

I N S T I T U T O
B R A S I L E I R O
D E D I R E I T O S
A U T O R A I S

INOVAÇÃO, INTELIGÊNCIA ARTIFICIAL e DIREITOS AUTORAIS: *A essencialidade da Mineração de Textos e Dados.*

Allan Rocha de Souza

IBDAAutoral



O QUE É MTD (Mineração de Textos e Dados)

“Em termos simples, mineração de dados se refere à extração ou "mineração" de grandes quantidades de dados (...). Muitas pessoas tratam a mineração de dados como sinônimo de outro termo popularmente usado, *Knowledge Discovery from Data*, ou KDD.

Alternativamente, outros veem a mineração de dados como simplesmente um passo essencial no processo de *knowledge discovery*.” (HAN, PEI e KAMBER, 2006)

“De fato, os termos ciência de dados, aprendizagem de máquinas e mineração de dados são frequentemente utilizados de forma intercambiável. O ponto em comum entre estas disciplinas é um foco na melhoria de tomada de decisões através da análise de dados” (KELLEHER e TIERNEY, 2017)

“Mineração de dados, também conhecida como *knowledge discovery in data* (KDD), é o processo de descobrir padrões e outras informações valiosas a partir de grandes conjuntos de dados” (IBM, 2022)

MTD É ESSENCIAL PARA:

- a **pesquisa** contemporânea (com dados)
- o desenvolvimento de **tecnologias digitais**
- o ecossistema de **inovação**
- **segurança jurídica** para novos negócios intensivos em dados)
- **posicionamento** do Brasil como importante parceiro comercial e de pesquisa

O QUE MTD NÃO É:

- **fruir** das obras protegidas
- **disponibilizar** ao público as obras protegidas
- **aproveitar-se** das obras protegidas
- **prejudicar** os autores
- **concorrer** com a exploração das obras protegidas
- **restringir** a utilização das obras protegidas

IBDA**Autoral**



O QUE É PROTEGIDO?

Art. 7º São obras intelectuais protegidas as **criações do espírito, expressas** por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível, conhecido ou que se invente no futuro, tais como:
XIII - as coletâneas ou compilações, antologias, enciclopédias, dicionários, bases de dados e outras obras, que, por sua **seleção, organização ou disposição** de seu conteúdo, constituam uma **criação intelectual**.

§ 2º A proteção concedida no inciso XIII **não abarca os dados** ou materiais em si mesmos e se entende sem prejuízo de quaisquer direitos autorais que subsistam a respeito dos dados ou materiais contidos nas obras.

§ 3º No domínio das ciências, a proteção **recairá sobre a forma literária ou artística**, não abrangendo o seu conteúdo científico ou técnico, sem prejuízo dos direitos que protegem os demais campos da propriedade imaterial.

O QUE NÃO É PROTEGIDO?

Art. 8º **Não são** objeto de proteção como direitos autorais de que trata esta Lei:

I - as **ideias**, procedimentos normativos, sistemas, **métodos**, projetos ou conceitos matemáticos como tais;

II - os **esquemas, planos** ou regras para realizar atos mentais, jogos ou negócios;

(...)

V - as **informações de uso comum** tais como calendários, agendas, cadastros ou legendas;

SOBRE OS BANCOS DE DADOS...

Art. 87. O **titular** do direito patrimonial sobre uma **base de dados** terá o **direito exclusivo, a respeito da forma de expressão** da estrutura da referida base, **de autorizar ou proibir:**

I - sua reprodução total ou parcial, por qualquer meio ou processo;

II - sua tradução, adaptação, reordenação ou qualquer outra modificação;

III - a distribuição do original ou cópias da base de dados ou a sua comunicação ao público;

IV - a reprodução, distribuição ou comunicação ao público **dos resultados** das operações mencionadas no inciso II deste artigo.

DIRECTIVE (EU) 2019/790 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC

Article 3

Text and data mining for the purposes of scientific research

1. Member States shall provide for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of this Directive for reproductions and extractions made by research organizations and cultural heritage institutions in order to carry out, **for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access.** [...]

Article 4

Exception or limitation for text and data mining

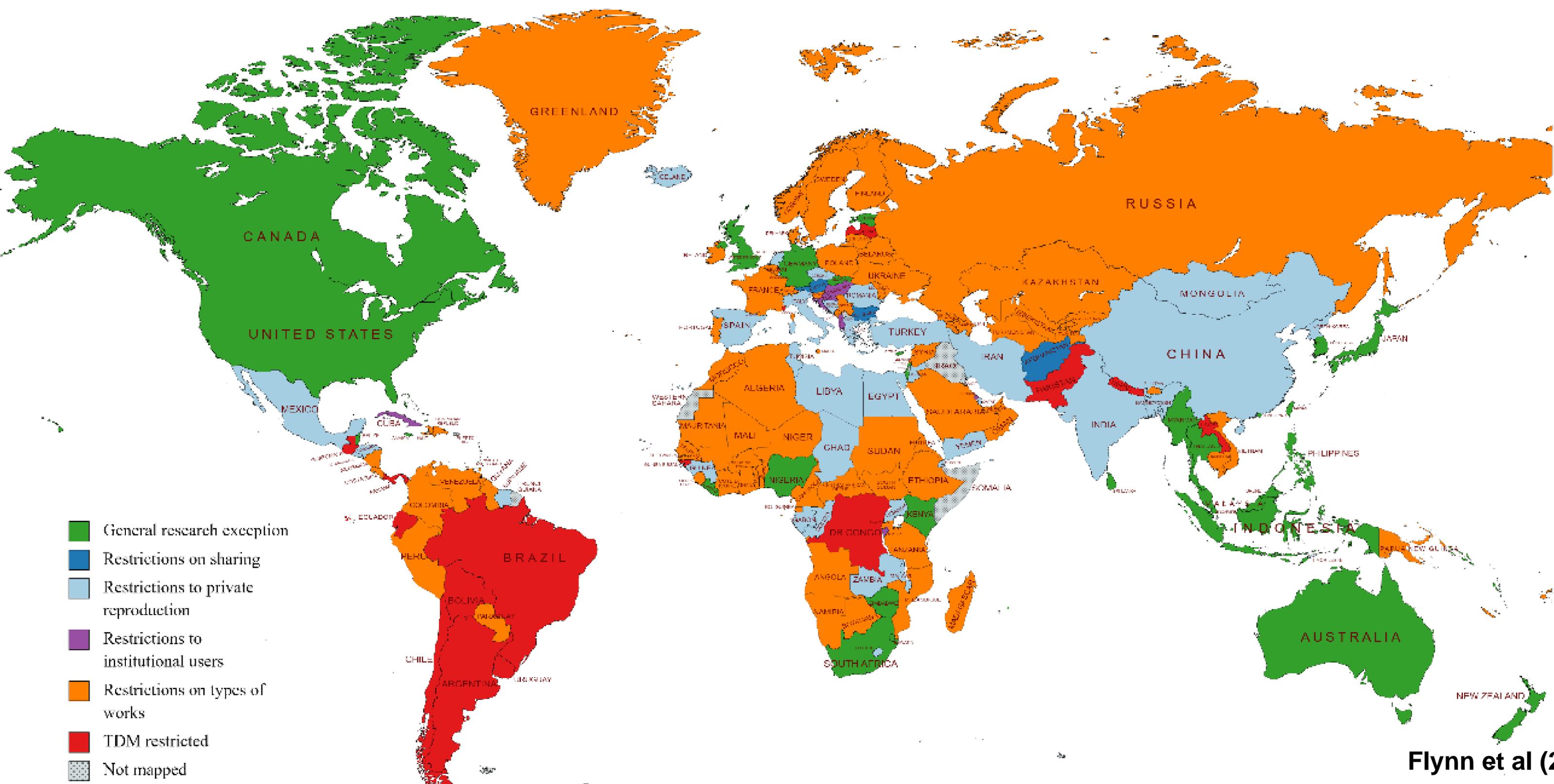
1. Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for **reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.** [...]

Copyright Act 2021 (Revised Edition 2020, Act No. 22 of 2021) (Singapura)

244.—(1) If the conditions in subsection (2) are met, it is a permitted use for a person (X) to make a copy of any of the following material: (a) a work; (b) a recording of a protected performance. (2) The conditions are : (a) the copy is made for the purpose of : (i) **computational data analysis**; or (ii) **preparing the work or recording for computational data analysis**;
(...)

Copyright Act, 1970 (Act No. 48 of May 6, 1970, as amended up to Act No. 72 of July 13, 2018) (Japão)

(“Article 30-4 (ii). Exploitations not for enjoying the ideas or emotions expressed in a work. **It is permissible to exploit work, in any way and to the extent considered necessary, in any of the following cases or other cases where such exploitation is not for enjoying (...): (ii) exploitation for using the work in a data analysis (meaning the extraction, comparison, classification, or other statistical analysis of language, sound, or image data, or other elements of which a large number of works or a large volume of data is composed;”)**)



- General research exception
- Restrictions on sharing
- Restrictions to private reproduction
- Restrictions to institutional users
- Restrictions on types of works
- TDM restricted
- Not mapped

DIRECTIVE (EU) 2019/790 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC

(3) A rápida evolução tecnológica continua a mudar a forma como as obras e outro material protegido são criados, produzidos, distribuídos e explorados. Continuam a surgir novos modelos empresariais e novos intervenientes. **Cumpra que a legislação aplicável esteja orientada para o futuro, para não limitar a evolução tecnológica.** (...)

(8) As novas tecnologias permitem a análise automática computacional de informações em formato digital, tais como texto, som, imagem ou dados, normalmente designada por prospeção de textos e dados. A prospeção de textos e dados torna possível o tratamento de grandes quantidades de informação para obter novos conhecimentos e descobrir novas tendências. **Embora as tecnologias de prospeção de textos e dados sejam predominantes em toda a economia digital, existe um amplo reconhecimento de que esta prospeção pode beneficiar nomeadamente, a comunidade científica e, ao fazê-lo, apoiar a inovação.** (...)

[...]

(11) **A insegurança jurídica no âmbito da prospeção de textos e dados deverá ser eliminada através da previsão de uma exceção** obrigatória para as universidades e outros organismos de investigação, bem como para as instituições responsáveis pelo património cultural, ao direito exclusivo de reprodução e ao direito de impedir a extração a partir de bases de dados. (...)

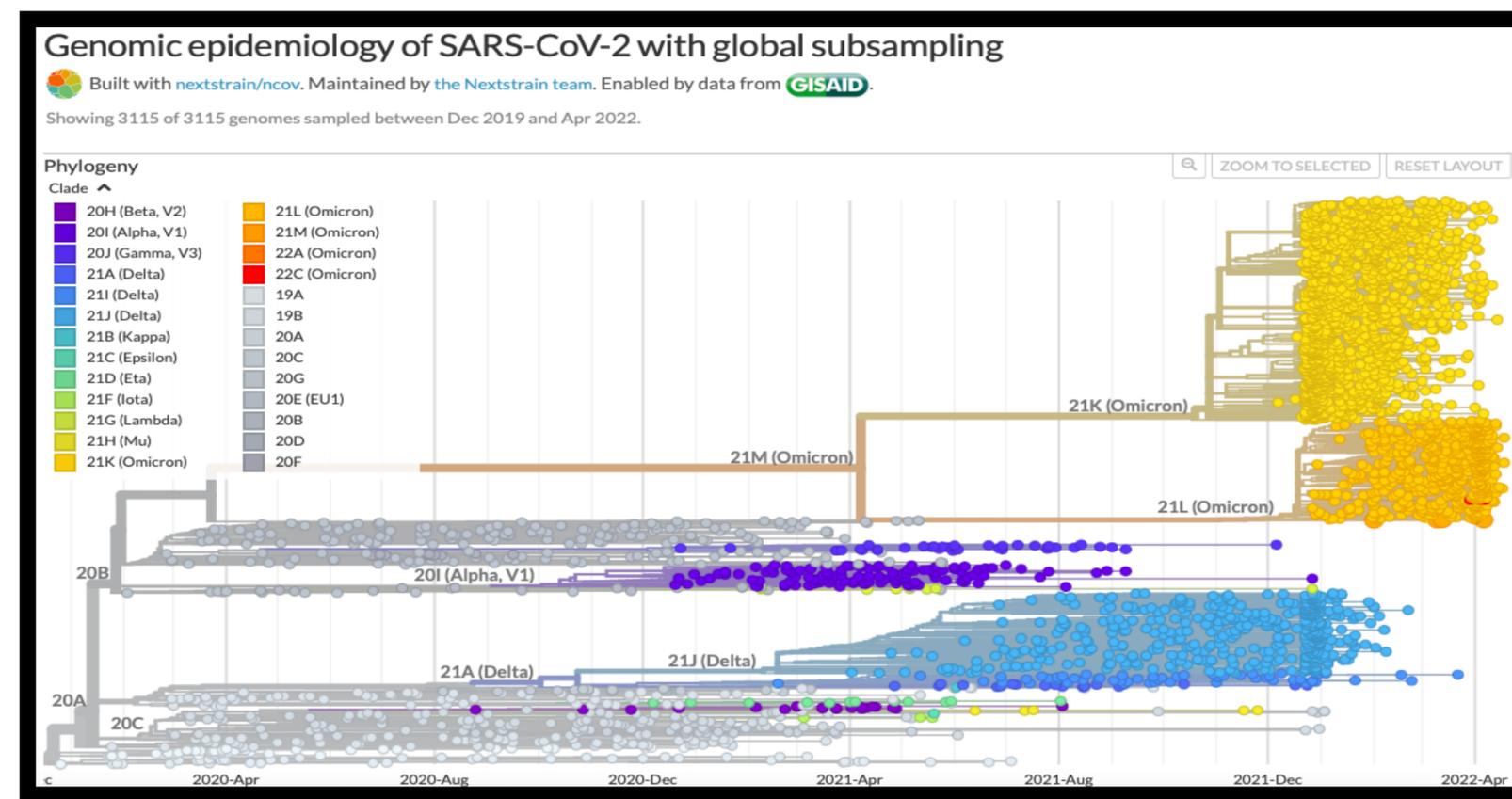
(17) Tendo em conta a natureza e o âmbito da exceção, que se limita às entidades que realizam investigação científica, **qualquer potencial prejuízo para os titulares de direitos criado através desta exceção deveria ser mínimo.** Por conseguinte, os Estados-Membros não deverão prever uma compensação para aos titulares de direitos no que se refere a utilizações ao abrigo das exceções de prospeção de textos e dados introduzidas pela presente diretiva.

PARA QUE SERVE MTD?

- Treinamento de inteligências artificiais e algoritmos;
 - Pesquisa científica;
- Cruzamento de dados oriundos de múltiplas fontes;
 - Análise linguística e de material artístico;
 - Pesquisas bibliográficas,
 - Análise de mídias sociais;
 - Pesquisas em humanidades digitais;
 - Checagem de fatos e notícias;
 - Auxílio na criação de imagens e áudio;
 - Etc.

The screenshot shows the medRxiv website interface. At the top, there are logos for medRxiv, CSH Cold Spring Harbor Laboratory, and BMJ Yale. Below the logos is a navigation menu with links for HOME, ABOUT, SUBMIT, NEWS & NOTES, and ALERTS / RSS. A search bar is visible with the text "Search" and a magnifying glass icon. Below the search bar, it says "Advanced Search". At the bottom of the screenshot, it displays "16,529 Results for term 'coronavirus' and posted between '01 Apr, 2020 and 28 Apr, 2022'".

The screenshot shows the medRxiv website interface. At the top, there are logos for medRxiv, CSH Cold Spring Harbor Laboratory, and BMJ Yale. Below the logos is a navigation menu with links for HOME, ABOUT, SUBMIT, NEWS & NOTES, and ALERTS / RSS. A yellow banner at the top states: "medRxiv is receiving many new papers on coronavirus SARS-CoV-2. A reminder: these are preliminary reports that have not been peer-reviewed. They should not be regarded as conclusive, guide clinical practice/health-related behavior, or be reported in news media as established information." Below the banner, the main heading is "COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv". Underneath, it says "23,219 Articles (17,709 medRxiv, 5,510 bioRxiv)". There is a "Subject Areas" button on the right. At the bottom, it says "Most recent first" and "Page 1: Articles 1-10 | Next".



IBDAutoral



HEALTH AND WELLNESS

How this Canadian start-up spotted coronavirus before everyone else knew about it

Published Tue, Mar 3 2020 10:27 AM EST • Updated Fri, Mar 6 2020 1:40 PM EST

 Cory Stieg
@CORYSTIEG

SHARE    

RADAR >>>

coronavírus

Nos últimos 7 dias, encontramos

2.652 publicações

de baixa qualidade dentre 45.876 analisadas.

A variação foi de **+54%** para a semana anterior

Automated analysis of the US presidential elections using Big Data and network analysis

Saatviga Sudhahar, Giuseppe A Veltri, Nello Cristianini

First Published March 2, 2015 | Research Article

 Check for updates

<https://doi.org/10.1177/2053951715572916>

[Article information](#) ▾



AI Trained on Old Scientific Papers Makes Discoveries Humans Missed

Scientists used machine learning to reveal new scientific knowledge hidden in old research papers.

 By Madeleine Gregory

Robot rhythms: the startups using AI to shake up the music business

When you think of great songwriting teams, music and artificial intelligence aren't an obvious mix, but a number of startups are bringing smart tech into the studio

RECURSO ESPECIAL Nº 1.921.769 - PR (2020/0222330-3)

RELATORA : **MINISTRA NANCY ANDRIGHI**

RECORRENTE : ALEXANDRE RICARDO PESSERL

ADVOGADOS : RICARDO WYPYCH - PR067159
ISIS KLÜBER - PR008072

RECORRIDO : UNIÃO BRASILEIRA DE COMPOSITORES

ADVOGADO : DANIEL CONDE FALCÃO RIBEIRO - PR050111

EMENTA

DIREITO AUTORAL. RECURSO ESPECIAL. AÇÃO DE OBRIGAÇÃO DE FAZER. INTERESSADO QUE PLEITEIA O ACESSO INTEGRAL ÀS INFORMAÇÕES RELATIVAS À PARTICIPAÇÃO INDIVIDUAL DE CADA ARTISTA NAS OBRAS MUSICAIS COLETIVAS. INFORMAÇÃO DE CARÁTER PÚBLICO. ACESSO QUE DEVE SER GRATUITAMENTE DISPONIBILIZADO POR MEIO ELETRÔNICO A QUALQUER INTERESSADO. ART. 98, § 6º E § 7º, DA LEI 9.610/98.

1. Ação de obrigação de fazer por meio da qual se objetiva que a União Brasileira de Compositores – UBC forneça informações relativas à participação individual de cada autor nas obras musicais coletivas (percentuais de titularidade).

2. Ação ajuizada em 25/04/2017. Recurso especial concluso ao gabinete em 11/12/2020. Julgamento: CPC/2015.

3. O propósito recursal é definir acerca da obrigatoriedade da recorrida – UNIÃO BRASILEIRA DE COMPOSITORES – em disponibilizar as informações pleiteadas pelo recorrente, a saber, a participação individual de cada artista nas obras musicais coletivas.

4. As associações de gestão coletiva de direitos autorais, a despeito de possuírem natureza jurídica de direito privado, exercem, tal qual dispõe o art. 97, § 1º, da Lei 9.610/98, atividade de interesse público, devendo atender a sua função social.

5. Nos termos do art. 98, § 6º, da Lei 9.610/98, introduzido pela Lei 12.853/13, as associações deverão manter um cadastro centralizado de todos os contratos, declarações ou documentos de qualquer natureza que comprovem a autoria e a titularidade das obras e dos fonogramas, bem como as participações individuais em cada obra e em cada fonograma, prevenindo o falseamento de dados e fraudes e promovendo a desambiguação de títulos similares de obras. Ainda, nos moldes do que dispõe o § 7º do mencionado dispositivo legal, tais informações são de interesse público e o acesso a elas deverá ser disponibilizado por meio eletrônico a qualquer interessado, de forma gratuita.

6. Do confronto entre o disposto no art. 98, § 6º e § 7º, da Lei 9.610/98 e o art. 6º, I, II, III e IV, e § 1º, da Instrução Normativa nº 3/2015 do Ministério da Cultura (IN nº 3/2015 do MinC), pode-se constatar que a suposta incompatibilidade entre as suas redações é meramente aparente, não

resistindo ante à necessidade de observância do próprio interesse público e função social das associações. Ainda que a instrução normativa não albergue expressamente a pretensão do recorrente, é bem verdade que ela também não a veda, convivendo harmonicamente com o disposto na Lei de Direitos Autorais.

7. Recurso especial conhecido e provido.

IBDAAutoral



The Direction of Innovation

1.6 The future direction of innovation

Innovation can certainly assist in solving or at least mitigating the world's grand challenges, be it climate change, inequality, the need for greater food production or better access to water, health and education. (...) p. 32

(...) Some hope comes from growing evidence that digital technologies will prove a new general-purpose technology (...) p.33

Fourth and finally, scientists and researchers that had never worked with one another started collaborating.³⁰ Some even crossed scientific fields to help with the endeavor. For example, epidemiologists joined with sociologists and economists to understand how the virus spread and how to contain it. In addition, researchers shared their work openly, even before it had been peer-reviewed, so as to speed up the diffusion of knowledge between scientists and researchers. This helped in the fast dissemination of the latest research findings.

p.64

CONDICIONANTES COMPLEMENTARES

Digitalização

However, the influence of these technologies across economic sectors and countries is uneven. Digitization of information is an essential component of how these technologies work. By accessing massive amounts of information, the technology is able to deduce patterns from the information provided, and with training learn to identify specific patterns and trends. But sufficient computing power is necessary to allow for the processing of large amounts of digitized information. This requirement could pose further difficulties for lesser developed economies in competing in the new economic age.

p. 74

IBDAutoral

Abertura e Acessibilidade

Direitos autorais e mineração de dados e textos no combate à Covid-19 no Brasil

<p>Allan Rocha de Souza Programa de Pós-Graduação em Políticas Públicas, Estratégias e Desenvolvimento, Instituto de Economia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil. Curso de Direito, Instituto Três Rios, Universidade Federal Rural do Rio de Janeiro, Três Rios, RJ, Brasil https://orcid.org/0000-0002-6549-0085</p> <p>Luca Schirru Escola de Direito e Ciências Sociais, Universidade Positivo, Curitiba, PR, Brasil https://orcid.org/0000-0002-4706-3776</p> <p>Miguel Bastos Alvarenga Núcleo de Pesquisa em Direitos Fundamentais, Relações Privadas e Políticas Públicas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil https://orcid.org/0000-0002-4079-7785</p> <p>DOI: https://doi.org/10.18617/liinc.v16i2.5536</p> <p>Palavras-chave: Direito autoral, Banco de Dados, Mineração de dados e textos, Limitações e Exceções, COVID-19</p>	<p>PDF</p> <p>Publicado 31/12/2020</p> <p>Como citar Souza, A. R. de, Schirru, L., & Alvarenga, M. B. (2020). Direitos autorais e mineração de dados e textos no combate à Covid-19 no Brasil. <i>Liinc Em Revista</i>, 16(2), e5536. https://doi.org/10.18617/liinc.v16i2.5536</p> <p>Fomatos de citação</p> <p>Edição v. 16 n. 2 (2020): Perspectivas e desafios informacionais em tempos da pandemia da Covid-19</p>	<p>English</p> <p>Español (España)</p> <p>Português (Brasil)</p> <p>A <i>Liinc em Revista</i>, ISSN 1808-3536, periódico científico de acesso aberto e avaliado por pares, é orientada para a reflexão crítica sobre dinâmicas de produção, circulação e apropriação da informação e do conhecimento, ante as transformações do mundo contemporâneo.</p> <p>É uma publicação do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), editada por seu Programa de Pós-Graduação em Ciência da Informação, desenvolvido em associação com a Universidade Federal do Rio de Janeiro (UFRJ).</p>
--	---	---



I N S T I T U T O
B R A S I L E I R O
D E D I R E I T O S
A U T O R A I S

OBRIGADO!

allan@rochadesouza.com

www.ibdautoral.org.br

IBDAAutoral

