SemeAd 2023
XXVI Seminários em Administração

07, 08, 09 e 10 de novembro de 2023
ISSN 2177-3866

SEGURANÇA, PROTEÇÃO E PRIVACIDADE DE DADOS: UM ESTUDO BIBLIOMÉTRICO E LEXICOGRÁFICO

SABRINA MEDIANEIRA DA SILVA AVILA
UNIVERSIDADE FEDERAL DE SANTA MARIA (UFSM)

MARCELO BATTESINI
UNIVERSIDADE FEDERAL DE SANTA MARIA (UFSM)

SEGURANÇA, PROTEÇÃO E PRIVACIDADE DE DADOS: UM ESTUDO BIBLIOMÉTRICO E LEXICOGRÁFICO

1. INTRODUÇÃO

Existe um aumento da preocupação com os dados, pois o mundo passa por uma transformação digital, e os dados passaram a ter uma importância mais relevante no cotidiano (de Lima, 2021). Indicando que no ambiente computacional a segurança de dados e a proteção de privacidade são questões das mais importantes (Li et al., 2023). Na era da informação são perseguidos velocidade e segurança, que sempre será uma batalha de ataque e defesa (Zhou, 2022). Com isso, o imediatismo no acesso a dados não pode permitir riscos de segurança mesmo que a sua coleta e uso tenha gradualmente se tornado diversificada (Li et al., 2023).

No Brasil, a Lei Geral de Proteção de Dados (LGPD), implementada em 2021, atinge todas as instituições públicas e privadas. Ela tem como princípio proteger os direitos de liberdade e privacidade dos cidadãos brasileiros (Donda, 2020). Pode-se entender como um aspecto essencial da regulamentação brasileira a mudança da busca incessante por dados para a captura apenas de informações que servem a um propósito e a garantia de um controle entre aqueles que acessam e processam dados (Masseno, 2020).

Apesar disso, a falta de segurança de dados tem sido evidenciada por um conjunto de incidentes nos últimos anos, tornando o cenário tão crítico e preocupante que os governos passaram a criar leis a fim de definir diretos de privacidades e penalidades para casos do não cumprimento da legislação (Bisso et al., 2020). Parte do problema é que o dado é considerado uma informação flexível e dependente do contexto, tornando desafiadora a sua aplicação e segurança nos mais diversos contextos (Sollins, 2019).

A bibliografia sobre o tema é vasta, tendo sido propostas vários estudos de revisão na literatura. Kumar et al. (2021) analisou a segurança de dados na telemedicina; Wang et. al (2021), realizaram uma revisão sistemática no contexto da saúde digital; Yalcin & Daim (2021) na estrutura do blockchain e suas patentes; Awan & Abbas (2023), o impacto e os indicadores em pesquisas sobre computação em nuvem; e Carvalho et. al (2022), identificou trabalhos relacionados à Lei Geral de Proteção de Dados no Brasil.

A relevância teórica e prática da temática demanda uma busca sistemática da literatura sobre a temática segurança, proteção e privacidade de dados, que atualize e amplie o escopo de trabalhos anteriores, dado não ter sido identificada revisão recente que utilize, ambas, análises bibliométrica e lexicográfica. Lacunas que endereçaremos por meio das questões de pesquisa: Qual(is) o panorama mundial da produção acadêmica e os principais periódicos, autores e suas redes colaborativas? Quais as relações semânticas entre as formas lexicais latentes? Como elas se agrupam? Como classificar os artigos estudados a partir delas? Com isso, este artigo tem o objetivo de caracterizar a literatura contemporânea sobre a segurança, proteção e privacidade de dados.

Nossos resultados apontam os principais periódicos e os autores mais produtivos, assim indicam quatro classes temáticas latentes podem ser utilizadas para organizar os artigos: legislação para segurança de dados; segurança de dados sensíveis; tecnologias de hardware; e tecnologias de software.

2. METODOLOGIA

Trata-se de uma pesquisa com propósito descritivo, no qual foram realizadas análises bibliométricas por meio do software RStudio e do IRaMuTeQ para análise lexicográfica. Delineamento que busca ampliar as possibilidades dos estudos unicamente bibliométricos, ao

estudar o tema proposto. Os estudos bibliométricos tratam dos aspectos relacionados a comunicação de trabalhos científicos, examinando impacto, autoria, publicações, citações e conteúdo (Araujo, 2022). Enquanto a análise lexicográfica favorece o estudo das palavras utilizadas em determinadas comunidades (Costa, 2019), ampliando e complementando a análise.

Os dados foram extraídos das bases de dados Web of Science (WoS) e Scopus em 2023 (26 de junho), considerado a produção de artigos no período de 2019 a junho de 2023 e os filtros: TITLE-ABS-KEY ("Data security" OR "Data protection" OR "Data privacity") AND PUBYEAR > 2019 AND PUBYEAR < 2023 AND (LIMIT-TO(DOCTYPE, "ar"). Configurações de busca, filtragem e extração de dados, que permitiu reunir 8.664 artigos para composição do banco de dados analisado.

As análises tratadas no RStudio foram: redes de cocitações, autores mais relevantes e periódicos mais citados. A partir do software IRaMuTeQ foram realizadas as análises lexicográficas: Classificação Hierárquica Descendente (CHD), Análise Fatorial por Correspondência (CHD) e Árvore de Similitude (ASim) da composição do banco de dados. Resultados que foram então discutidos.

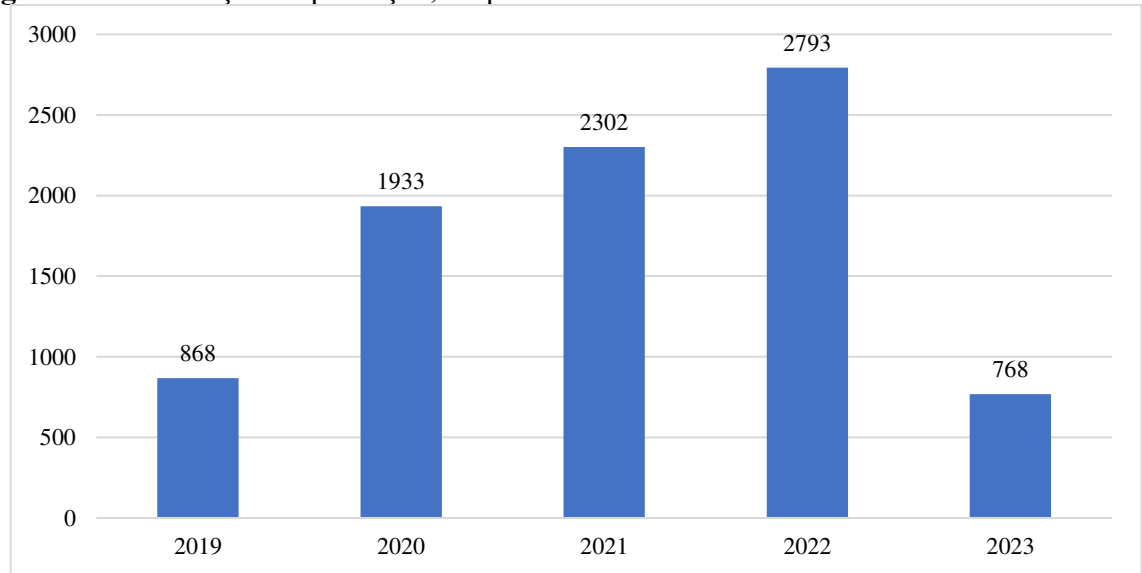
3. RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados dos estudos bibliométrico (seção 3.1) e lexicográfico (seção 3.2).

3.1. Estudo Bibliométrico

Foi observada uma grande e crescente produção de artigos no período investigado, ver Figura 1. Apresentando, em 5 anos, um total de 8.664 artigos e uma média anual de 1.732 artigos relacionado a segurança, proteção e privacidade de dados. Ainda, foi possível identificar que a China teve uma grande concentração de publicações, com um total de 1.529 artigos publicados no período, seguido da Índia com 868 artigos e da Alemanha com 518 artigos.

Figura 1 - Distribuição da produção, no período 2019-2023



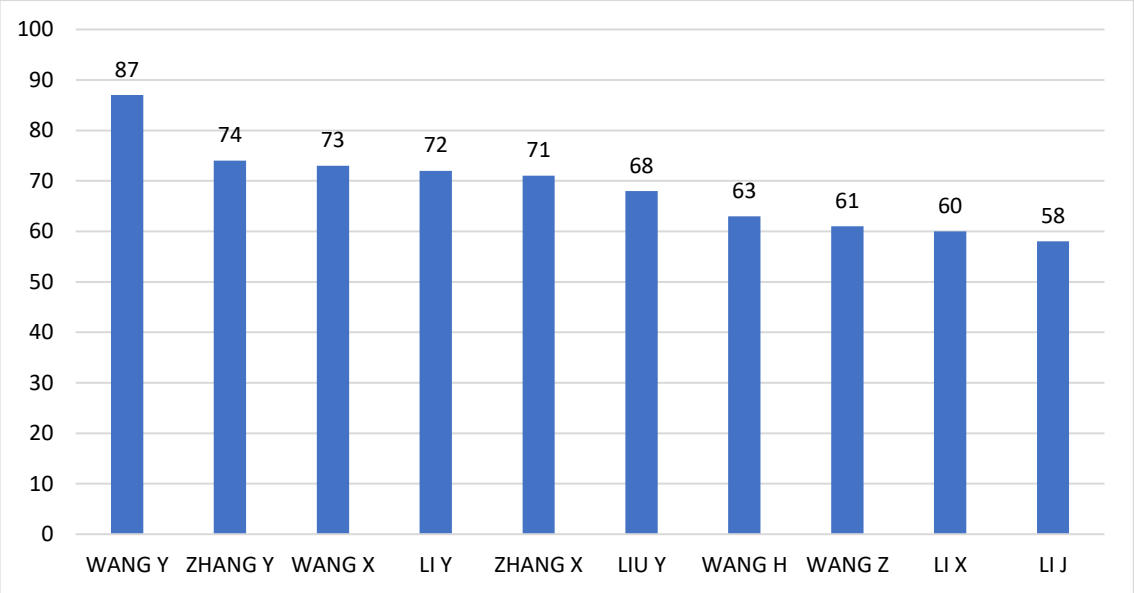
Fonte: Autores, a partir da base dados analisada no bibliometrix.

Nos primeiros artigos publicados sobre o assunto em 2019, tem-se um escrito por Reillo et. al., com o título “How to implement EU data protection regulation for R&D in biometrics”,

no periódico ScienceDirect, visando descrever e fornecer procedimentos para tratar e adquirir dados pessoais na União Europeia. Em 2020, a produção de artigos relacionados ao assunto duplicou, mantendo-se estável em 2021 e com um acréscimo grande em 2022. O crescimento nos últimos cinco anos se explica pela aplicação da GPDR (General Data Protection Regulation) pela União Europeia, em 2018, (GPDR, 2016) e da LGPD no Brasil, em 2020 (Brasil, 2018).

Em seguida foi realizada a busca dos autores com maior número de publicações, sendo os dez mais produtivos conforme indicados na Figura 2. Nela, se observa que o autor Wang Y. apresentou a maior quantidade, tendo publicado 87 artigos no período. Entre eles o artigo mais citado do autor, intitulado "Cloud-Assisted EHR Sharing With Security and Privacy Preservation via Consortium Blockchain", com o objetivo de fornecer um protocolo de compartilhamento de dados de registros médicos. O segundo mais produtivo foi Zhang Y. tendo publicado 74 artigos sendo aquele com maior destaque "3D Fluorescent Hydrogel Origami for Multistage Data Security Protection", que apresenta uma plataforma para criptografia de dados baseado em plano 3D. Wang X foi o terceiro autor em quantidade de publicações e seu artigo "Survey on blockchain for Internet of Things" (Wang, 2019), que destaca a necessidade da segurança de dados na adoção da Internet das Coisas (IoT), apresentou alto número médio anual de citações (TCperYear = 214).

Figura 2 - Autores mais relevantes, no período 2019-2023

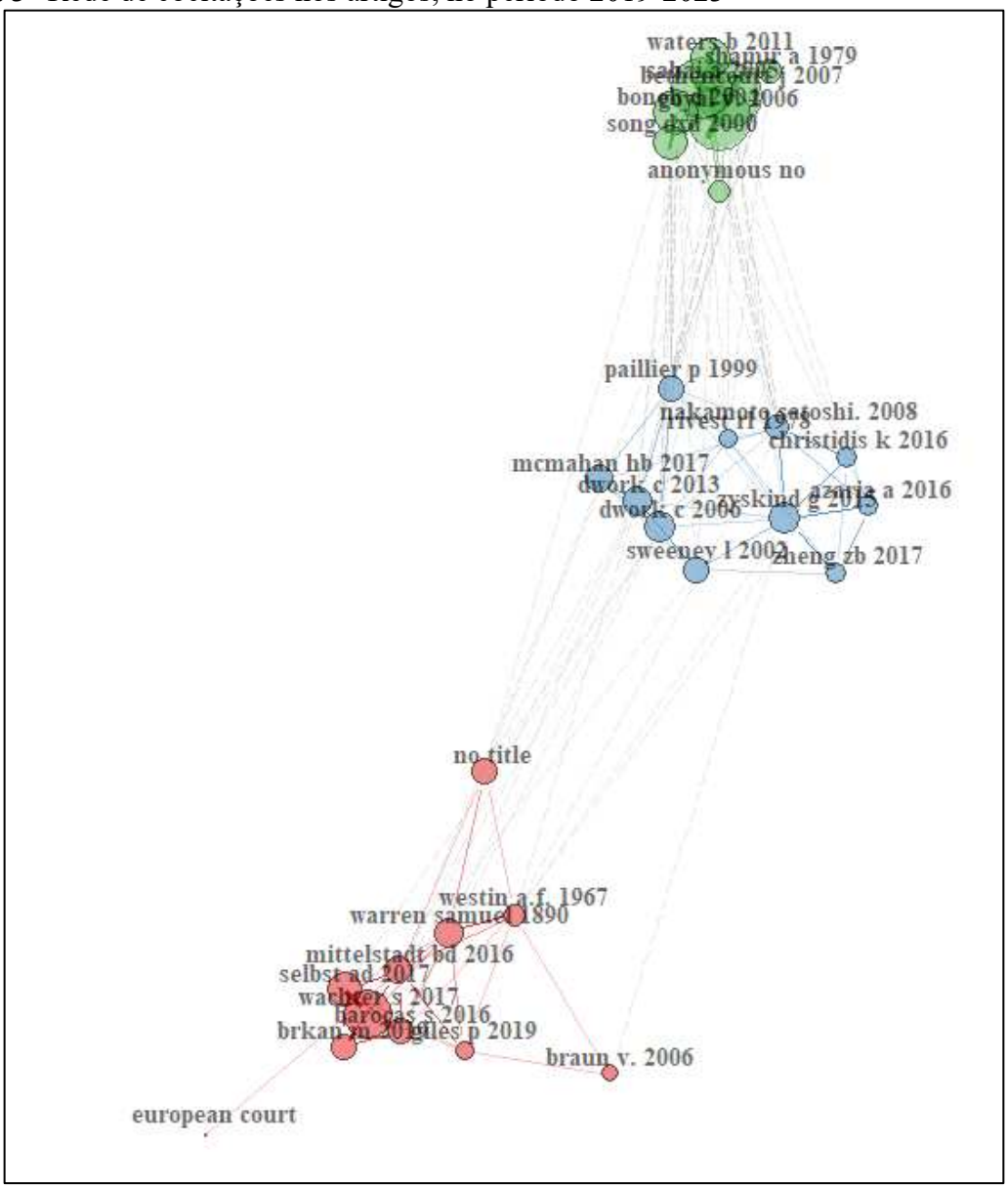


Fonte: Autores, a partir da base dados analisada no bibliometrix.

Além disso, foi gerada a rede de cocitações dos documentos recuperados, ver Figura 2, que de acordo com Castanha (2023) quantifica a frequência que dois autores foram citados concomitantemente. Evidenciando que quanto maior a frequência de cocitação, mais próxima será a relação entre os citados ou temas (Grácio, 2016).

Na Figura 3 se observa três redes principais de cocitações, demonstrando uma maior aproximação de temas, ideias e conceitos por esses autores. Dado que elas permitem entender os padrões de cocitação, identificar as literaturas potencialmente relevantes para pesquisas futuras, bem como a história, o reconhecimento e o impacto acadêmico das publicações (Hjørland, 2013).

Figura 3 - Rede de cocitações nos artigos, no período 2019-2023



Fonte: Autores, a partir da base dados analisada no bibliometrix.

Por fim, também foi estudada a quantidade de artigos sobre a segurança de dados por periódico, como apresentado na Tabela 1, sendo que 325 artigos se referem a periódicos com classificação Qualis A1 na Capes. Nela, é possível identificar os dez periódicos com maior número de artigos, sendo os cinco primeiros extratos A pela CAPES, classificação que não guarda relação direta com os escores SJR e H-Index.

Dentre os periódicos indicados na Tabela 1, destacamos os três com classificação A1 na Qualis. O artigo "Artificial intelligence-based mining of electronic health record data to accelerate the digital transformation of the national cardiovascular ecosystem: design protocol of the CardioMining study" publicado no BMJ Open, visa desenvolver um estudo estruturado em IA para transformar dados não estruturados em conjunto de dados interpretável para pacientes cardíacos (Samaras et al., 2023).

Tabela1 – Periódicos com maior número de artigos publicados e seus escores, no período 2019-2023

Periódico	Número de artigos	SJR	H-INDEX	Qualis
IEEE Access	287	0.926	204	A3
BMJ Open	155	1.059	139	A1
IEEE Internet Of Things Journal	88	3.747	149	A1
Multimedia Tools and Applications	83	0.720	93	A2
Computer Law and Security Review	82	0.718	49	A1
Sensors	77	0.764	219	-
Security and a Communication Networks	74	1.968	58	A3
International Journal of Advanced Computer Science	68	0.258	35	C
European Data Protection Law Review	59	0.185	5	-
Journal of Data Protection and Privacy	54	0.148	6	-

Fonte: Autores, a partir da base dados analisada no bibliometrix.
Legenda: SJR e H-Index, Scimago Journal & Country Rank | Qualis, Periódicos da CAPES.

No periódico IEEE Internet Of Things Journal, o artigo "Verifiable Multikeyword Search Encryption Scheme With Anonymous Key Generation for Medical Internet of Things", descreve a segurança como uma consideração indispensável em um sistema, devendo evitar o vazamento das informações ou adulterações dos registros (Liu, 2021). No mesmo periódico, o artigo "A DQN-Based Consensus Mechanism for Blockchain in IoT Networks", Liu et al. (2021), defende que a rede multicamada poderá atender as necessidades de privacidade de diferentes serviços e garantir a segurança dos dados.

Enquanto no periódico Computer Law and Security Review, o artigo "Data protection and artificial intelligence inequalities and regulations in Latin America" destaca que a inovação tecnológica e proteção de direitos humanos devem funcionar em conjunto e a regulamentação deve realizar o equilíbrio entre estimular a inovação e garantir a plena implementação das medidas de segurança (Belli e Zingales, 2021). No mesmo periódico, Karjalainen (2022) defende que a regulamentação da proteção de dados provocou uma mudança cultural de confiança nas novas tecnologias e a necessidade de regulá-las.

3.2. Análise Lexicográfica

A análise lexicográfica realizada IRaMuTeQ, utilizou um corpus textual constituído por 8.547 textos, que se desmembraram em 53.079 segmentos. Foram localizadas 1.891.662 ocorrências (palavras, formas ou vocábulos), sendo destas 14.978 formas distintas e 118 com ocorrência única.

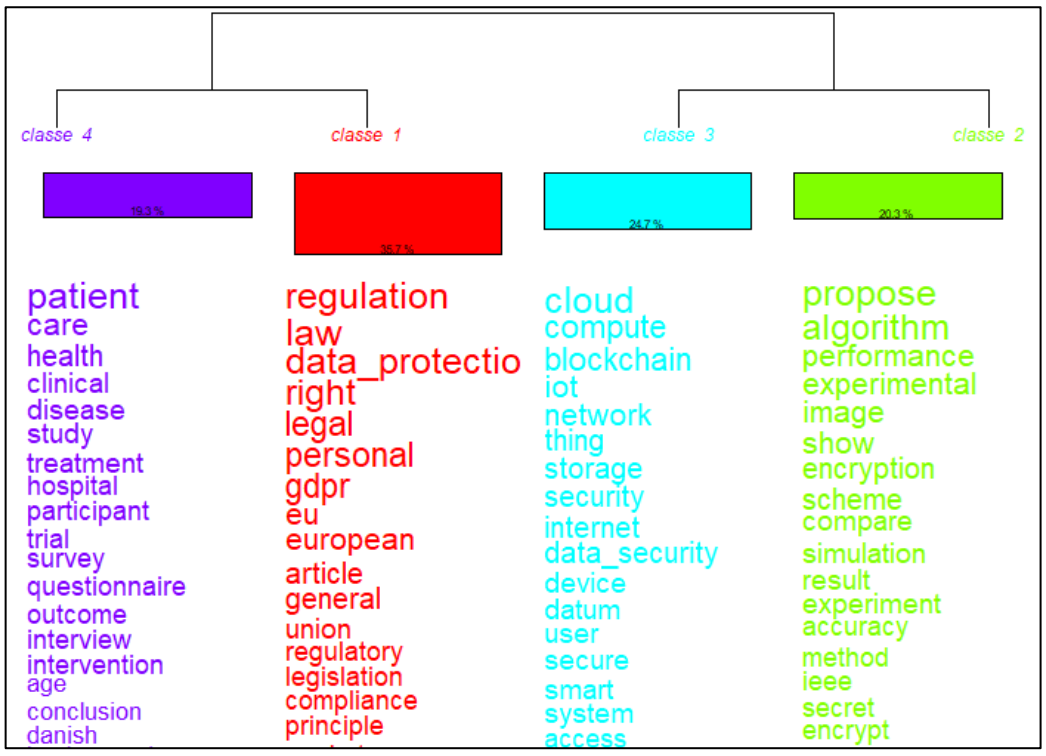
Como resultado da Classificação Hierárquica Descendente (CHD) conteúdo foi dividido em quatro classes latentes de formas lexicais indicadas na Figura 7. Nela são apresentadas as classes temáticas identificadas por cores distintas identificadas pela classificação dos segmentos de texto e, ainda, as palavras que sobressaíram em cada classe ordenadas e destacadas por tamanho da fonte.

A Classe 1 com um poder de representar 35,7% do corpus foi denominada "legislação para segurança de dados". Se destacando nela os resumos que propõem o atendimento às regulamentações de segurança de dados, com destaque para as formas textuais principais regulação, lei, direito etc. O que é consistente com o entendimento de Sun & Lu (2022), dado que a falta de segurança de dados resulta em graves perdas financeiras e danos à reputação, o que requer modelos e uso de seguros para mitigar riscos ou acidentes de violação de dados.

Ainda, para Souza e Bulzico (2022), a entrada da LGPD chegou para impor um novo enquadramento para os direitos constitucionais a liberdade e a intimidade.

Derivada dela, a Classe 4 (19,3%) foi denominada "segurança de dados sensíveis" e trata dos tipos de dados que requerem uma maior atenção no tratamento e segurança de dados, com destaque para as formas paciente, cuidado e saúde, dentre outras. Faz parte dessa classe o artigo "Data makes the story come to life: understanding the ethical and legal implications of Big Data research involving ethnic minority healthcare workers in the United Kingdom a qualitative study". Estando alinhada com a necessidade de que dados de saúde que sejam tratados, ou compartilhados, com consentimento do titular de forma específica e destacada (Brasil, 2018). Para Dove et al. (2022), o aspecto significativo para segurança de dados em pesquisa de saúde é a desconfiança nas comunidades participantes.

Figura 4 - Dendograma das classes de formas lexicais principais, artigos no período 2019-2023



Fonte: Autores, a partir da base dados analisada no IRaMuTeQ.

A Classe 3 (24,7%), foi denominada "tecnologias de hardware", se destacando as formas nuvem (armazenagem), blockchain, inteligência artificial etc. Nesta classe, no artigo "Blockchain data privacy protection modeling based on CP-ABE algorithm", os autores defendem que para aumentar a força da proteção de dados, pode-se utilizar dois métodos de criptografia para criptografar e descriptografar uma mensagem (Dang et. al., 2022). Sendo que a interconectividade digital, leva a uma maior eficiência das cidades o que exigem certos mecanismos de segurança para garantir a confidencialidade e integridade das informações pessoais e críticas (Aslam et al., 2022).

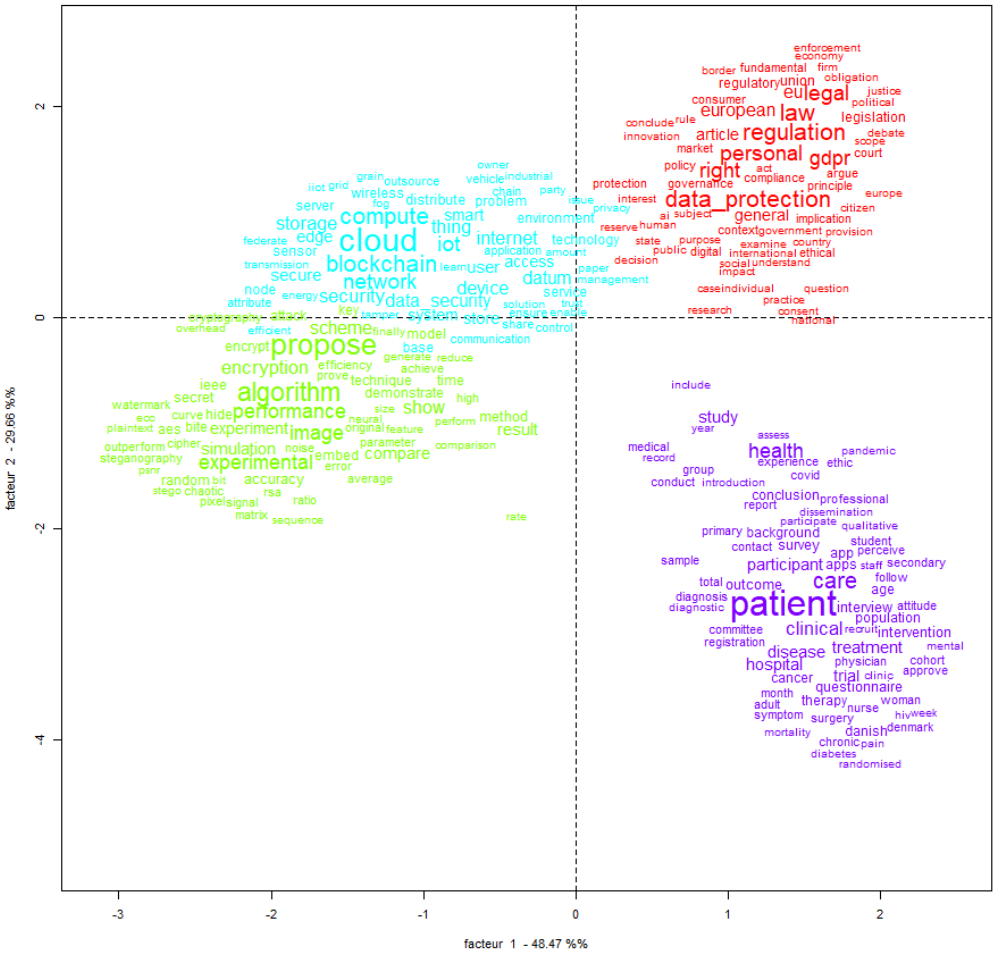
A Classe 2 (20,3%), também relacionada a tecnologias, foi denominada "tecnologias de software" se destacando pelas formas algoritmo, performance, encriptação, etc. Nessa classe, no artigo "Molecular Visual Sensing, Boolean Logic Computing, and Data Security Using a Droplet-Based Superwetting Paradigm", a garantia de segurança de dados foi aprimorada a partir da utilização de banco de dados com desenvolvimento contínuo de tecnologias (Li et al.,

2022). Enquanto Juan (2022), aponta que o uso crescente de tecnologias de internet na vida cotidiana aumenta a vulnerabilidade de informações de dados pessoais, pois se usa comunicação segura por canais inseguros.

Essas classes também foram analisadas por meio da Análise Fatorial por Correspondência (AFC), que de acordo com (Teixeira e Brandalise, 2020), nela são realizados cruzamentos entre vocabulários, a partir da incidência e frequência das palavras e as classes.

A Figura 5 apresenta o plano fatorial da AFC realizada, merecendo destaque a grande capacidade explicativa conjunta dos dois fatores nela representados (78,13%), na qual as cores identificam as classes e o tamanho da fonte indica a importância da forma textual para a classe.

Figura 5 – Gráfico de dois fatores das formas lexicais principais do corpus, artigos no período 2019-2023



Fonte: Autores, a partir da base dados analisada no IRaMuTeQ.

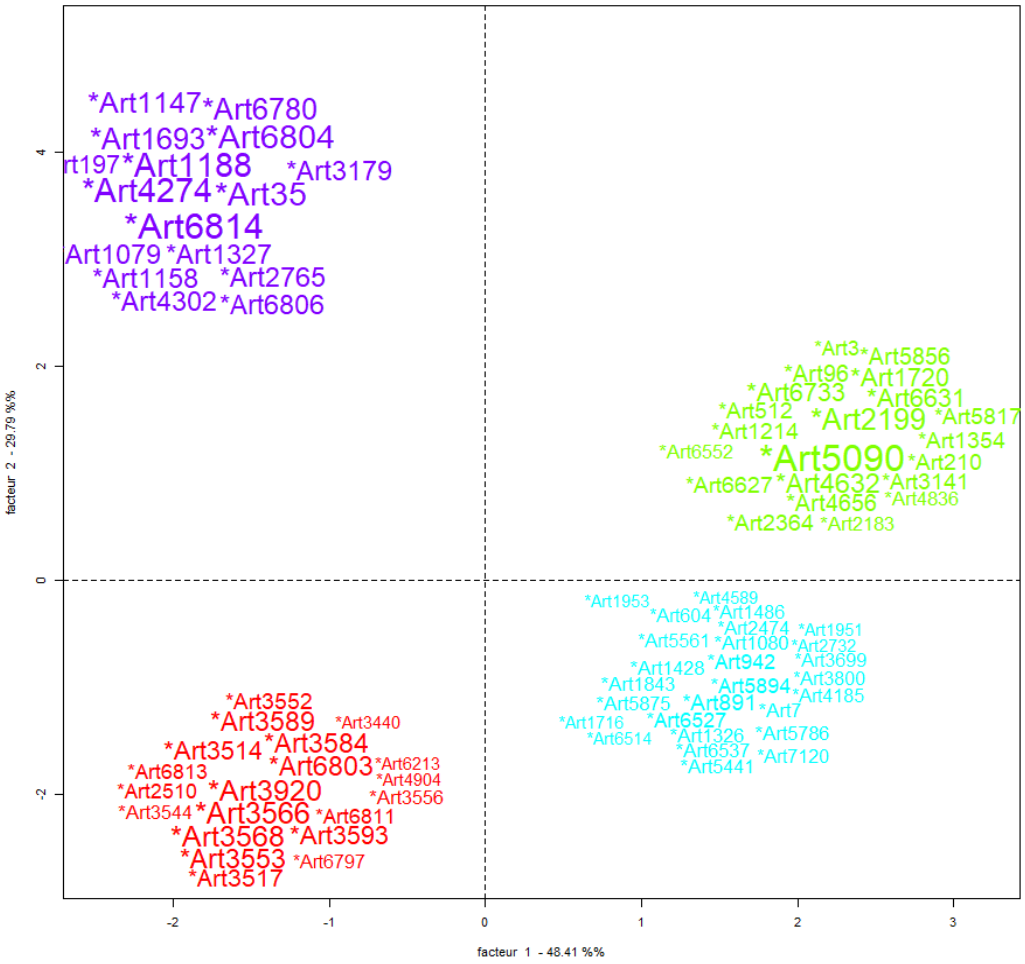
Nela, no eixo horizontal (fator 1), é possível observar que as formas das Classes 1 e 4 (positivas) apresentam estruturas discursivas compostas por segmentos de texto que se diferenciam daquelas das Classes 2 e 3 (negativas). Ou seja, apresentam segmentos de texto com grande contraste e que expressam delimitações distintas relacionadas ao grande tema segurança, proteção e privacidade de dados. Ainda em relação a Figura 5, no eixo vertical (fator 2), as Classes 1 (positivo) e 4 (negativo) apresentam grande contraste, que também pode ser observado em relação às Classes 3 (positiva) e 2 (negativa).

Assim, a Figura 5 fornece um importante critério para a organização da literatura, com delimitações latentes em torno das classes: legislação para segurança de dados; segurança de dados sensíveis; tecnologias de hardware; e de software. Em cada uma delas, os segmentos de

texto dos resumos se valem das formas pertencentes a cada classe, indicando que elas são utilizadas conjuntamente pelos autores. Informação que pode ser útil para direcionar a leitura dos artigos completos sobre a temática.

Especialmente, ao considerar que a AFC pode gerar um gráfico de dois fatores para os artigos, a partir da sua numeração na forma de variáveis, como indicado na Figura 6. Ela permite análise similar àquela da Figura 6 ao mostrar agrupamentos de artigos (resumos) e em destaque àqueles mais importantes para a formação das classes latentes.

Figura 6 – Gráfico de dois fatores dos artigos do corpus, artigos no período 2019-2023



Fonte: Autores, a partir da base dados analisada no IRaMuTeQ.

Por exemplo, na: Classe 1, o Artigo 3566 têm título “A Complete User Authentication and Key Agreement Scheme Using Cancelable Biometrics and PUF in Multi-Server Environment” que discute a proteção de dados biométricos; na Classe 2 o artigo 5090 “A Context Model for Intelligible Explanations in Adaptive Personalized Learning Environments” discute a solução para aplicações inteligíveis sobre os processos dos sistemas e o armazenamento de dados pessoais; Classe 3, o Artigo 6527 “Privacy Laws, Genomic Data and Non-Fungible Tokens” que analisa as intersecções entre as legislações de privacidade de dados e contratos entre fornecedores; e Classe 4, o Artigo 6814 “A Blockchain Framework for Patient-Centered Health Records and Exchange (HealthChain): Evaluation and Proof-of-Concept Study” trata de um novo framework blockchain centrado no paciente.

Por fim, foi realizada uma análise de similitude que é fundamentada na teoria dos grafos e, de acordo com (Silva e Nascimento, 2019), representa a ligação entre as formas ativas do

corpus textual por meio das coocorrências entre as palavras, resultando em uma árvore que representa conectividade entre elas.

A Figuras 7, apresenta a análise de similitude, obtida pelo algoritmo da árvore máxima, considerado o conjunto de resumos de artigos analisados. Na árvore é possível identificar, o destaque dado as comunidades de palavras, com destaque para as formas ativas: “security”, “privacy”, “data_security”, “data_protection”, “proposed”, “system”, “information”, “study”, “health” e “patient”. E de suas conexidades comuns. Ainda nela, a palavra “security” ocupa uma posição central, possuindo conexidades mais expressivas com as formas “data_security”, “data_protection”, “system” e “information”.

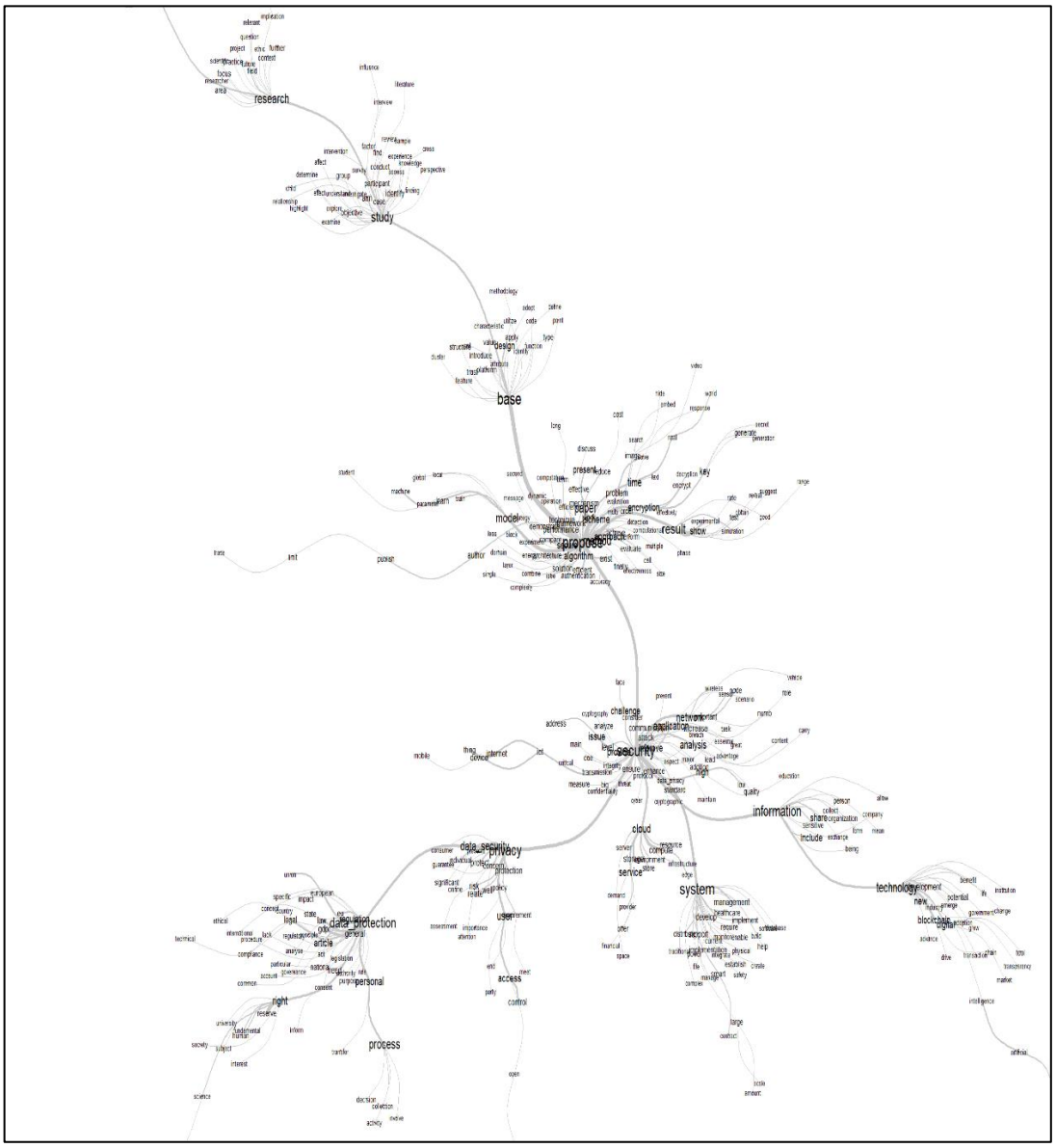
Além disso, há um conjunto de comunidades de palavras conexas que representam as delimitações em torno do tema central “segurança” de dados, sendo elas: segurança da informação e tecnologia; segurança sistemas e armazenagem de dados; segurança privacidade e proteção de dados; estudo da segurança (modelo, delineamento, pesquisa etc.). Comunidades que indicam “ramos” de estudo que podem ser destacados nos resumos dos artigos analisados, sendo alguns exemplos indicados a seguir.

No ramo “proposed”, os autores do artigo “Development of Security Rules and Mechanisms to Protect Data from Assaults”, definem que diferentes regras em diferentes fases podem ser utilizadas para criptografar os dados, assim, mesmo que ocorra o vazamento dos dados, estes estarão criptografados (Zahra et. al., 2022). No ramo “security”, o artigo “A Localized Bloom Filter-Based CP-ABE in Smart Healthcare”, defende a técnica de acesso a dados urgentes de saúde em caso de emergência uma vez sem credencial de login, baseado em datagramas e acesso a dados orientado por sessão para suprir as necessidades de segurança, urgência e disponibilidade de dados (Remamany et al., 2022).

A seguir, no ramo “privacy”, Kumar et al. (2022), defende que os serviços relacionados a internet das coisas, inteligência artificial e computação em nuvem, necessitam de arquitetura distribuída, fatores de autenticação e protocolos para evitar ações ilegais nos sistemas. Ligando a ele, no ramo “data_protection”, Arulprakash e Jebakumar (2022), a computação e o armazenamento em nuvem é seguro desde que a proteção de dados seja tratada como item essencial.

No ramo “information”, mais à direita na árvore, o artigo “Automatic Detection of Sensitive Data Using Transformer- Based Classifiers”, Petrolini et al. (2022) defendem que os gestores dos dados devem estar cientes que se possuírem informações confidenciais, deverão seguir uma disciplina rígida de proteção de dados. Por fim, ligado a ele têm-se o ramo “technology”, no qual a publicação “A novel metaheuristics with deep learning enabled intrusion detection system for secured smart environment”, apresenta como um problema desafiador a necessidade de segurança continua estabelecida com a profunda integração da internet e da computação onipresente (Malibari et al., 2022).

Figura 7 - Árvore de Similitude, artigos no período 2019-2023



Fonte: Autores, a partir da base dados analisada no IRaMuTeQ.

4. CONCLUSÃO

Este artigo teve o objetivo de caracterizar a literatura contemporânea sobre a segurança, proteção e privacidade de dados, por meio da análise bibliográfica e Lexicográfica, considerando o período entre 2019 e 2023. O estudo quantitativo realizado no ambiente RStudio e as análises lexicais no IRaMuTeQ, permitiram ampliar a compreensão sobre a temática, elucidando similaridades e diferenças presentes na literatura.

Nossos resultados indicam que a relevância do tema no período investigado, bem como mostraram o crescimento de publicações relacionadas a ele (bibliometria). E, também, demonstraram que a produção acadêmica pode ser delimitada em torno das classes lexicais latentes (lexicografia), denominadas: legislação para segurança de dados, segurança de dados sensíveis, tecnologias de hardware e tecnologias de software. Caracterização que pode oferecer importante auxílio a estudos posteriores que exploram artigos relacionadas à temática.

Como limitação desta pesquisa se destaca o fato de terem sido utilizados artigos de apenas dois repositórios (Wos e Scopus) na composição do banco de dados. Em estudos futuros poderão ser utilizados artigos de outros repositórios, a fim de verificar a similaridade com os resultados obtidos.

REFERÊNCIAS

Ali, I., Sabir, S., & Ullah, Z. (2019). Internet of things security, device authentication and access control: a review. https://arxiv.org/ftp/arxiv/papers/1901/1901.07309.pdf

Araújo, C. S. de. (2022). Urban circular economy as a resource for a circular city: a bibliometric study. Revista Produção e Desenvolvimento, 8(1), e627. https://doi.org/10.32358/rpd.2022.v8.627

Arulprakash, M., & Jebakumar, R. (2022). Towards developing a block chain based advanced Data Security-Reward Model (DSecCS) in mobile crowd sensing networks. Egyptian Informatics Journal, 23(3), 405-415. https://doi.org/10.1016/j.eij.2022.03.002

Aslam, M., Khan Abbasi, M. A., Khalid, T., Shan, R. us, Ullah, S., Ahmad, T., Saeed, S., et al. (2022). Getting Smarter about Smart Cities: Improving Data Security and Privacy through Compliance. Sensors, 22(23), 9338. MDPI AG. Retrieved from http://dx.doi.org/10.3390/s22239338

Awan, W. A., & Abbas, A. (2023). Mapping the quantity, quality and structural indicators of Asian (48 countries and 3 territories) research productivity on cloud computing. Library Hi Tech, 41(2), 309-332. 10.1108/lht-07-2021-0233

Ayaz, A., Celik, K., & Ozyurt, O. (2021). Pattern detection in cloud computing: Bibliometric mapping of publications in the field from past to present. COLLNET Journal of Scientometrics and Information Management, 15(2), 469-494. https://doi.org/10.1080/09737766.2021.2007038

Belli, L., & Zingales, N. (2022). Data protection and artificial intelligence inequalities and regulations in Latin America. Computer Law & Security Review, 47, 105761. https://doi.org/10.1016/j.clsr.2022.105761

Bisso, R., Kreutz, D., Rodrigues, G., & Paz, G. (2020). Vazamentos de dados: histórico, impacto socieconomico e as novas leis de proteção de dados. Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação, 3(1). doi:10.5281/zenodo.3833275

Brasil. Lei Nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Diário Oficial da União, 2018.

Carvalho, H. E. R. H. de., Freitag, A. E. B., & Santos, D. R. dos. (2022). Impactos da implantação da Lei Geral de proteção de dados pessoais no brasil: uma análise bibliométrica: Impacts of the implementation of the General Law for the protection of personal data in brazil: a bibliometric analysis. Revista de Gestão e Secretariado (Management and Administrative Professional Review), 13(3), 1398–1411. https://doi.org/10.7769/gesec.v13i3.1412

Castanha, R. G. (2023). The Coupler: uma nova ferramenta bibliométrica para análises relacionais de citação, acoplamento bibliográfico e cocitação. RDBCI, Revista Digital de Biblioteconomia e Ciência da Informação, n. 20. https://doi.org/10.20396/rdbci.v20i00.8671208

Costa, D. de S. S. (2019). Tratamento lexicográfico de dados geolinguísticos: discussões a partir da elaboração do Vocabulário dialetal do Centro-Oeste. A Cor Das Letras, 20(1), 127–142. https://doi.org/10.13102/cl.v20i1.4742

Dang, Q., Qiu, Y., Sun, B., Yang, Z. & Liu, X. (2022). Blockchain data privacy protection modeling based on CP-ABE algorithm. International Journal of Emerging Electric Power Systems. https://doi.org/10.1515/ijeeps-2022-0094

De Lima, A. C. (2021). Segurança de dados e Big Data. Editora Senac, São Paulo.

Donda, D. (2020). Guia prático de implementação da LGPD. Editora Labrador.

Dove, E.S., Reed-Berendt, R., Pareek, M. et al. "Data makes the story come to life:" understanding the ethical and legal implications of Big Data research involving ethnic minority healthcare workers in the United Kingdom-a qualitative study. BMC Med Ethics 23, 136 (2022). https://doi.org/10.1186/s12910-022-00875-9

Goram, M., & Veiel, D. (2020). A Context Model for Intelligible Explanations in Adaptive Personalized Learning Environments. International Journal of Information and Education Technology, 10(5). doi: 10.18178/ijiet.2020.10.5.1388

Grácio, M. C. C. (2016). A coplamento bibliográfico e análise de cocitação: revisão teórico-conceitual. Encontros Bibli, revista eletrônica de biblioteconomia e ciência da informação, 21(47), 82-99. https://www.redalyc.org/journal/147/14746959008/movil

Hjørland, B. (2013). Citation analysis: A social and dynamic approach to knowledge organization. Information Processing & Management, 49(6), 1313-1325. https://doi.org/10.1016/j.ipm.2013.07.001

Hylock, R. H., & Zeng, X. (2019). A blockchain framework for patient-centered health records and exchange (HealthChain): evaluation and proof-of-concept study. Journal of medical Internet research, 21(8), e13592. doi:10.2196/13592

Juan, W. (2022). Resource cache sharing system of education information center network based on internet of things. Mobile Information Systems, 2022. https://doi.org/10.1155/2022/4947586

Kumar, A., Sinha, S., Jameel, J., & Kumar, S. (2022). Telemedicine trends in orthopaedics and trauma during the COVID-19 pandemic: A bibliometric analysis and review. Journal of Taibah University Medical Sciences, 17(2), 203-213. https://doi.org/10.1016/j.jtumed.2021.09.003

Kumar, V., Mahmoud, M. S., Alkhayyat, A., Srinivas, J., Ahmad, M., & Kumari, A. (2022). RAPCHI: Robust authentication protocol for IoMT-based cloud-healthcare infrastructure. The Journal of Supercomputing, 78(14), 16167-16196. https://doi.org/10.1007/s11227-022-04513-4

Li, F., Wang, J., & Song, Z. (2023). Privacy Protection of Cloud Computing Based on Strong Forward Security. International Journal of Cloud Applications and Computing (IJCAC), 13(1), 1-9. http://doi.org/10.4018/IJCAC.323804

Li, F., Wang, J., & Song, Z. (2023). Privacy Protection of Cloud Computing Based on Strong Forward Security. International Journal of Cloud Applications and Computing (IJCAC), 13(1), 1-9. 10.4018/IJCAC.323804

Liu, X., Yang, X., Luo, Y., & Zhang, Q. (2021). Verifiable multikeyword search encryption scheme with anonymous key generation for medical internet of things. IEEE Internet of Things Journal, 9(22), 22315-22326. 10.1109/JIOT.2021.3056116

Liu, Z., Hou, L., Zheng, K., Zhou, Q., & Mao, S. (2021). A DQN-based consensus mechanism for blockchain in IoT networks. IEEE Internet of Things Journal, 9(14), 11962-11973.

Malibari, A. A., Alotaibi, S. S., Alshahrani, R., Dhahbi, S., Alabdan, R., Al-wesabi, F. N., & Hilal, A. M. (2022). A novel metaheuristics with deep learning enabled intrusion detection system for secured smart environment. Sustainable Energy Technologies and Assessments, 52, 102312. https://doi.org/10.1016/j.seta.2022.102312

Masseno, M. D. (2020). A segurança dos dados na LGPD, brasileira: uma perspectiva europeia, desde Portugal. Proteção de dados pessoais em perspectiva, 39.

Petrolini, M., Cagnoni, S., & Mordonini, M. (2022). Automatic Detection of Sensitive Data Using Transformer- Based Classifiers. Future Internet, 14(8), 228. MDPI AG. Retrieved from http://dx.doi.org/10.3390/fi14080228

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46. Official Journal of the European Union, 59(1-88), 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

Reillo, R. S., Ortega-Fernandez, I., Ponce-Hernandez, W., & Quiros-Sandoval, H. C. (2019). How to implement EU data protection regulation for R&D in biometrics. Computer Standards & Interfaces, 61, 89-96. https://doi.org/10.1016/j.csi.2018.01.007

Remamany, K. P., Maheswari, K., Ramesh Babu Durai, C., Anushkannan, N. K., Victoria, D. R. S., Ben Othman, M. T., Hamdi, M., et al. (2022). A Localized Bloom Filter-Based CP-ABE in Smart Healthcare. Applied Sciences, 12(24), 12720. MDPI AG. Retrieved from http://dx.doi.org/10.3390/app122412720

Samaras, A., Bekiaridou, A., Papazoglou, A. S., Moysidis, D. V., Tsoumakas, G., Bamidis, P., Tsigkas G. & Giannakoulas, G. (2023). Artificial intelligence-based mining of electronic health record data to accelerate the digital transformation of the national cardiovascular ecosystem: design protocol of the CardioMining study. BMJ open, 13(4), e068698

Silva, M. A., Hedler, H. C., & Nascimento, T. G. (2019) Gestão participativa no instituto federal de goiás: análise qualitativa do estatuto com uso do software Iramuteq. Revista de Administração Educacional, 10, 40-55. https://periodicos.ufpe.br/revistas/ADED/article/download/242650/33180

Sollins, K. (2019). IoT Big Data Security and Privacy Versus Innovation. IEEE Internet of Things Journal, 6(2), 1628-1635. 10.1109/JIOT.2019.2898113

Sousa, D. R., & Bulzico, B. A. (2022). O princípio da publicidade dos atos processuais e as novas regras de privacidade e proteção de dados pessoais no Brasil. Revista Brasileira de Políticas Públicas, 12(3). https://www.publicacoesacademicas.uniceub.br/RBPP/article/view/7825

Sun, M., & Lu, Y. (2022). A Generalized Linear Mixed Model for Data Breaches and Its Application in Cyber Insurance. Risks, 10(12), 224. MDPI AG. Retrieved from http://dx.doi.org/10.3390/risks10120224

Teixeira, O., & Brandalise, M. A. T. (2020). Conhecimento pedagógico do conteúdo: cenário das pesquisas brasileiras nos contextos da licenciatura e da docência em matemática (2001-2018). Actio, Docência em Ciências, 5(2), 1-21. https://doi.org/10.3895/actio.v5n2.11287

Uribe, D., & Waters, G. (2020). Privacy laws, genomic data and non-fungible tokens. The Journal of The British Blockchain Association. https://doi.org/10.31585/jbba-3-2-(5)2020

Wang, Q., Su, M., Zhang, M., & Li, R. (2021). Integrating digital technologies and public health to fight Covid-19 pandemic: key technologies, applications, challenges and outlook of digital healthcare. International Journal of Environmental Research and Public Health, 18(11), 6053. 10.3390/ijerph18116053

Wang, X., Gu, W., Wang, F. et al. (2022). A potential controlling approach on surface ozone pollution based upon power big data. SN Appl. Sci. 4, 164. https://doi.org/10.1007/s42452-022-05045-5

Wang, X., Zha, X., Ni, W., Liu, R. P., Guo, Y. J., Niu, X., & Zheng, K. (2019). Survey on blockchain for Internet of Things. Computer Communications, 136, 10-29. https://doi.org/10.1016/j.comcom.2019.01.006

Wang, Y., Chen, C., Chen, Z., & He, J. (2020). Attribute-based user revocable data integrity audit for internet-of-things devices in cloud storage. Security and Communication Networks, 2020, 1-10. https://doi.org/10.1155/2020/8837456

Xu, Y., Tao, Y., Zhang, C., Xie, M., Li, W., & Tai, J. (2022). Review of digital economy research in China: a framework analysis based on bibliometrics. Computational Intelligence and Neuroscience, 2022. https://doi.org/10.1155/2022/2427034

Yalcin, H., & Daim, T. (2021). Mining research and invention activity for innovation trends: case of blockchain technology. Scientometrics, 126(5), 3775-3806. https://doi.org/10.1007/s11192-021-03876-4

Zahra, S. W., Arshad, A., Nadeem, M., Riaz, S., Dutta, A. K., Alzaid, Z., Alabdan, R., et al. (2022). Development of Security Rules and Mechanisms to Protect Data from Assaults. Applied Sciences, 12(24), 12578. MDPI AG. Retrieved from http://dx.doi.org/10.3390/app122412578

Zhang, H., Bian, W., Jie, B., Xu, D., & Zhao, J. (2021). A complete user authentication and key agreement scheme using cancelable biometrics and PUF in multi-server environment. IEEE Transactions on Information Forensics and Security, 16, 5413-5428. 10.1109/TIFS.2021.3128826

RESUMO ESTRUTURADO

SEGURANÇA, PROTEÇÃO E PRIVACIDADE DE DADOS: UM ESTUDO BIBLIOMÉTRICO E LEXICOGRÁFICO

Segurança de Dados; Revisão Bibliométrica; Lexicografia

Introdução (no máximo 600 caracteres, incluindo espaços)
Existe um aumento da preocupação com os dados, pois o mundo passa por uma transformação digital, e os dados passaram a ter uma importância mais relevante no cotidiano (de Lima, 2021). Indicando que no ambiente computacional a segurança de dados e a proteção de privacidade são questões das mais importantes (Li et al., 2023). No Brasil, a Lei Geral de Proteção de Dados (LGPD), implementada em 2021, atinge todas as instituições públicas e privadas. Ela tem como princípio proteger os direitos de liberdade e privacidade dos cidadãos brasileiros (Donda, 2020).

Problema de Pesquisa e Objetivo (no máximo 600 caracteres, incluindo espaços)
A relevância teórica e prática da temática segurança, proteção e privacidade de dados demanda uma busca sistemática da literatura que atualize e amplie o escopo de trabalhos anteriores. Lacunas que endereçaremos por meio das questões de pesquisa: Qual(is) o panorama mundial da produção acadêmica e os principais periódicos, autores e suas redes colaborativas? Quais as relações semânticas entre as formas lexicais latentes? Como elas se agrupam? Como classificar os artigos estudados a partir delas? Com isso, este artigo tem o objetivo de caracterizar a literatura contemporânea sobre a segurança, proteção e privacidade de dados.

Fundamentação Teórica (no máximo 600 caracteres, incluindo espaços)
Esta é uma pesquisa com propósito descritivo, no qual foram realizadas análises bibliométricas (RStudio) e análise lexicográfica (IRaMuTeQ). Delineamento que busca ampliar as possibilidades dos estudos unicamente bibliométricos, ao estudar o tema proposto. Os estudos bibliométricos tratam dos aspectos relacionados a comunicação de trabalhos científicos, examinando impacto, autoria, publicações, citações e conteúdo (Araujo, 2022). A análise lexicográfica favorece o estudo das palavras utilizadas em determinadas comunidades (Costa, 2019), ampliando e complementando a análise.

Discussão (no máximo 600 caracteres, incluindo espaços)
Um conjunto de resultados que foram discutidos. Em relação às análises bibliométricas, foram: redes de cocitações, autores mais relevantes e periódicos mais citados. Em relação às análises lexicográficas: Classificação Hierárquica Descendente, Análise Fatorial por Correspondência e Árvore de Similitude. Nossos resultados apontam os principais periódicos e os autores mais produtivos, assim indicam quatro classes temáticas latentes podem ser utilizadas para organizar os artigos presentes na literatura.

Conclusão (no máximo 600 caracteres, incluindo espaços)
Alcançamos nosso objetivo de caracterizar a literatura contemporânea sobre a segurança, proteção e privacidade de dados, considerando o período 2019-2023, o que permitiu ampliar a compreensão da temática, elucidando similaridades e diferenças nas relações dos temas.

Contribuição / Impacto (no máximo 600 caracteres, incluindo espaços)
Nossos resultados indicam que a relevância do tema no período investigado, bem como mostraram o crescimento de publicações relacionadas a ele (bibliometria). E, também, demonstraram que a produção acadêmica pode ser delimitada em torno das classes lexicais latentes (lexicografia), denominadas: legislação para segurança de dados, segurança de dados

sensíveis, tecnologias de hardware e tecnologias de software. Caracterização que pode oferecer importante auxílio a estudos posteriores que exploram artigos relacionadas à temática.

Referências Bibliográficas (no máximo 600 caracteres, incluindo espaços)

Araújo, C. S. de. (2022). Urban circular economy as a resource for a circular city: a bibliometric study. Rev Prod e Desenv, 8(1), e627.
Costa, D. de (2019). Tratamento lexicográfico de dados geolinguísticos: discussões [...]. A Cor Das Letras, 20(1), 127–142.
De Lima, A. C. (2021). Segurança de dados e Big Data. Editora Senac, São Paulo.
Donda, D. (2020). Guia prático de implementação da LGPD. Editora Labrador.
Li, F., Wang, J., & Song, Z. (2023). Privacy Protection of Cloud Computing Based on Strong Forward Security. Internat J Cloud Appand Comp (IJCAC), 13(1), 1-9.