



XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT 8 – Informação e Tecnologia

INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS NA GESTÃO DE DOCUMENTOS: UMA REVISÃO SISTEMÁTICA DE LITERATURA

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE TECHNOLOGIES IN DOCUMENT MANAGEMENT: A LITERATURE SYSTEMATIC REVIEW

Edenilza Valéria da Silva Magalhães – Universidade Federal de São Carlos (UFSCar)

Dênis Leonardo Zaniro – Universidade Federal de São Carlos (UFSCar)

Vinícius Rafael Micali Soares – Universidade Federal de São Carlos (UFSCar)

Roniberto Morato do Amaral – Universidade Federal de São Carlos (UFSCar)

Bárbara Coelho Neves – Universidade Federal da Bahia (UFBA)

Modalidade: Trabalho Completo

Resumo: A Gestão Eletrônica de Documentos é uma área que vem recebendo cada vez mais atenção nas organizações, em razão do papel estratégico da informação perante a competitividade e as tendências do mercado atual. De um lado, trata-se de uma área que envolve a aplicação de diferentes recursos tecnológicos para a recuperação de informações levando a melhorias significativas em relação às práticas manuais de gestão de documentos. De outro, a evolução tecnológica cria condições para que volumes maiores de dados sejam tramitados exigindo a adoção de novas estratégias para lidar com aspectos como eficiência e segurança. Assim, o objetivo foi investigar como as tecnologias oriundas da inteligência artificial e da ciência de dados podem ser utilizadas na gestão eletrônica de documentos com base em uma revisão sistemática da literatura. Como método, definiu-se um protocolo de revisão e, a partir deste, conduziu-se a primeira etapa de análise dos estudos em uma perspectiva qualitativa. Como resultado inicial, foram selecionados 54 estudos, cuja maior parte envolve a aplicação de inteligência artificial em vez de ciência de dados na gestão de documentos. A análise ainda revelou que boa parte dos estudos são recentes, estão concentrados na área de saúde e são oriundos de países como Estados Unidos e África do Sul. Conclui-se que existe quantidade considerável de estudos que tratam da interseção entre a gestão de documentos e a computação, e o protocolo de revisão proposto pode contribuir para a confiabilidade e replicação da pesquisa, com base nas medidas de revocação e precisão.

Palavras-chave: gestão eletrônica de documentos; inteligência artificial; ciência de dados; revisão sistemática da literatura.

Abstract: Electronic Document Management is an area that has been receiving increasing attention in organizations due to the strategic role of information in competitiveness and current market trends. On one hand, it involves the application of different technological resources for information retrieval,

leading to significant improvements over manual document management practices. On the other hand, technological evolution creates conditions for larger volumes of data to be processed, necessitating the adoption of new strategies to address aspects such as efficiency and security. Thus, the objective was to investigate how technologies stemming from artificial intelligence and data science can be utilized in electronic document management based on a systematic literature review. As a method, a review protocol was defined, and from this, the first stage of analysis of the studies was conducted from a qualitative-quantitative perspective. As an initial result, 54 studies were selected, most of which involve the application of artificial intelligence rather than data science in document management. The analysis also revealed that many of the studies are recent, concentrated in the health sector, and originate from countries such as the United States and South Africa. It is concluded that there is a considerable amount of research addressing the intersection between document management and computing, and the proposed review protocol can contribute to the reliability and replication of research based on recall and precision measures.

Keywords: electronic document management; artificial intelligence; data science; systematic literature review.

1 INTRODUÇÃO

No contexto organizacional, basicamente toda a informação sobre a empresa e para a empresa é armazenada por meio de documentos (Medeiros; Amaral, 2010). O papel do documento tem se tornado ainda mais crítico, a partir das necessidades mercadológicas e da evolução tecnológica que tem sido notada nos últimos anos (Santos; Krawszuk, 2020). Conforme a Associação Brasileira de Normas Técnicas (ABNT, 2015), um atributo distintivo dos documentos em relação a outros recursos informacionais como atas de reunião, relatórios etc. é que seu conteúdo pode ser submetido a várias revisões ao longo do tempo, conforme a necessidade de informação da organização. Essa condição reforça a necessidade da adoção de práticas eficientes de gestão desses documentos contribuindo para a eficácia operacional nas organizações.

A área de gestão de documentos originou-se no fim do século XIX e, nessa época, mesmo aqueles documentos que não tinham mais relevância em um contexto específico continuavam sendo armazenados nos arquivos (Medeiros; Amaral, 2010). Como ainda não havia métodos sistemáticos para lidar com grandes quantidades de arquivos e documentos, criou-se um cenário crítico, que se intensificou durante a Segunda Guerra Mundial (Indolfo, 2007).

Segundo Vianna e Freitas (2019), a gestão de documentos compreende processos, métodos e operações para que as informações sejam organizadas e recuperadas da maneira mais eficiente possível. Medeiros e Amaral (2010) acrescentam que a gestão de documentos

propõe formas de intervir no ciclo vital de um documento buscando otimizar o processamento de grandes volumes de informação. Nessa direção, segundo Moreno (2008), em virtude da evolução tecnológica nas últimas décadas, especialmente no âmbito computacional, emergiu uma subárea denominada Gestão Eletrônica de Documentos (GED).

Santos (2005) e, mais recentemente, Weiss (2019) afirmam que a GED envolve um conjunto de tecnologias e ferramentas para garantir que todas as etapas no processo de gestão sejam implementadas de forma eficiente, segura e confiável fomentando a capacidade de uma organização em gerar conhecimento e adaptar-se às novas dinâmicas e tendências de mercado. Muitas dessas tecnologias são investigadas e providas pela área da Computação. Percebe-se, portanto, que a área de GED fortaleceu a necessidade de um olhar interdisciplinar considerando tanto a Ciência da Informação (CI), que abriga como uma de suas subáreas a gestão de documentos, quanto a Computação.

Na dimensão tecnológica, há duas grandes áreas cujas tecnologias têm recebido muita atenção nos últimos anos: Ciência de Dados (CD) e Inteligência Artificial (IA). O termo *Data Science* (Ciência de Dados) apareceu pela primeira vez, segundo CAO (2017), no livro "*Concise Survey of Computer Methods*" de Peter Naur, publicado em 1974. Conforme apresentado por Marchionini (2016), a área de Ciência de Dados é um campo interdisciplinar, que emerge a partir de quatro pilares fundamentais: Domínio do Conhecimento; Ciência da Computação; Estatística; e Ciência da Informação.

Fonseca (2021) ressalta que a Ciência de Dados, a partir de grandes volumes de dados e informações, permite revelar e extrair conhecimento valioso que servirá de base para o planejamento estratégico de uma organização visando à competitividade e à inovação. Dessa forma, a relação entre essa área e a GED é promissora e, conforme os resultados da revisão conduzida neste estudo, ainda não foi devidamente explorada no meio científico e no meio industrial.

Quanto à IA, sua origem remonta ao ano de 1956, quando o termo *Artificial Intelligence* foi adotado pela Faculdade de Dartmouth, nos Estados Unidos, em um evento sobre redes neurais e inteligência (Nakamiti, 2009). Conforme apontado por Vinagre e Moniz (2020), uma questão relevante sobre IA é justamente a sua definição, uma vez que conceituá-la exige a compreensão de dois elementos que a constituem: "inteligência" e "artificial". Verifica-se que não existe consenso ainda na literatura sobre essas definições.

Embora os primeiros conceitos de IA sejam provenientes da área da Computação, trata-se de um campo multidisciplinar (Sichman, 2021), envolvendo, como exemplos, as áreas de matemática, psicologia, linguística, filosofia, entre outras. Skalfist, Mikelsten e Teigens (2019) definem a IA como a capacidade de sistemas computacionais em aprender e executar tarefas que não podem ser realizadas a partir de algoritmos convencionais simulando, em muitos casos, a habilidade de resolver problemas e tomar decisões dos humanos.

Neste cenário, o objetivo foi analisar estudos da literatura para compreender os meios pelos quais as áreas de CD e, especialmente, IA podem ser incorporadas às práticas de gestão eletrônica de documentos, levando a melhorias no tratamento, organização, recuperação, e uso de informações. Para cumprir este objetivo, conforme ainda será detalhado, uma revisão sistemática da literatura tem sido conduzida, com base nas diretrizes propostas pelo Conselho Nacional de Pesquisa Médica e de Saúde da Austrália (NHMRC, 2000), para estudos na área da saúde, e adaptadas por Kitchenham (2004), para aplicação na computação, e por Petticrew e Roberts (2005), na área de ciências sociais.

Para guiar a condução da revisão sistemática, conforme estabelecido na literatura (Kitchenham, 2004; Kitchenham; Charters, 2007), foi elaborado o protocolo de revisão. Todas as etapas, escolhas e critérios adotados estão documentados no protocolo de revisão. Ressalta-se também que a condução desta revisão sistemática ainda está em andamento e os resultados iniciais do processo de análise são apresentados neste artigo.

2 REVISÕES SISTEMÁTICAS DA LITERATURA

Uma revisão sistemática da literatura é um tipo de revisão da literatura que define um meio mais rigoroso e replicável para coletar e sintetizar estudos científicos referentes a uma determinada área ou tópico de pesquisa. Conforme mencionado, todo o processo é guiado por um plano denominado protocolo de revisão, permitindo que os interessados na pesquisa avaliem a completude e a consistência da revisão e dos resultados obtidos (NHMRC, 2000). Há três fases que devem ser seguidas na condução de uma revisão sistemática da literatura (Kitchenham, 2004): Fase 1 - Planejamento da revisão; Fase 2 - Condução da revisão; e Fase 3 - Publicação dos resultados.

Na Fase 1, o protocolo de revisão é elaborado e revisado pelos pesquisadores. No mínimo, os seguintes elementos devem ser descritos: i) objetivo da pesquisa; ii) questões de

pesquisa; iii) idioma dos estudos; iv) bases de dados; v) *string* (expressão) de busca; vi) critérios de inclusão e exclusão; vii) critérios para avaliação da qualidade; e viii) métodos para extração e síntese dos dados.

É importante verificar que as questões de pesquisa (ii) são derivadas do objetivo (i) ou, colocado de outra forma, representam uma estratégia para dividir o objetivo em objetivos menores. A escolha do idioma (iii), das bases de dados (iv) e a definição da *string* de busca (v) são etapas básicas de qualquer tipo de revisão da literatura. Os critérios de inclusão e exclusão (vi) devem ser definidos de acordo com o domínio sob investigação e o escopo da pesquisa, e são aplicados aos estudos retornados, durante a revisão, para selecionar aqueles que sejam relevantes do ponto de vista das questões de pesquisa.

Os critérios para avaliação da qualidade (vii) servem para estabelecer um determinado nível de qualidade dos estudos retornados e, como os critérios de inclusão e exclusão, são definidos de acordo com as características da pesquisa. Os métodos para extração e síntese dos dados (viii) representam todos os meios, qualitativos e quantitativos, que serão utilizados para a extração e organização dos dados obtidos.

Na Fase 2, a *string* de busca é passada nas bases de dados selecionadas, e os estudos retornados são analisados a partir de duas etapas. Na primeira etapa, para cada estudo encontrado, são analisados título, palavras-chave, resumo e, em alguns casos, as seções de introdução e conclusões. Em seguida, os critérios de inclusão e exclusão são aplicados para que os pesquisadores decidam pela inclusão ou não do estudo. Na segunda etapa, os estudos selecionados são lidos de maneira completa, e os critérios de inclusão e exclusão são aplicados novamente.

Os dados desses estudos são extraídos e organizados, a partir dos métodos escolhidos, para responder às questões de pesquisa e discutir achados, lacunas e tendências relevantes. Esses dados também podem servir de base para a avaliação da qualidade do conjunto final de estudos. Finalmente, na Fase 3, o protocolo de revisão e os resultados são divulgados a partir de publicações em revistas e eventos científicos.

3 METODOLOGIA

No desenvolvimento e aplicação do protocolo de pesquisa deste artigo, foi utilizada, como base de conhecimento, a revisão sistemática da literatura proposta por Petticrew e

Roberts (2005), Kitchenham (2004) e Conselho Nacional de Pesquisa Médica e de Saúde da Austrália (NHMRC, 2000), que compreende caráter exploratório-descritivo e, para a análise e síntese dos estudos, adota uma abordagem combinando aspectos qualitativos e quantitativos.

Antes mesmo que o protocolo de revisão fosse elaborado, foi determinado o principal objetivo da pesquisa, que se refere ao mapeamento da interdisciplinaridade entre GED, IA e CD. Nesse ponto, foi avaliada a necessidade de elaboração de uma revisão sistemática. Para esse propósito, foram buscados estudos com o mesmo objetivo, isto é, estudos apresentando revisões sistemáticas que tratassem dessa temática, e nenhum trabalho foi encontrado.

A partir do objetivo de pesquisa, foram derivadas três questões: P1) Quais são os principais pontos de interseção entre GED, IA e CD?; P2) Quais desafios, tendências e lacunas de pesquisa estão presentes na interseção dessas áreas?; P3) Como as aplicações de IA e CD têm influenciado as práticas na GED? Cada pergunta foi definida com base na abordagem PICO (*Population - Intervention - Comparison - Outcome*), seguindo as orientações de Petticrew e Roberts (2005) e Kitchenham e Charters (2007).

Population (População) se refere às pessoas ou grupos afetados pelos resultados na condução da revisão sistemática. No caso, de forma ampla, a população é formada pelos profissionais e pesquisadores da Ciência da Informação, e, de forma específica, pelas equipes que atuam diretamente na GED. *Intervention* (Intervenção) é um termo genérico que designa quais são os principais objetos alvos da investigação, isto é, abordagens como, por exemplo, métodos, processos, ferramentas etc. Neste estudo, a intervenção é representada pelas tecnologias das áreas de IA e CD relacionadas com a GED.

Comparison (Comparação) não se aplica aqui, uma vez que o foco do estudo não é comparar as tecnologias identificadas na literatura com alguma abordagem de referência. *Outcome* (Resultado) é o principal resultado esperado com a revisão sistemática e, neste artigo, buscou-se obter resultados iniciais a respeito do panorama da interseção entre as áreas de IA e CD, com a área de GED, criando condições para investigações futuras em diferentes perspectivas.

Em relação às bases de dados, foram escolhidas as plataformas *Web of Science* e *Scopus*, porque ambas contemplam uma vasta cobertura de revistas científicas consolidadas. A escolha dessas bases também segue as diretrizes propostas por Dyba, Dingsoyr e Hanssen (2007), no contexto de uma revisão sistemática. Outra recomendação, conforme Kitchenham

(2004), é para que sejam evitadas exclusões de estudos unicamente com base no idioma. Dessa forma, foram consideradas publicações em português, inglês e espanhol.

Conforme enfatizado por Zhang e Muhammad (2013), o processo de busca em uma revisão sistemática é um fator crítico para que o protocolo de revisão seja adequadamente executado, e depende fortemente da qualidade da *string* de busca. Entretanto, conforme apresentado por Boell e Cezec-Kecmanovic (2011), a definição da *string* precisa envolver discussões entre especialistas no domínio investigado e exige um processo de refinamento contínuo. Além disso, conforme ainda os autores apresentam, quanto mais abrangente uma *string* de busca for, mais trabalhos irrelevantes poderão ser retornados; por outro lado, quanto mais específica uma *string* de busca for, mais documentos relevantes serão deixados de fora.

A partir das questões de pesquisa, foram definidas três palavras-chave para compor a *string* de busca: "*artificial intelligence*", "*data science*" e "*electronic document management*". Sinônimos e derivações foram definidos posteriormente. A *string* foi pensada de tal forma que seja possível encontrar estudos que contenham o termo "*electronic document management*" (ou derivados) obrigatoriamente e, no mínimo, um de dois termos: "*artificial intelligence*" ou "*data science*". A versão final da *string* de busca é apresentada no Quadro 1 a seguir.

Quadro 1 – Versão final da *string* de busca

("electronic document management system" OR "electronic record management system" OR "electronic recordkeeping system" OR "electronic document management" OR "electronic records management" OR "document management" OR "records management") AND ("data science" OR "artificial intelligence")

Fonte: Autores

Como pode ser observado pelo Quadro 1, a palavra-chave "*electronic document management*" possui derivações, que foram identificadas como sinônimos relevantes a partir da análise de trabalhos sobre GED na literatura. Em relação às tecnologias computacionais, foram especificadas apenas os termos significativos "*data science*" e "*artificial intelligence*" (sem derivações), por dois motivos básicos. O primeiro é que ambas as áreas envolvem uma ampla gama de abordagens, e estão em constante evolução, portanto, determinar tecnologias ou estratégias específicas dessas áreas seria não apenas inviável, mas poderia tornar a *string* semanticamente imprecisa.

Ligado ao motivo anterior, outro aspecto relevante é que uma *string* composta por termos menos abstratos situados nas áreas de IA e CD perderia a "validade" em um curto período afetando negativamente a reprodutibilidade da revisão. Portanto, na perspectiva do que é defendido por Boell e Cezec-Kecmanovic (2011), decidiu-se manter a *string* abrangente para garantir que mais trabalhos relevantes sejam recuperados, sob um custo de tempo maior na seleção posterior dos estudos. No entanto, observa-se que o custo é minimizado a partir da quantidade de revisores envolvidos (autores deste trabalho).

Para ajustar e validar a *string*, uma busca piloto foi conduzida nas bases de dados, em um fluxo iterativo, conforme as recomendações apresentadas em Brereton *et al.* (2007) e reforçadas em Fabbri *et al.* (2013). Esses autores propõem que o protocolo de revisão seja "pilotoado" com o objetivo de se descobrirem erros antes mesmo que a revisão sistemática seja oficialmente iniciada. Ainda conforme Fabbri *et al.* (2013), no contexto desta pesquisa, a *string* foi refinada até o ponto em que os revisores consideraram que o conjunto de estudos retornados era suficientemente adequado para responder às questões de pesquisa.

As decisões documentadas no protocolo de revisão e, em especial, as questões de pesquisa, permitiram orientar a especificação dos critérios de inclusão (CI) e de exclusão (CE), apresentados no Quadro 2 a seguir. Conforme pode ser observado, foram especificados dois critérios de inclusão e quatro critérios de exclusão. Também é importante observar que os critérios de exclusão prevalecem sobre os critérios de inclusão, isto é, se algum critério de exclusão for satisfeito, o estudo não será incluído, mesmo que algum critério de inclusão seja satisfeito também.

Quadro 2 – Critérios de inclusão e de exclusão

Tipo	Critério	Descrição
Inclusão	CI1	O estudo discute ou propõe técnica, estratégia, ferramenta ou qualquer abordagem de IA ou de CD ligada à GED.
Inclusão	CI2	O estudo não propõe uma abordagem específica, mas discute possibilidades, tendências, benefícios e implicações na intersecção entre GED, IA e CD.
Exclusão	CE1	O estudo é uma versão anterior de um trabalho mais completo sobre a mesma pesquisa com os mesmos autores.
Exclusão	CE2	O estudo está escrito em um idioma diferente de português, inglês e espanhol.
Exclusão	CE3	O texto completo não possui acesso aberto ou não pode ser acessado por outro motivo.
Exclusão	CE4	O estudo é um livro, capítulo de livro ou resumo expandido.

Fonte: Autores

Com base na especificação dos critérios, foram definidos os procedimentos para a seleção dos estudos e avaliação da qualidade do conjunto final. Conforme descrito, o processo de análise dos estudos foi organizado em duas etapas, e os critérios de inclusão e exclusão foram aplicados para a primeira etapa de seleção. Os mesmos critérios também devem ser aplicados para a segunda etapa de seleção, mas esta etapa não fez parte do escopo deste trabalho.

Para avaliar a qualidade do conjunto final de estudos, foram especificados quatro critérios de avaliação, conforme é mostrado no Quadro 3 a seguir. Foi adotado o instrumento de avaliação utilizado por Dyba, Dingsoyr e Hanssen (2007), composto por uma escala de três pontos: 1, se o critério é completamente atendido; 0,5, se o critério é parcialmente atendido; e 0, se o critério não é atendido. A pontuação final é determinada a partir do somatório das pontuações atribuídas a todos os estudos, estabelecendo uma medida quantitativa.

Quadro 3 – Critérios para avaliação da qualidade

Critério	Descrição
CQ1	Existe uma razão fundamental pela qual o estudo foi conduzido.
CQ2	As contribuições do estudo estão claras e de acordo com o objetivo apresentado.
CQ3	As limitações do estudo são discutidas.
CQ4	São apresentadas perspectivas de trabalhos futuros com base nas contribuições.

Fonte: Autores

Em relação aos métodos para síntese dos dados, serão adotadas técnicas descritivas para análise de conteúdo e técnicas apoiadas pela bibliometria (Hayashi *et al.*, 2010; Faria *et al.*, 2010). Vale ressaltar que o passo referente à síntese dos dados corresponde à segunda etapa de seleção e análise dos estudos, portanto, não será tratado aqui.

4 RESULTADOS

A condução da revisão sistemática (Fase 2) resultou em um número substancial de trabalhos recuperados e, na avaliação dos autores, uma proporção significativa de estudos selecionados para análise. Conforme apresentado, trata-se de uma revisão sistemática em andamento, e a primeira etapa de seleção e análise foi finalizada. Dessa forma, esta seção discute os resultados iniciais obtidos. Como resultado da busca nas bases de dados, foram retornados 213 estudos, dos quais, 41 são provenientes da base Web of Science e 172 são da

base Scopus. Após a retirada de trabalhos duplicados, foram obtidos 185 estudos, ao todo, para análise e seleção.

Inicialmente, uma pequena amostra dos estudos foi analisada de forma conjunta entre os autores para reduzir as possíveis diferenças entre as interpretações dos critérios de seleção, isto é, estabelecer uma calibragem e, conseqüentemente, aumentar a consistência da análise. Em seguida, os estudos foram divididos, atribuídos aos revisores, e o processo foi conduzido de maneira paralela. Na etapa de seleção e análise dos estudos, os revisores coletaram os seguintes dados para cada estudo: situação (incluído ou excluído), critério de inclusão (CI1 ou CI2), área abordada pelo estudo (IA, CD ou ambas), setor econômico alvo do estudo (por exemplo, Saúde; Ciência da Informação; Legislação etc.), comentários e dúvidas. As dúvidas foram anotadas para discussão posterior dos revisores acerca da inclusão ou não de um dado estudo.

Considerando o total de estudos (185), foram selecionados 54 estudos (29,2%). Em outras palavras, de todos os estudos retornados, foram identificados, na primeira etapa de seleção, 54 estudos que atenderam, no mínimo, a um critério de inclusão sem envolver nenhum critério de exclusão. Dois aspectos devem ser observados a partir deste quantitativo.

O primeiro aspecto é que muitos estudos excluídos não satisfizeram à *string* de busca, considerando-se os campos título, resumo e palavras-chave, mas foram retornados porque a expressão "*records management*" - que faz parte da *string* - foi encontrada no campo de palavras-chave indexado pela própria base de dados. O segundo aspecto é que, conforme apresentado na Seção de Metodologia, optou-se pela abrangência da *string* de busca, com o objetivo de identificar estudos que estivessem na "borda" do escopo desta pesquisa, isto é, estudos que pudessem conter quaisquer elementos que contribuíssem para responder às questões de pesquisa - perspectiva da revocação. Dessa forma, são aspectos que contribuem para justificar o alto índice de exclusão observado.

Em relação aos critérios de inclusão, 30 estudos (56%) foram selecionados com base no critério CI1, enquanto 24 estudos (44%) foram selecionados com base no critério CI2. Embora a diferença seja pequena, é interessante notar que há mais estudos dedicados à proposição e/ou discussão de abordagens específicas de IA ou CD aplicadas à GED. De fato, esses números ajudam a corroborar que o interesse por tecnologias ou técnicas específicas em ambas as áreas tem aumentado ao longo dos últimos anos.

Outro achado é que 90,74% dos estudos, portanto, 49 (soma da categoria IA + Ambas), envolvem a aplicação de tecnologias provenientes da área de IA na GED. Alguns exemplos encontrados nos estudos são aprendizagem de máquina (*machine learning*), aprendizagem profunda (*deep learning*) e redes neurais (*neural networks*). Esse dado permite inferir que existe não somente uma tendência de aproximação cada vez maior entre as áreas de IA e GED, mas revela uma área de pesquisa provavelmente muito pouco explorada ainda na literatura - há poucos trabalhos que investigam a área de CD no contexto da GED.

Apenas 19 (35,19%) dos estudos envolvem a aplicação de CD (soma da categoria CD + Ambas). Em síntese, tem-se: 35 trabalhos que abordam somente IA (64,81%); 5 trabalhos que lidam apenas com CD (9,26%); e 14 trabalhos relacionados a ambas, tanto IA quanto CD, correspondendo a 25,93% do número total. Portanto, esses quantitativos ajudam a responder parcialmente à questão de pesquisa P1 (Quais são os principais pontos de interseção entre GED, IA e CD?).

Uma constatação que vale ressaltar é que falta uniformização na forma de citar certas abordagens de IA e CD. Um exemplo disso é que a abordagem de aprendizagem de máquina, identificada em alguns trabalhos, é descrita de maneiras diferentes, dependendo da área de conhecimento envolvida no estudo. Em alguns casos, a aprendizagem de máquina é descrita como uma técnica e, em outros, como um processo, e assim por diante. Verifica-se, portanto, que essas diferenças semânticas criam dificuldades na categorização do conteúdo.

Em relação aos países associados aos estudos, os Estados Unidos são responsáveis por aproximadamente 11 (20%) dos estudos selecionados, seguidos pela África com 6 (11%) e China com 4 (7%). Vale observar que todos os estudos da África tratam de questões locais relacionadas à integração entre IA e GED, reforçando a preocupação do país com as melhorias que podem ser alcançadas na área de GED. A Europa responde por 26% (14) dos estudos, e não foram identificados quaisquer trabalhos provenientes do Brasil.

Considerando os anos de publicação, verificou-se que aproximadamente 80% (43) dos estudos selecionados foram publicados a partir de 2020. Além disso, há 9 estudos publicados no ano de 2024, o que contribui para reforçar que o interesse pela temática aqui tratada tem se tornado cada vez maior nos últimos anos. Observa-se que os resultados agrupados por distribuição geográfica e por ano fornecem elementos para responder à questão de pesquisa P2 (Quais desafios, tendências e lacunas de pesquisa estão presentes na interseção dessas áreas?).

A análise dos estudos também permitiu identificar os setores econômicos envolvidos. O setor da saúde tem sido o setor mais beneficiado pela integração entre IA, CD e GED, sendo o alvo de pesquisa em 23 estudos (42,59%). Os principais assuntos tratados nesse setor estão relacionados com a automação e previsão de diagnósticos, classificação de doenças, aspectos de privacidade de dados e interoperabilidade entre diferentes sistemas de informação de pacientes.

Os setores empresarial, público e científico foram alvo de pesquisa em 7 estudos cada um (12,96%). Os estudos envolvendo esses setores enfatizam a adoção de tecnologias de IA para melhorar aspectos de qualidade e segurança nos sistemas de gestão de documentos eletrônicos. Esses achados ajudam a responder às questões de pesquisa P2 (Quais desafios, tendências e lacunas de pesquisa estão presentes na interseção dessas áreas?) e P3 (Como as aplicações de IA e CD têm influenciado as práticas na GED?).

5 CONSIDERAÇÕES FINAIS

Diante do imediatismo e das pressões de mercado, as organizações têm sido levadas a adotar, cada vez mais, práticas de melhoria contínua no desenvolvimento de seus produtos e serviços. A informação contida em documentos e consequentemente a gestão de documentos fornecem a sustentação necessária para que as organizações adotem estratégias, incorporem melhorias e se adaptem às novas dinâmicas de maneira ágil.

Nesse contexto, surgem as tecnologias computacionais, capazes de fornecer múltiplas soluções para a gestão de documentos, tendo impulsionado o nascimento da área de gestão eletrônica de documentos. Por um lado, essas tecnologias fornecem meios para a resolução de problemas; por outro, criam a necessidade de uma compreensão maior sobre os impactos de sua utilização nos processos de gestão e no aumento do volume das informações e dados organizacionais.

Assim, neste artigo foi apresentada uma revisão sistemática da literatura para buscar, compreender e sintetizar estudos que abordem diferentes técnicas e estratégias das áreas de IA e CD, aplicadas à GED. A condução da revisão sistemática está em andamento, e é baseada no protocolo de revisão descrito aqui. Os resultados iniciais obtidos corroboram a relevância desta pesquisa, e mostram uma tendência acentuada especialmente na integração entre IA e

GED, com o objetivo de melhorar a qualidade na gestão de documentos e, como resultado, a usabilidade na recuperação da informação.

É importante destacar a necessidade de padronização nos termos relacionados às áreas de IA e CD, constituindo-se como um desafio a ser superado em outras pesquisas. Além disso, a distribuição geográfica dos estudos e a concentração no setor da Saúde sublinham áreas prioritárias e regiões líderes em pesquisa, enquanto a ausência de estudos do Brasil indica uma possível lacuna a ser explorada. Os resultados obtidos pavimentam um caminho para outras investigações e estudos na área.

Conclui-se também que o protocolo proposto e aplicado à recuperação de documentos pode contribuir para a confiabilidade e replicação de toda a pesquisa, seguindo as medidas de revocação e precisão.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO 9001:2015**: Sistemas de gestão da qualidade - Requisitos. Rio de Janeiro, 2015.

BOELL, S. K.; CEZEC-KECMANOVIC, D. Are systematic reviews better, less biased and of higher quality? EUROPEAN CONFERENCE ON INFORMATION SYSTEMS, 19., 2011, Helsinki, Finland. **Proceedings [...]**: ECIS 2011. Helsinki: [s.n.], 2011. Disponível em: https://www.researchgate.net/publication/221407655_Are_systematic_reviews_better_less_biased_and_of_higher_quality. Acesso em: 8 jun. 2024.

BRERETON, P.; KITCHENHAM, B. A.; BUDGEN, D.; TURNER, M.; KHALIL, M. Lessons from applying the systematic literature review process within the software engineering domain. **Journal of systems and software**, [S. l.], v. 80, n. 4, p. 571-583, Apr. 2007. Disponível em: <https://doi.org/10.1016/j.jss.2006.07.009>. Acesso em: 8 jun. 2024.

CAO, L. Data science: a comprehensive overview. **Acm Computing Surveys**, New York, v. 50, n. 3, p. 1-42, 29, Jun. 2017. Disponível em: <http://dx.doi.org/10.1145/3076253>. Acesso em: 7 jan. 2024.

DYBA, T.; DINGSOYR, T.; HANSSEN, G. K. Applying systematic reviews to diverse study types: an experience report. INTERNATIONAL SYMPOSIUM ON EMPIRICAL SOFTWARE ENGINEERING AND MEASUREMENT, 1., 2007, Madrid. **Proceedings [...]**: ESEM 2007. Madrid: [s. l.], 2007. p. 225-234. DOI: <http://dx.doi.org/10.1109/esem.2007.59>.

FABBRI, S. C. P. F.; FELIZARDO, K. R.; FERRARI, F. C.; HERNANDES, E. C. M.; OCTAVIANO, F. R.; NAKAGAWA, E. Y.; MALDONADO, J. C. Externalising tacit knowledge of the systematic review process. **IET Software**, Stevenage, v. 7, n. 6, p. 298-307, ago. 2013. Disponível em: <https://doi.org/10.1049/iet-sen.2013.0029>. Acesso em: 8 jun. 2024.

FARIA, L. I. L.; GREGOLIN, J. A. R.; HOFFMANN, W. A. M.; QUONIAM, L. Análise da produção científica a partir de publicações em periódicos especializados. Em: BRENTANI, R. R.; CRUZ, C. H. D. E. B.; SUZIGAN, W.; FURTADO, J. E. D. E. M. P.; GARCIA, R. D. E. C. (ed.). **Indicadores de ciência, tecnologia e inovação em são paulo 2010**. São Paulo: FAPESP, 2010. p. 4.1-4.71.

FONSECA, F. Whether or when: the question on the use of theories in data science. **Journal of the Association for Information Science and Technology**, [s. l.], v. 72, n. 12, p. 1593-1604, 11 jun. 2021. Disponível em: <http://dx.doi.org/10.1002/asi.24537>. Acesso em: 15 fev. 2024.

HAYASHI, M. C. P. I.; ZAUITH, G.; BELLO, S. F.; GOMES, C.; GUTIERREZ, R. F.; GUIMARÃES, V. A. L.; RIGOLIN, C. C. D. Sociologia da ciência: primeiras aproximações ao campo. **Revista Tecnologia e Sociedade**, Curitiba, v. 6, n. 11, p. 72-85, jul./dez. 2010.

INDOLFO, A. C. Gestão de documentos: uma renovação epistemológica no universo da arquivologia. **Arquivística.net**, Rio de Janeiro, v. 3, n. 2, p. 28-60, jul./dez. 2007.

KITCHENHAM, B. **Procedures for performing systematic reviews**. United Kingdom: Keele University; Australia: National ICT Australia, 2004.

KITCHENHAM, B.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. **Technical Report EBSE 2007-001**. United Kingdom: Keele University; Durham University, 2007.

MARCHIONINI, G. Information Science Roles in the Emerging Field of Data Science. **Journal of data and information science**, [s. l.], v. 1, n. 2, p. 1-6, 2016. Disponível em: <https://doi.org/10.20309/jdis.201609>. Acesso em: 15 fev. 2024.

MEDEIROS, N. L.; AMARAL, C. M. G. do. A representação do ciclo vital dos documentos: uma discussão sob a ótica da gestão de documentos. **Em Questão**, Porto Alegre, v. 16, n. 2, 2010.

MORENO, N. A. **Gestão documental ou gestão de documentos: trajetória histórica**. Londrina: Eduel, 2008.

NAKAMITI, E. K. **Agentes inteligentes artificiais**. 2009. Dissertação (Mestrado em Comunicação) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2009. Disponível em: <https://tede2.pucsp.br/handle/handle/5240>. Acesso em: 5 mar. 2024.

NATIONAL HEALTH AND MEDICAL RESEARCH COUNCIL (NHMRC). **How to review the evidence: systematic identification and review of the scientific literature**. Handbook series on preparing clinical practice guidelines. Canberra: National Health and Medical Research Council, 2000. 122 p. ISBN 186-4960329.

PETTICREW, M.; ROBERTS, H. **Systematic Reviews in the Social Sciences: a practical guide**. Oxford: Blackwell Publishing, 2005.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

SANTOS, H. M.; KRAWSZUK, G. L. Gestão do conhecimento organizacional: tratamento arquivístico para reuso da informação administrativa. **Investigación Bibliotecológica**, México, v. 34, n. 83, p. 103-127, abr./jun. 2020.

SANTOS, V. B. **Gestão de documentos eletrônicos: uma visão arquivística**. 2. ed. Brasília: ABARQ, 2005.

SICHMAN, J. S. Inteligência artificial e sociedade: avanços e riscos. **Estudos Avançados**, São Paulo, v. 35, n. 101, p. 37-50, jan./abr. 2021. Disponível em: <http://dx.doi.org/10.1590/s0103-4014.2021.35101.004>. Acesso em: 5 mar. 2024.

SKALFIST, P.; MIKELSTEN, D.; TEIGENS, V. **Inteligência artificial: a quarta revolução industrial**. Tradução de C.S.B Equipment. Cambridge: Cambridge Stanford Books, 2019.

VIANNA, W. B.; FREITAS, M. C. V. Gestão da informação e ciência da informação: elementos para um debate necessário. **Ciência da Informação**, Brasília, v. 48, n. 2, p. 191-208, maio/ago. 2019. Disponível em: <https://revista.ibict.br/ciinf/article/view/4800>. Acesso em: 5 mar. 2024.

VINAGRE, J.; MONIZ, N. Inteligência artificial: riscos e promessas. **Revista ciência elementar**, Porto, v. 8, n. 4, p. 1-9, 2020. DOI: 10.24927/rce2020.052. Disponível em: <https://rce.casadasciencias.org/rceapp/art/2020/052/>. Acesso em: 10 mar. 2024.

WEISS, M. C. Sociedade sensoriada: a sociedade da transformação digital. **Estudos avançados**, São Paulo, v. 33, n. 95, p. 203-214, 2019. Disponível em: <https://doi.org/10.1590/s0103-4014.2019.3395.0013>. Acesso em: 9 mar. 2024.

ZHANG, H., MUHAMMAD, A. Systematic reviews in software engineering: an empirical investigation, **Information and software technology**, [s. l.], v. 55, n. 7, p. 1341-1354, Jul. 2013. DOI: <https://doi.org/10.1016/j.infsof.2012.09.008>.