.

de los lectores individuales y deberían tomar las medidas necesarias para proteger la privacidad de dichos lectores, ante la interferencia de cualquier agente externo. Cualquier excepción, que, por ejemplo ocasionara la invasión de la privacidad individual, deberá proporcionarse y estar prevista por la ley. El uso de hechos, datos e ideas no debe perjudicar en ningún caso, el derecho a la intimidad y a la vida privada, derechos inalienables de las personas.

3. **Las licencias y términos de los contratos no deberían restringir el uso de hechos datos e ideas.** Normalmente las licencias y los términos de los contratos, que regulan y restringen cómo los individuos pueden analizar y utilizar hechos, datos e ideas, son inaceptables e inhiben la innovación y la creación de nuevo conocimiento, y por tanto, no deberían adoptarse. De forma similar, es inaceptable que las medidas técnicas en los sistemas de gestión de derechos digitales (DRM, en inglés) deban inhibir el derecho legítimo para llevar a cabo la minería contenido.
4. **La ética en torno al uso de técnicas de minería de contenidos necesita continuar su evolución, en respuesta al cambio tecnológico.** El estudio de las normas éticas bien establecidas en el ámbito de la investigación y de los negocios, así como el continuo desarrollo de tales normas y leyes, debe de apoyarse y motivarse de tal forma que se desplieguen tecnologías de minería de contenido en beneficio de la sociedad.
5. La Ley de propiedad intelectual no debería restringir la innovación y la investigación comercial en el uso de hechos, datos e ideas

En tanto que los hechos, datos e ideas no son susceptibles de derechos de autor, no tiene sentido restringir el uso comercial ético de aquellos hechos, datos e ideas extraídos de un contenido que se ha obtenido legalmente. Aunque la ley de patentes está diseñada para proteger innovaciones e invenciones, esto no significa que abarque también hechos y datos. Las restricciones en el uso de hechos, datos e ideas pueden tener un impacto importante en la innovación y en el desarrollo económico a nivel mundial. También puede reducir la capacidad de utilizar herramientas y procesos que pueden beneficiar a los ciudadanos en los ámbitos de la salud, la ciencia, el empleo, la investigación, el medio ambiente y la cultura.

## Hoja de ruta para la acción

### Aceptar la visión y los principios

1. Esta declaración recoge la visión de que la propiedad intelectual no fue diseñada para regular el libre flujo de hechos, datos e ideas, sino que tiene el objetivo de promocionar las actividades de investigación.
2. Aunque el marco de los derechos de autor actuales no son compatibles con esta visión, los legisladores deben trabajar inmediatamente para apoyar la introducción de los cambios que permitirían a los usuarios llevar a cabo la minería de contenido en los materiales a los que tienen acceso legal.
3. Cuando se introduzcan excepciones o limitaciones<sup>6</sup> a la ley de derecho de autor para permitir la minería de contenido, deberían ser obligatorias, y no poder ser anuladas por los contratos.
4. Es inaceptable que las medidas técnicas en los sistemas de gestión de derechos digitales puedan inhibir el derecho legal a realizar la minería de contenido.
5. No debería de ser necesario firmar licencias separadas para llevar a cabo actividades de minería de contenido, en tanto que el derecho a leer es el derecho a realizar minería de contenido, donde las actividades de minería que se realizan ya tienen acceso legal a los contenidos relevantes.

---

<sup>6</sup>Ver, por ejemplo (texto del enlace en inglés) <http://www.wipo.int/copyright/en/limitations>.

## Políticas e incentivos

6. Los responsables políticos deben tratar de proporcionar claridad jurídica asegurando que la minería de contenido no constituye una infracción de los derechos de autor u otros derechos conexos. Creemos que el derecho a leer incluye el derecho a la minería, pero sólo cuando los individuos tengan acceso legal a ese contenido.
7. Cuando las instituciones que financian la investigación u otros organismos lo requieran, y cuando los autores así lo deseen, los resultados de investigación pueden ponerse a disposición bajo licencias específicas, que normalmente deberían ser CC-BY para las publicaciones y CC0 para los datos de la investigación<sup>7</sup>.
8. Todas las universidades, organizaciones de investigación, entidades financiadoras de la investigación y negocios comerciales deben asegurar que sus políticas abogan por la minería de contenidos como metodología de investigación que tiene el potencial de transformar la forma en que se realiza la investigación. El crecimiento del acceso abierto y de los datos abiertos, ha sido y continuará siendo un elemento clave para facilitar la minería de contenido.
9. Los investigadores deben reconocer el derecho de los autores de las publicaciones a ser reconocidos como tales, y a ser respetados y reconocidos como productores de los datos cuando sea posible.
10. Las políticas de minería de contenidos deben respetar el derecho legítimo de los autores y editores, y ser impulsadas a su vez por las necesidades de los investigadores y de las empresas en la era digital.
11. Para motivar la aceptación de la actividad de minería de contenidos, universidades, organismos de investigación, financiadores de la investigación y empresas, deben considerar la introducción de incentivos para recompensar a quienes utilicen estas nuevas técnicas, por ejemplo, las actividades de minería de contenido deben tenerse en cuenta y valorarse positivamente en los procesos de evaluación.

## Infraestructura y herramientas

12. Las organizaciones de investigación, universidades y empresas deben asegurar el mantenimiento y desarrollo de infraestructuras de depósito que proporcionen el almacenamiento de, y el acceso a, publicaciones legalmente disponibles para minería de contenido. Los investigadores independientes, deben tratar de utilizar dichas facilidades [*content mining*] siempre que sea posible.
13. Los organismos de investigación, universidades y empresas deben proporcionar acceso a las infraestructuras adecuadas para que los datos de investigación estén disponibles para la minería de datos, cuando sea legal y éticamente posible. Los investigadores independientes, deben tratar de utilizar dichas facilidades [*content mining*] siempre que sea posible.
14. Todos estos desarrollos implican infraestructuras tecnológicas, estándares, normas éticas y requisitos de financiación para poner los resultados de investigación en abierto.
15. Siempre que sea posible se deben usar estándares abiertos tales como XML y JSON para el transporte de datos, ORCID para la identificación de autores y licencias CC (Creative Commons) para las necesidades de licencias abiertas.
16. Si un material está disponible bajo licencia CC BY, los creadores de contenidos deben permitir descargar: el XML u otro formato de fichero de alta calidad, imágenes de alta

---

<sup>7</sup> Por ejemplo, la iniciativa de Ciencia Abierta e Investigación en Finlandia ha publicado una hoja de ruta en este sentido (contenido del enlace en inglés) <http://openscience.fi>. Recomienda una licencia CC4.0 BY para todos los resultados de investigación. Para los metadatos de los repositorios, la recomendación nacional es CC0.

resolución, los datos que apoyan a las imágenes, y el artículo aceptado sin formato por parte del autor.

### **Apoyo**

17. Organismos como universidades, centros de investigación, asociaciones de bibliotecas, la comunidad médica, las empresas, los miembros de la comunidad científica de la minería de contenidos deben defender y apoyar los beneficios de la minería de contenidos.
18. Las bibliotecas de investigación están en condiciones de asumir esta función de promoción como parte de sus actividades de apoyo a la investigación.
19. Las bibliotecas deben proporcionar formación a los investigadores en minería de contenidos, incluyendo la asesoría legal.

### **Desarrollo y seguimiento de la declaración de la Haya y Hoja de ruta**

20. LIBER (La Asociación de las Bibliotecas Europeas de investigación) ha sido fundamental en el trabajo con los agentes implicados en la Declaración de la Haya y el diseño de una hoja de ruta para su desarrollo. LIBER continuará dando seguimiento y supervisando el progreso para obtener la firma y refrendo de la Declaración, y defenderá la implementación de esta hoja de ruta.