

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

EDENILZA VALÉRIA DA SILVA MAGALHÃES

**O PAPEL DA INTELIGÊNCIA ARTIFICIAL E DA CIÊNCIA DE DADOS
NA GESTÃO ELETRÔNICA DE DOCUMENTOS**

São Carlos - SP
2024

EDENILZA VALÉRIA DA SILVA MAGALHÃES

O PAPEL DA INTELIGÊNCIA ARTIFICIAL E DA CIÊNCIA DE DADOS NA GESTÃO
ELETRÔNICA DE DOCUMENTOS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação ao Departamento de Ciência da Informação da Universidade Federal de São Carlos, para obtenção do título de Mestre em Ciência da Informação.



Orientadora: Dra. Bárbara Coelho Neves

São Carlos - SP
2023

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência da Informação

Folha de aprovação.

Assinatura dos membros da comissão examinadora que avaliou e aprovou a Defesa de Mestrado da candidata Edenilza Valéria da Silva Magalhães, realizada em 20/12/2024:

Prof. Dr. Barbara Coelho Neves
PPGCI da Universidade Federal de São Carlos - USFCar

Prof. Dr. Roniberto Morato do Amaral
PPGCI da Universidade Federal de São Carlos - USFCar

Profa. Dra. Telma de Carvalho
PPGCI da Universidade Federal do Seará - UFS

Profa. Dra. Elaine Mosconi
Université de Sherbrooke - USD

RESUMO

A Gestão Eletrônica de Documentos é uma área da Ciência da Informação que incorpora um conjunto de tecnologias, ferramentas e sistemas, fundamentais na organização, preservação, proteção, distribuição e uso informações e recursos informacionais em contextos corporativos e institucionais, especialmente com o avanço da era digital. A Inteligência Artificial e a Ciência de Dados são tecnologias emergentes, que oferecem inovação por meio da automação de processos, análises avançadas e melhoria na recuperação de informações, aumentando eficiência e apoio à tomada de decisões. Diante dessa problematização, uma revisão sistemática da literatura foi conduzida à questão principal: qual a contribuição da Inteligência Artificial e Ciência de Dados na Gestão Eletrônica de Documentos, com o intuito de identificar padrões, desafios e oportunidades que possam contribuir para a evolução das práticas de gestão documental. Para tanto foi definido e aplicado um protocolo de revisão nas bases de dados *Web of Science* e *Scopus* e nas bases de dados Brapci e LISTA, que são específicas da área de Ciência da Informação, com um marco temporal de 2000 a 2024. Como resultado final, foram selecionados 18 artigos, em sua maioria pesquisas experimentais ou práticas, que permitiu uma visão ampla de abordagens desenvolvidas para aplicação de Inteligência Artificial e Ciência de Dados na Gestão Eletrônica de Documentos. Algumas lacunas foram identificadas, dentre elas falta de padronização nas terminologias, dificuldade na obtenção de grandes volumes de dados treinamento e testes, questões éticas e baixa adoção pelos profissionais da informação, ainda são fatores que impedem a implementação plena. Como possibilidade de trabalhos futuros, apresenta-se a possibilidade de ampliação desta pesquisa para incorporar as áreas de gerenciamento de registros da saúde e de documentos técnicos de projeto, áreas que não foram contempladas neste estudo. Outra possibilidade é a elaboração de um tesouro de Inteligência Artificial e Ciência de Dados, envolvendo uma equipe multidisciplinar, ou ainda a exploração de ferramentas como GitHub ou base de dados de patentes para identifica ferramentas desenvolvidas e disponíveis para uso. Com isso esta pesquisa busca contribuir para identificar oportunidades no uso de Inteligência artificial e Ciência de Dados na Gestão Eletrônica de Documentos.

Palavras-Chave: Gestão Eletrônica de Documentos. Inteligência Artificial. Ciência de Dados. Revisão Sistemática de Literatura. Ciência da Informação.

ABSTRACT

Electronic Document Management is a discipline within Information Science that encompasses technologies, tools, and systems critical to the organization, preservation, protection, distribution, and use of informational resources in corporate and institutional settings, particularly in the digital age. Artificial Intelligence and Data Science emerge as transformative technologies, enabling innovation through process automation, advanced analytics, and enhanced information retrieval, thus improving efficiency and supporting decision-making. To explore their contributions to this field, a systematic literature review addressed the core question: How do Artificial Intelligence and Data Science contribute to Electronic Document Management? The study aimed to uncover patterns, challenges, and opportunities that advance document management practices. A protocol was applied across databases, including Web of Science, Scopus, Brapci, and LISTA, covering the period from 2000 to 2024. Eighteen articles were selected, mostly experimental or practical studies, revealing diverse applications of Artificial Intelligence and Data Science in document management. However, the review identified key gaps, including the lack of standardized terminology, challenges in acquiring large datasets for model training and testing, ethical issues, and low adoption rates among information professionals, which limit full implementation. Future work could extend this research into specialized areas like health record management and technical project documentation. Another potential direction involves creating a multidisciplinary thesaurus for Artificial Intelligence and Data Science or utilizing tools like GitHub and patent databases to identify developed and available solutions. This study seeks to uncover opportunities for applying Artificial Intelligence and Data Science in Electronic Document Management.

Keywords: Electronic Document Management. Artificial Intelligence. Data Science. Systematic Literature Review. Information Science.

LISTA DE ILUSTRAÇÕES

<u>Figura 1 – Espiral do Conhecimento – Processo SECI</u>	21
<u>Figura 2 – Fundação da Ciência de Dados</u>	28
<u>Figura 3 – Interdisciplinaridade da Ciência de Dados</u>	34
<u>Figura 4 – Etapas de desenvolvimento da pesquisa</u>	42
<u>Figura 5 – Processo de definição de <i>string</i> de busca</u>	50
<u>Figura 6 – Fluxograma Prisma 2020</u>	62
<u>Figura 7 – Pesquisa nas bases de dados</u>	63
<u>Figura 8 – Duplicidade por base</u>	64
<u>Figura 9 – Representação das bases de dados</u>	65
<u>Figura 10 – Distribuição dos critérios aplicados em cada base de dados</u>	67
<u>Figura 11 – Aplicação dos critérios</u>	68
<u>Figura 12 – Países de Origem das pesquisas</u>	73
<u>Figura 13 – Abordagem de pesquisa</u>	76
<u>Figura 14 – Aplicações Práticas</u>	79

LISTA DE QUADROS

<u>Quadro 1 – Objetivo de pesquisa</u>	43
<u>Quadro 2 – Questões de pesquisa</u>	43
<u>Quadro 3 – Descrição dos elementos PICOC da Pesquisa</u>	44
<u>Quadro 4 – Descrição dos Critérios de Exclusão</u>	46
<u>Quadro 5 – Descrição dos Critérios de Inclusão</u>	47
<u>Quadro 6 – Descrição dos Critérios de Qualidade</u>	48
<u>Quadro 7 – Lista de termos e definições</u>	50
<u>Quadro 8 – Quantitativos dos termos de GD</u>	53
<u>Quadro 9 – Porcentagem da análise de relevância dos termos de GD</u>	54
<u>Quadro 10 – Porcentagem da análise de relevância dos termos de GD</u>	55
<u>Quadro 11 – String de busca para GED</u>	56
<u>Quadro 12 – Lista de termos d</u>	57
<u>Quadro 13 – String final de busca</u>	61
<u>Quadro 14 – Lista de termos encontrados</u>	70
<u>Quadro 15 – Universidades produtoras das pesquisas</u>	74
<u>Quadro 16 – Classificação quanto aos benefícios</u>	83
<u>Quadro 17 – Relação dos principais benefícios percebidos</u>	100

LISTA DE SIGLAS

Sigla	Descrição
AI	Artificial Intelligence
BD	Banco de Dados
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
Brapci	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
Capes	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CI	Ciência da Informação
DCI	Departamento de Ciência da Informação
DS	Data Science
CVD	Ciclo de Vida dos Dados
EDM	Electronic Document Management
EDMS	Electronic Document Management System
ENANCIB	Encontro Nacional de Pesquisa em Ciência da Informação
GED	Gestão Eletrônica de Documentos
GD	Gestão de Documentos
IA	Inteligência Artificial
IS	Information Science
TIC	Tecnologia de Informação e Comunicação
PC	Palavra-Chave
SGED	Sistema/Software de Gestão Eletrônica de Documentos
UFSCar	Universidade Federal de São Carlos

SUMÁRIO

1. INTRODUÇÃO.....	11
1.1. Estrutura da Dissertação	12
1.2. Caracterização do problema.....	14
1.3. Problema de Pesquisa.....	14
1.4. Hipótese.....	15
1.5. Objetivos da pesquisa	15
1.6. Justificativa	16
1.7. Delimitação do Escopo e Amostragem teórica	17
2. FUNDAMENTAÇÃO TEÓRICA	19
2.1. Ciência da Informação.....	19
2.2. Relação da CI com a IA e CD.....	30
2.3. Inteligência Artificial.....	31
2.4. Ciência de Dados	35
3. TRABALHOS RELACIONADOS.....	38
4. METODOLOGIA DE PESQUISA.....	40
5. DESENVOLVIMENTO DA PESQUISA.....	44
5.1. Protocolo de Revisão Sistemática de Literatura (Etapa 1.1).....	45
5.2. Definição e execução da String de busca.....	51
5.3. Tratamento e Seleção de Artigos (Etapa 2).....	66
6. ANÁLISE DE DADOS	75
7. DISCUSSÃO DOS RESULTADOS.....	85
7.1. P1 - Pontos de interseção entre GED e IA	86
7.2. P2 - Desafios, tendências e lacunas de pesquisa	100
7.3. P3 - Influência da IA nas práticas de GED.....	107
7.4. Trabalhos futuros.....	113
8. CONSIDERAÇÕES FINAIS	115
REFERÊNCIAS.....	119
APÊNDICE 1 – Prototipagem da String – IA, CD e GED	131
APÊNDICE 2 – Ficha de Anotações	132
APÊNDICE 3 – Artigo: MAGALHÃES, E. V. S.; ZANIRO, D. L.; SOARES, V. R. M. AMARAL, R. M.; NEVES, B. C. Inteligência Artificial e Ciência de Dados na Gestão de Documentos: uma Revisão Sistemática de Literatura.....	133

1. INTRODUÇÃO

A Gestão Eletrônica de Documentos (GED) é reconhecida como uma área essencial para a organização e preservação da informação em ambientes corporativos e institucionais. Sua relevância cresce diante do aumento exponencial de dados digitais, exigindo sistemas mais eficientes e inteligentes que possam atender às demandas modernas. Nesse contexto, compreender e aprimorar a GED é fundamental para garantir a integridade, segurança e acessibilidade das informações, fortalecendo sua aplicação em múltiplos setores.

Com a evolução tecnológica, a Inteligência Artificial (IA) e a Ciência de Dados (CD) emergem como forças transformadoras no campo da GED. Essas tecnologias não apenas potencializam a automação de tarefas repetitivas, mas também oferecem capacidades avançadas de análise, organização e recuperação de informações. Tais inovações podem impactar positivamente os sistemas, aprimorando sua eficiência, reduzindo custos e permitindo decisões mais informadas e ágeis.

Entretanto, a implementação dessas tecnologias na GED enfrenta desafios significativos. A ausência de padrões bem estabelecidos, os riscos éticos e sociais ainda pouco explorados limitam o pleno aproveitamento de seu potencial. Essas questões apontam para a necessidade de pesquisas mais aprofundadas que abordem tanto os aspectos técnicos quanto os impactos humanos dessas inovações.

A metodologia adotada baseia-se no planejamento e condução de uma Revisão Sistemática da Literatura (RSL), que se destaca tanto como processo quanto como produto, contribuindo para o avanço do conhecimento na área.

A pesquisa, de caráter exploratório-descritivo, combinou abordagens qualitativas e quantitativas para a análise e síntese dos estudos. A análise temática e de conteúdo simples foram utilizadas para identificar padrões e categorias emergentes, proporcionando uma visão detalhada das tendências e lacunas existentes. Esse enfoque permitiu uma compreensão profunda dos desafios e oportunidades relacionados à aplicação de IA e CD na GED.

Diante disso, esta pesquisa teve como objetivo mapear as principais técnicas e ferramentas de IA e CD aplicadas à GED, destacando suas contribuições,

limitações e possíveis direções futuras. Além disso, buscou avaliar os impactos éticos, sociais e organizacionais associados a essas tecnologias, apresentando um panorama com informações para a melhoria contínua da área. Assim, espera-se contribuir para o avanço da GED como um campo estratégico e interdisciplinar, essencial para o gerenciamento de informações na era digital.

Os resultados da pesquisa revelaram importantes pontos de interseção entre GED, IA e CD, destacando-se a automatização de processos, otimização do acesso à informação e suporte à tomada de decisão. Além disso, foram identificados benefícios como maior eficiência operacional e novas possibilidades de análise documental, bem como lacunas significativas, incluindo a falta de padrões, diversidade de técnicas, métodos e desafios éticos e sociais. Essas descobertas reforçam o papel estratégico dos profissionais da informação na construção, em cooperação com profissionais de outras áreas, de soluções tecnológicas mais inclusivas e eficazes.

1.1. Estrutura da Dissertação

A pesquisa está estruturada em sete seções. Na Introdução é apresentada a contextualização sobre do tema, o problema de pesquisa, que apresenta a questão central que orienta este trabalho, seguido da hipótese que sugere que a integração dessas técnicas na Gestão Eletrônica de Documentos proporcionará melhorias significativas em termos de efetividade e eficiência nas organizações.

Os objetivos da pesquisa são também apresentados nesta seção, de forma a destacar a busca por alternativas inovadoras que permitam lidar com o desafio da crescente produção documental. Na sequência justifica a relevância da pesquisa, apontando para sua contribuição ao desenvolvimento de soluções e tecnologias para a gestão eletrônica de documentos, além de apresentar novos caminhos para a área. Para finalizar o capítulo da introdução, a amostragem teórica e a delimitação do escopo são também discutidas, delineando o enfoque adotado.

Na segunda seção, intitulada Fundamentação Teórica, são apresentados os principais conceitos que conduzirão esta investigação, com foco na interconexão entre as áreas de Inteligência Artificial, Ciência da Informação e Gestão de

Documentos, visando aprimorar o processo de recuperação de informações estratégicas.

A terceira seção, Considerações Metodológicas, aborda a metodologia utilizada na pesquisa, que se trata de Revisão Sistemática de Literatura (RSL), com abordagem qualitativa-quantitativa. Na seção seguinte, intitulada Desenvolvimento da Pesquisa, são descritas e desenvolvidas as etapas da RSL e o protocolo de aplicação estabelecido. As etapas de refinamento e prototipagem da *string* com a seleção dos termos de GED, IA e CD, as buscas nas bases de dados são executadas e o protocolo de RSL é aplicado.

Nesta seção são também descritas todas as etapas ligadas ao tratamento e processamentos do resultado das pesquisas e são apresentados os quantitativos e algumas pré-análises. Como resultado dessa etapa temos o *corpus* é definido. Todos os textos escolhidos são lidos integralmente e são realizadas anotações de elementos e padrões.

Na sequência, na seção Análise de Dados, é apresentada a relação de todos os artigos do *corpus* e um breve resumo do assunto tratado. Em seguida são realizadas as análises quantitativas e alguns comparativos de dados. A partir da análise temática e de conteúdo, as questões de pesquisa são respondidas e os resultados são discutidos, trazendo pontos positivos e uma análise crítica sobre as descobertas, além de propostas de trabalhos futuros.

Por fim, na conclusão é realizado um apanhado de tudo o que foi apresentado e são feitas as considerações finais.

Neste trabalho, busca-se explorar e compreender as iniciativas em desenvolvimento no campo da Gestão Eletrônica de Documentos, com foco na aplicação de tecnologias emergentes como a Inteligência Artificial e a Ciência de Dados. O objetivo é identificar como essas tecnologias estão sendo utilizadas e de que maneira podem contribuir para aprimorar processos documentais, um cenário caracterizado pela crescente digitalização e pela necessidade de práticas mais eficazes na gestão de documentos e informações.

1.2. Caracterização do problema

A crescente digitalização e o aumento exponencial na produção de documentos destacam desafios críticos na gestão eficaz da informação. Frente ao volume, diversidade e complexidade crescentes dos documentos, a Gestão Eletrônica de Documentos precisa se adaptar para atender às exigências de eficiência e precisão. Embora os sistemas de GED sejam essenciais para organizar informações, eles encontram limitações diante das transformações na estrutura, formato e suporte dos documentos.

A descentralização na geração de informações exige repensar padrões e diretrizes, promovendo uma visão mais fluida, interconectada e estratégica da informação, essencial para organizações públicas e privadas. Tecnologias emergentes, como Inteligência Artificial e Ciência de Dados, oferecem novas abordagens para interagir com coleta e padronização de grandes volumes e diversidade de dados. Métodos como aprendizado de máquina, processamento de linguagem natural e mineração de dados permitem a automação de tarefas como classificação, indexação e extração de informações contextuais, otimizando a gestão e apoiando a tomada de decisões estratégicas.

Contudo, obstáculos como a ausência de padrões integrativos para sistemas de GED, complexidade de desenvolvimento, e lacunas éticas e sociais limitam a adoção dessas tecnologias. A transformação digital exige que a GED evolua de um repositório de documentos para uma ferramenta estratégica de gestão da informação. Nesse cenário, a integração de IA e CD apresenta oportunidades de automação, redução de custos, suporte preditivo à decisão, e maior eficiência organizacional.

1.3. Problema de Pesquisa

Tendo em vista o exposto e alinhado ao desafio que tem se intensificado pela demanda por sistemas mais inteligentes e eficientes, capazes de extrair valor estratégico dos dados e informações contidos nesses documentos, surgiu a oportunidade de explorar como essas tecnologias emergentes podem transformar a Gestão Eletrônica de Documentos. Neste contexto, a pesquisa buscou responder à seguinte questão central:

Qual a contribuição das tecnologias, como a Inteligência Artificial e a Ciência de Dados, para a Gestão Eletrônica de Documentos?

A partir das lacunas e desafios identificados na literatura e prática da área, esta investigação propõe explorar o potencial dessas tecnologias na construção de sistemas de GED mais eficientes, estratégicos e alinhados às demandas contemporâneas. Essa abordagem também visa formular hipóteses sobre a integração de IA e CD como elementos transformadores para o futuro da gestão documental.

1.4. Hipótese

A hipótese central é que a integração de Inteligência Artificial e Ciência de Dados na Gestão Eletrônica de Documentos tem o potencial de revolucionar o gerenciamento de documentos e informações, proporcionando soluções mais eficientes, estratégicas e alinhadas às demandas contemporâneas de organizações públicas e privadas.

Essa abordagem não apenas otimiza processos, mas também com o debate no campo das Tecnologias Informacionais na Ciência da Informação como um elemento-chave no fomento à inovação, posicionando-a como protagonista no desenvolvimento de tecnologias transformadoras e sustentáveis.

1.5. Objetivos da pesquisa

Com o intuito de delimitar o caminho a ser percorrido e alinhado ao problema de pesquisa, foram definidos os seguintes objetivos:

1.5.1. Objetivo geral

Investigar a integração de tecnologias, como a Inteligência Artificial e a Ciência de Dados, na Gestão Eletrônica de Documentos, com o intuito de identificar padrões, desafios e oportunidades que possam contribuir para a evolução das práticas de gestão documental.

1.5.2. Objetivos Específicos

- Levantar as pesquisas existentes sobre o uso de técnicas de Inteligência Artificial e Ciência de Dados no processo de Gestão Eletrônica de Documentos;

- Mapear e classificar essas técnicas de Inteligência Artificial e Ciência de Dados utilizadas na Gestão Eletrônica de Documentos a partir da análise de conteúdo;
- Analisar como essas técnicas podem melhorar a eficiência da Gestão Eletrônica de Documentos no acesso, recuperação, tratamento, análise, uso e segurança de informações.

1.6. Justificativa

Os temas relacionados à e Inteligência Artificial Ciência de Dados, embora pareçam recentes, remontam da década de 1950. Novas tecnologias foram desenvolvidas e estão avançando em um ritmo acelerado, trazendo inovações inimagináveis há uma década.

Essas tecnologias deixaram de ser apenas promessas de futuro e tornaram-se parte integral do presente. Elas já influenciam diretamente nosso cotidiano de formas sutis, muitas vezes imperceptíveis, seja por meio de dispositivos inteligentes, algoritmos de recomendação, ou automação de processos.

A lacuna identificada na revisão sistemática de Dorneles (2024), onde constatou-se um baixo volume de publicações científicas na área de Ciência da Informação (CI), evidenciando poucas inovações tecnológicas no campo.

Isso reflete uma questão crítica para o campo da Ciência da Informação: a necessidade de ampliar e diversificar o entendimento sobre as possibilidades de integração entre Inteligência Artificial e Gestão Eletrônica de Documentos. Embora a gestão documental seja uma área essencial para instituições públicas e privadas, sua evolução tecnológica tem sido limitada, conforme evidenciado pelo foco quase exclusivo no Sistema Eletrônico de Informações (SEI) e pela escassez de publicações científicas inovadoras.

Essa carência de estudos compromete a capacidade de explorar soluções mais avançadas que atendam às demandas contemporâneas, como o tratamento de grandes volumes de dados, a segurança da informação e a eficiência no acesso e uso de documentos. O avanço de tecnologias de IA e Ciência de Dados oferece oportunidades únicas para superar essas limitações, permitindo a identificação de padrões, a automação de processos e a melhoria da tomada de decisão no âmbito da gestão documental.

Dessa forma, investigar como essas tecnologias podem ser aplicadas à GED contribui não apenas para preencher um vazio no conhecimento acadêmico, mas também para oferecer subsídios teóricos que ampliem o impacto da Ciência da Informação na transformação digital de instituições públicas e privadas.

1.7. Delimitação do Escopo e Amostragem teórica

Para esta pesquisa serão utilizadas as bases de dados Brapci, LISTA, *Scopus* e *Web of Science*, sendo as duas primeiras da área da Ciência da Informação e as outras duas abarcam diversas áreas de conhecimento. Com essa abordagem espera ser possível ter maior cobertura e abrangência na busca em relação às publicações científicas, pois determinados artigos ou periódicos podem estar presentes em uma base e ausentes em outra, isso será levado em consideração, visto que pode haver variações nos resultados encontrados.

Braga (2007, p. 34) apresenta que existem quatro tipos de triangulação metodológica (de dados, do investigador, de metodologia, de teorias), que consiste no uso de diferentes abordagens na tentativa de compreender de maneira mais efetiva o objeto ou fenômeno estudado. Para esta pesquisa escolheu-se adotar a triangulação de dados, visto que uma única fonte ou base de dados não seria suficiente para fornecer o aprofundamento necessário para esta pesquisa.

Com isso, o uso de múltiplas bases de dados bibliográficas pode enriquecer o levantamento bibliográfico e a análise das pesquisas, ampliando a cobertura e diversidade dos dados. No entanto, ciente das diferenças entre as bases de dados faz-se necessário adaptações metodológicas adequadas para minimizar problemas como duplicação de registros e garantir resultados precisos e confiáveis.

Faz-se necessário adaptar a metodologia de busca para cada base de dados utilizada, levando em consideração suas particularidades, recursos de pesquisa e critérios de indexação. As palavras-chave, operadores booleanos e filtros utilizados também podem variar entre as bases e exige análise por parte do pesquisador.

O foco da literatura estará em artigos científicos, publicados no Brasil ou no exterior, em português, inglês ou espanhol, por autores brasileiros e estrangeiros,

que estejam disponíveis gratuitamente em texto integral, que versem sobre o uso de técnicas de Inteligência Artificial na Gestão Eletrônica de Documentos.

Como recorte temporal, optou-se por buscar artigos publicados entre os anos de 2000 à 2024, visto que as aplicações de inteligência artificial, como conhecemos hoje, segundo Cozman e Neri (2021), iniciou-se no novo milênio. Com a “[...] explosão do poder computacional, não apenas embutido em computadores pessoais, mas também em câmeras e telefones de todos os tipos, veículos, eletrodomésticos [...]”, como também no volume de dados coletados de automaticamente de equipamentos e pessoas conectados (Cozman; Neri, 2021).

Na seção seguinte é apresentada a fundamentação teórica que esclarece conceitos importantes para o entendimento de desenvolvimento essa dissertação.

2. FUNDAMENTAÇÃO TEÓRICA

Esta pesquisa explora o papel da Inteligência Artificial e Ciência de Dados na Gestão Eletrônica de Documentos com o intuito como essas tecnologias podem ser usadas para otimizar processos e maximizar o uso de informações para a tomada de decisões.

Essa convergência permite não apenas a automação de tarefas operacionais, como a classificação, indexação e recuperação de documentos, mas também a identificação de padrões e insights estratégicos a partir da análise de dados. Enquanto a Ciência da Informação oferece o contexto necessário para transformar dados em informações relevantes, focando na organização, no acesso e no uso efetivo desses elementos.

A aplicação conjunta dessas técnicas e tecnologias busca transformar a gestão documental tradicional em uma abordagem mais dinâmica, preditiva e alinhada com as necessidades estratégicas das instituições públicas e privadas.

Por meio da Inteligência Artificial, a Ciência da Informação pode identificar padrões e tomar decisões baseadas em dados, criando sistemas e estratégias para gerenciar todo o ciclo de vida documental. Isso garante sua disponibilidade e utilidade, tornando a gestão documental essencial para otimizar tempo e recursos ao fornecer informações estratégicas.

Nas próximas seções, serão apresentados os pilares fundamentais das áreas de Inteligência Artificial, Ciência de Dados, Ciência da Informação e Gestão de Documentos. Estes conceitos formam a base essencial para uma compreensão do presente trabalho. Contudo, é importante ressaltar que o objetivo não é exaurir todos os conhecimentos associados a cada uma dessas disciplinas. Em vez disso, será apresentada uma visão introdutória sólida, abrindo caminho para as discussões sobre a integração destas áreas no contexto da gestão documental e da tomada de decisões estratégicas nas organizações.

2.1. Ciência da Informação

A Ciência da Informação como campo científico pode ser considerada uma área relativamente nova, segundo Freire (2006) sua origem remonta ao início do século XX, período em que Paul Otlet e Henri La Fontaine preocupados em tornar acessível todo conhecimento científico produzido, vislumbraram o conceito da universalidade da

documentação, que deu origem ao Tratado de Documentação, considerado um dos primeiros documentos preocupados em discutir as problemáticas que envolvem a informação e identificar uma forma de tornar todo conhecimento registrado disponível para quem precisa ter acesso.

Após a Segunda Guerra Mundial, houve uma explosão na produção científica. Em diversos países despertou-se o desejo em discutir os aspectos da informação e conhecimento explícito, por meio da criação de instituições de pesquisa, periódicos ou artigos científicos. Alguns tinham seu interesse voltado nos processos de produção e disseminação, outros nos métodos de reprodução, outros ainda, nas tecnologias de armazenamento e disseminação da informação para grupos de cientistas (Freire, 2006).

O objetivo da CI está em investigar a informação quanto às suas propriedades, comportamentos, fluxos e como é processada, visando a forma ideal de acessibilidade e usabilidade. Borko (1968, p. 2, tradução nossa) complementa que a “Ciência da Informação está preocupada com o corpo de conhecimentos relacionados à origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação, e utilização da informação”.

Estas práticas são estudadas pela CI tanto através de sistemas naturais, quanto artificiais, por meio da utilização de técnicas e tecnologias voltadas à sua transmissão e recuperação (Borko, 1968). O autor descreve a informação como o centro de estudos da CI, sendo ela responsável por realizar a comunicação e disseminar conhecimento explícito em diversas áreas do conhecimento.

Nesse sentido, a informação tem papel transformador na sociedade, segundo Freire (2006). Já nos primórdios dos estudos de CI, Otlet acreditava que ela seria capaz, por meio da comunicação intelectual, de provocar mudanças sociais e poderia, inclusive, trazer a paz mundial.

A CI por natureza é interdisciplinar possui relacionamento com diversas disciplinas que estão em constantes mudanças e evolução, esta definição é apresentada por Saracevic (1995) como a primeira de três classificações características da Ciência da Informação.

Neste sentido, Borko (1968) reafirma o caráter interdisciplinar da CI e complementa que esta característica está presente desde sua origem, e tem se

perpetuado até hoje. Ela se dá a partir das diferenças de formação dos profissionais e pesquisadores e deriva “de campos relacionados, tais como a Matemática, Lógica, Linguística, Psicologia, Ciência da Computação, Engenharia da Produção, Artes Gráficas, Comunicação, Biblioteconomia, Administração, e outros campos científicos semelhantes” (Borko, 1968, p. 2, tradução nossa).

A Ciência da Informação permeia pela ciência pura “visto que investiga seu objeto sem considerar sua aplicação” e pela ciência aplicada uma vez que “desenvolve serviços e produtos” (Borko, 1968, p. 2, tradução nossa). A disciplina responsável por estudar a informação e seu ciclo de vida é a Gestão da Informação.

De forma geral, a Ciência da Informação é uma área dinâmica e de grande importância, cujas raízes se estendem desde os visionários Otlet e La Fontaine até as atuais complexidades do mundo da informação digital. A interdisciplinaridade, característica fundamental da CI, permanece como seu alicerce, incorporando uma ampla gama de disciplinas para abordar as complexidades inerentes à gestão e à disseminação da informação, transformando a forma como acessamos informações, gerando grande impacto na forma como a sociedade se desenvolve e evolui.

2.1.1. Tríade: Dado, Informação e Conhecimento

Para estabelecer uma base conceitual sólida é importante explorar a tríade dado, informação e conhecimento. Esses conceitos constituem a espinha dorsal da Ciência da Informação, fornecendo a estrutura necessária para analisar como dados são processados e transformados em informações e, posteriormente, em conhecimento evoluindo para o desenvolvimento de competências e gerando inovações.

Compreender suas distinções e inter-relações possibilita uma melhor compreensão dos processos informacionais e o papel da gestão de documentos e tecnologias no cenário contemporâneo. Assim, esta seção abordará os significados específicos de cada conceito e suas implicações no contexto da Ciência da Informação.

Os dados foram identificados por Davenport e Prusak (1998) como um conjunto de fatos ligados a eventos e observações, que podem ser facilmente estruturados, quantificáveis e passíveis de transferência e utilização por máquinas. Os dados podem ser registrados em documentos, registros, formulários, páginas da internet etc. e podem estar em formato físico ou digital.

Eles podem ser extraídos destes documentos, como também de produtos, fluxos de trabalho e processos. Eles podem ser utilizados para gerar valor em tomadas de decisões que visam melhorar o desempenho da empresa e otimizar os processos (Saarikko, 2020).

Em um computador, os dados podem ser organizados e vinculados entre si, formando estruturas de dados. O processamento de dados em um computador envolve principalmente manipulações estruturais dos dados, como formatação, ordenação, comparação, estatísticas, e é executado por meio de programas, que são funções matemáticas também consideradas dados. Portanto, os dados são entidades matemáticas e puramente sintáticas que podem ser manipuladas e processadas por um computador (Setzer, 2015).

Estes dados tornam-se informação quando são imbuídos de sentido, significado e passam a ser utilizados pelos indivíduos ou grupos (Davenport; Prusak, 1998). Por isso, o dado pode ser considerado a matéria prima para a informação, ou ainda uma informação bruta (Hoffman, 2009).

A informação pode ser classificada, segundo Moresi (2000), de acordo com a função que desempenha, pode ser classificada como informação crítica, informação mínima, informação potencial ou informação sem interesse. A partir desta classificação será possível estabelecer os esforços para seu tratamento e processamento. O autor apresenta que a busca e manutenção da informação crítica, mínima e potencial devem ser priorizadas respectivamente, enquanto a informação sem interesse deve despender esforços apenas para evitar desperdício de recursos.

O dado leva à informação, que por conseguinte, leva ao conhecimento (Hoffman, 2009). Quando os dados e informações são coletados, processados e interpretados gera, oportunidade para a criação de novos conhecimentos, que está associada à sua capacidade de desenvolver novos métodos, processos e ferramentas, disseminá-los nos diferentes níveis organizacionais e incorporá-los aos demais produtos, serviços e sistemas.

Para Takeuchi e Nonaka, (2008) o conhecimento pode ser classificado em dois tipos: Conhecimento Explícito que são dados e informações passíveis de serem registrados, armazenados e compartilhados e Conhecimento Tácito que está relacionado às experiências, habilidades e valores inerentes aos indivíduos, difícil de

ser registrado ou compartilhado. A interação entre os diferentes tipos de conhecimento permite a geração e compartilhamento do conhecimento, conforme apresentado na **Figura 1**.

Figura 1 – Espiral do Conhecimento – Processo SECI



Fonte: Adaptada de Takeuchi e Nonaka (2008)

Segundo Valetim (2005), acredita-se que o conhecimento pode realmente ser construído a partir da interação entre as pessoas. Em organizações essa interação pode se dar através de reuniões, projetos, grupos de trabalho, através de análises e reflexões, na execução e desenvolvimento de atividades e na elaboração de metodologias e procedimentos de trabalho. São diversas as formas em que uma organização pode fazer uso de dados e informações existentes para desenvolver novos conhecimentos.

Nas organizações, a inovação depende da transmissão e difusão de dados, informações e conhecimentos existentes. Além disso, ela depende de sua associação a conhecimentos que ainda deverão ser construídos em função das novas demandas que estão surgindo para alcançar inteligência e vantagem competitiva. Ainda vale observar que a produção de novos conhecimentos nos tempos atuais “é muito mais rápida do que em outros tempos” (Weiss, 2019)

Grandes volumes de dados podem ser registrados em documentos, registros e sistemas quando tratados e processados transformam-se em informação, que por sua vez, pode ser reorganizada e disposta em documentos, registros e sistemas. Quando há uma compreensão e análise dessa informação, que podem alimentar ou ser alimentados por processos de trabalho, novos conhecimentos são gerados.

A partir da síntese do conhecimento e dos processos de trabalho combinados a outros dados e informações, é gerada a inteligência, que abre caminho para inovação e alcançar a inteligência competitiva. As experiências, em consequente, são armazenadas em forma de registros, normas e procedimentos, melhorando os processos de trabalho, criando novos dados e informações, retroalimentando a base da pirâmide.

Se os dados gerados a partir de processos de trabalho, produtos e serviços conectados forem combinados (informação), os fluxos e processos podem ser analisados para encontrar padrões e comportamentos (conhecimento). A partir do processo de análises, decisões podem ser tomadas, como por exemplo, como melhorar produtos, serviços e otimizar processos Saarikko (2020).

Compreender os conceitos fundamentais da tríade dado, informação e conhecimento consolida a análise dos processos informacionais, facilitando a compreensão de seu papel na Gestão Eletrônica de Documentos (GED). Essa área utiliza técnicas e ferramentas para gerenciar dados, informações e conhecimentos, bem como seus fluxos nos diferentes suportes.

2.1.2. Gestão Eletrônica de Documentos

A Gestão Eletrônica de Documentos (GED) desempenha um papel central na organização, preservação e acessibilidade da informação em organizações contemporâneas, unindo ferramentas tecnológicas e práticas documentais. Com o avanço das Tecnologias da Informação e Comunicação (TIC), tornou-se essencial gerenciar o crescente volume de documentos digitais de forma eficiente.

O documento é considerado um suporte para apresentação de dados, informações e conhecimentos explícitos. Ao passo que a informação se tornou um ativo importante para o desenvolvimento das organizações, os documentos assumiram o papel de registrar, armazenar, transmitir e comprovar situações, ações, eventos e processos em seu cotidiano (Santos; Krawszuk, 2020).

Os documentos são gerados e recebidos pelas organizações, sendo definido como “[...] um conjunto de informações registradas em um suporte. Para que esses cumpram sua função social, administrativa, técnica, jurídica, cultural” (Calderon *et al.*, 2004, p.97). Para tanto, é essencial garantir sua organização, preservação e acesso.

No entanto, nem tudo deve ser chamado de documento, segundo a Associação Brasileira de Normas Técnicas (ABNT 9001:2015) existe uma diferença entre documento e registro, de acordo com sua função. O documento possui conteúdo que estabelece um direcionamento, instruções e orientações para ações das organizações e funcionários, como exemplo estão os procedimentos, normas, instruções, documentos técnicos, desenhos técnicos, cronogramas etc.

Uma característica importante dos documentos é a possibilidade de sofrerem revisões ao longo do tempo para atualização de seu conteúdo, enquanto os registros possuem um caráter mais definitivo, eles retratam uma situação específica em um determinado momento do tempo e espaço e, uma vez finalizados, não podem ser alterados. Exemplos de registros incluem atas de reunião, relatórios, cartas, listas de participação/presença em Treinamentos etc. (ABNT 9001:2015).

Com o avanço das Tecnologias da Informação e Comunicação (TIC) tornou-se mais rápido e mais fácil a produção, edição, disseminação ou exclusão de documentos e registros em meios físicos e digitais (Santos; Krawszuk, 2020). O que desencadeou um crescimento exponencial e algumas vezes caótico de documentos, dificultando a recuperação da informação.

A transição dos suportes de informação para as mídias digitais intensificou a relevância da gestão documental no contexto informacional, atribuindo-lhe um papel central e transformando-a em um componente essencial para a produção e organização do conhecimento. (Santos; Silva; Bari, 2021, p.3)

No entanto, o manejo de grandes volumes documentais, abrangendo uma diversidade de formatos, tornou-se um desafio significativo para as instituições. A falta de conhecimento especializado frequentemente resulta em arquivamentos sem critérios ou tratamento adequado (Calderon *et al.*, 2004).

Como consequência, a agilidade na recuperação da informação fica comprometida e o acesso a essa documentação fica restrito aos responsáveis pelo seu arquivamento, o que limita sua utilização e dificulta a extração plena de seu potencial informacional, “propiciando situações frustrantes nas tomadas de decisões” (Elias, 2012, p. 15)

A afirmação de Calderon *et al.* (2004) reforça que nas organizações há grandes massas documentais, onde estão registradas parcelas significativas de sua história.

Calderon *et al.* (2004, p.100) considera também que “[...] existirem valores técnicos, científicos, jurídicos, probatórios”. No entanto, é comum que essa riqueza de informações esteja diluída em meio a uma grande quantidade de documentos sem valor aparente.

A gestão de documentos envolve uma série de atividades essenciais para garantir a eficiência e a organização do fluxo de informações em uma organização. Isso inclui a análise da tipologia documental, sua relevância, vigência e qualidade informativa, bem como a padronização de documentos e formulários. Além disso, a informatização dos processos de tramitação documental, a coordenação entre os produtores de documentos e os arquivos, e a regulamentação das transferências são fundamentais. (Calderon *et al.*, 2004).

Com as novas Tecnologias digitais o acesso e o compartilhamento de dado, informação e conhecimento ocorre em tempo real e em grandes volumes. Faz-se necessário desenvolver metodologias, critérios e ferramentas, que irão garantir seu armazenamento de forma organizada, estruturada e sistemática. Com isso seu compartilhamento torna-se eficiente e pode facilitar a tomada de decisão, garantindo o acesso à informação correta, no momento em que se precisa dela (Weitzen, 1991).

Essa é a função da Gestão Eletrônica de Documentos, uma área que incorpora um conjunto de tecnologias, ferramentas e sistemas responsáveis por criar, organizar e conservar documentos e recursos informacionais, possibilitando acesso seguro a estes recursos (Vianna; Freitas, 2019).

A Gestão Eletrônica de Documentos “[...] é ao mesmo tempo: um método, um sistema e uma tecnologia” (Andrade, 2002). A GED deve assegurar o uso inteligente de dados, informação e conhecimento fazendo a conexão entre o físico e o digital, contribuindo para o aprimoramento dos processos de trabalho nas organizações e, com seu amadurecimento, gerar valor a partir da criação e compartilhamento de conhecimentos.

Mas uma grande lacuna na literatura está em apresentar ações concretas que evidenciem como dados, informações e conhecimentos explícitos, gerados a partir de documentos organizacionais, podem ser extraídos, estruturados, classificados e utilizados como fonte para gerar informações estratégicas para auxiliar na tomada de decisão e apoiar no avanço da Transformação Digital nas organizações.

2.1.3. O Profissional da Informação

Os profissionais da informação são especialistas responsáveis por gerenciar, organizar, recuperar e analisar informações em diversos formatos, contextos organizacionais e sociais. Eles incluem bibliotecários, arquivistas, cientistas da informação, especialistas em gestão de documentos, documentalistas e museólogos. Esses profissionais possuem conhecimentos técnicos e metodológicos para garantir o acesso eficiente a informações, atuando na indexação, recuperação, gestão de acervos e desenvolvimento de estratégias informacionais (Almeida Júnior, 2002).

Segundo Richard Mason (1990 *apud* Lima *et al.*, 2017) o profissional da informação é “capaz de fornecer a informação correta de fonte confiável ao cliente certo, no momento certo e de forma assertiva a um custo em que seu uso seja justificável”, e afirma ainda que este termo engloba várias profissões provenientes de diversas áreas como sociais, tecnológicas e administrativas.

No IV Encontro de Diretores das Escolas de Biblioteconomia e Ciência da Informação do Mercosul, realizado em Montevideu, em 2000, (Programa, 2000 *apud* Valentim, 2002, p. 122) foram definidas uma série de competências do profissional da informação divididas em quatro categorias:

Competência de Comunicação e Expressão: ligadas aos processos de gestão de projetos de informação, de desenvolvimento de produtos e serviços voltados à disseminação da informação, atendimento ao usuário e divulgação de produtos e serviços de informação;

Competências Técnico-Científicas: associadas à especificidade da área com o desenvolvimento de métodos, técnicas, processos de trabalho e ferramentas que visam o tratamento, processamento, registro, recuperação e disseminação da informação em diversos meios e independentemente do suporte;

Competências Gerenciais: voltadas ao planejamento e gerenciamento de unidades, sistemas, serviços de informação;

Competências Sociais e Políticas: ligadas ao desenvolvimento de relacionamento com os diversos atores da sociedade para identificação de demandas sociais de informação, articulação para desenvolvimento, promoção e defesa da profissão.

O profissional da informação desempenha um papel essencial na criação de valor dentro das organizações, sendo responsável por gerenciar todo o ciclo de vida da informação. Eles possuem acesso a múltiplos setores institucionais, interagindo

com diferentes departamentos e contribuindo para a integração organizacional. Dotados de habilidades avançadas, esses profissionais têm a capacidade de localizar informações ainda desconhecidas, utilizando técnicas especializadas (Neves; Longo, 2000).

Com habilidades estratégicas para atender às demandas e interações entre diversas áreas institucionais, são considerados canais de distribuição de conhecimento por toda organização (Davenport; Prusak (1998).

Essas habilidades incluem métodos e técnicas essenciais para a organização e recuperação da informação, especialmente voltados para superar barreiras culturais e geográficas no acesso ao conhecimento e à informação (Morais; Ramalho, 2019).

Com a chegada das tecnologias digitais os profissionais precisam estar atentos e desenvolver constantemente habilidades e competências para se adaptarem às mudanças e colaborarem de forma ativa com o crescimento e desenvolvimento das organizações. Segundo Neves (2018, p. 332), é essencial, para os profissionais da informação, este aprofundamento, “se desejam continuar uma interação com seus usuários”, mantendo um ambiente vivo tanto no contexto físico como digital.

Os avanços tecnológicos alcançados pela Inteligência Artificial podem ser percebidos em diversas áreas como as de redes neurais, processamento de linguagem natural, aprendizado de máquina e agentes inteligentes, que impactaram em transformações econômicas, sociais e políticas.

Essas transformações refletem uma era social impulsionada pelos avanços tecnológicos e por uma nova cultura informacional, cujo valor está cada vez mais no acesso e uso inteligente da informação para a construção do conhecimento (Belluzzo, 2011, p. 59). Nesse cenário, o profissional da informação deve estar preparado e qualificado para enfrentar os desafios e oportunidades desse contexto em constante evolução.

Segundo a autora, a informação e o conhecimento têm recebido atenção especial na sociedade e nas organizações, o que evidencia a importância do profissional da informação que deve atuar como aliado às novas tecnologias, com participação ativa como colaborador em seu desenvolvimento, como mediador ao acesso à informação e ao conhecimento, além de atuar de forma crítica na elaboração de políticas de implementação e monitoramento dessas tecnologias.

Este profissional tem como seu principal objeto de trabalho a informação como processo, “algo construído, essencialmente histórico e cultural, que só pode ser apreendido na perspectiva dos sujeitos que a produzem, e disseminam e a utilizam” (Araújo, 2009, p. 203). Sendo assim, a informação não é vista apenas como um objeto estático e sim como algo inerente ao contexto em que se insere.

Com isso, é preciso possuir competências e habilidades para lidar com a informação “subjéctiva, dotada de sentidos diversos e inserida no terreno da experiência histórico-cultural” (Araújo, 2009, p. 203).

No que se refere às competências técnico-científicas, alguns conhecimentos de domínio do profissional da informação se destacam como fundamentais para o desenvolvimento e contínuo aprimoramento de ferramentas e técnicas de IA:

Ontologias, Tesouros e Vocabulários Controlados: muito utilizada pela Ciência da Informação, apresenta a “possibilidade de melhorar significativamente a representação de um domínio do conhecimento” (Ferneda; Dias, 2015, p. 3) por meio de estruturação, definição, significação e contextualização de termos que permite identificar o verdadeiro significado, eliminando as ambiguidades dos termos em um dado contexto, o que tem consequência direta na qualidade da recuperação das informações realizadas por agentes inteligentes e outras ferramentas de IA.

Classificação e Indexação de informações: são técnicas largamente utilizadas pelos profissionais da informação que mostrou-se muito eficiente nos processos de automatização da recuperação de informação no âmbito da IA por agentes inteligentes, através do processamento de linguagem natural, onde dentre suas etapas principais está o “pré-processamento, no qual um conjunto de ações transformadoras é aplicado, para garantir que a informação em formato não estruturado (texto) se torne passível de compreensão pela máquina.” (Guimarães; Meireles; Almeida, 2019 p. 171).

Desta forma “a capacidade de pesquisar, acessar, mover, manipular e explorar esses dados será um requisito central – ou uma vantagem competitiva” no âmbito das novas tecnologias e aplicações colaborativas. (Hey, 2006, p. 519, tradução nossa)

Como a informação é o elo entre o trabalho do profissional da informação e a base das tarefas desenvolvidas em Inteligência Artificial, é fundamental que o profissional da informação desenvolva novas competências e habilidades ligadas que possibilitem a atuação em equipes multidisciplinares criando um diálogo eficiente de forma a contribuir com ambas as áreas.

2.2. Relação da CI com a IA e CD

A Ciência de dados, enquanto disciplina emergente, surge a partir de quatro pilares fundamentais, ilustrado na **Figura 2**, sendo o primeiro a área de domínio do conhecimento, derivam da esfera da Ciência, dentre elas está a Ciência da Informação. (Marchionini, 2016).

Figura 2 – Fundação da Ciência de Dados



Fonte: Adaptado de Marchionini (2016).

Marchionini (2016) afirma ainda que os Cientistas da Informação são responsáveis por gerir todo o ciclo de vida dos dados, como também estão envolvidos com questões socioculturais atreladas à coleta e uso desses dados.

Essa abordagem abrangente da Ciência da Informação contribui de maneira significativa para a compreensão e otimização da gestão de dados, considerando não apenas os aspectos técnicos, mas também os elementos éticos e sociais envolvidos nesse processo dinâmico e crucial para a tomada de decisões estratégicas. Portanto, a interseção entre Ciência de Dados e Ciência da Informação representa um campo fértil para inovações e avanços no tratamento da informação no contexto contemporâneo.

Quanto a Inteligência Artificial, uma disciplina que caminha lado a lado da Ciência da Informação (CI), que tem como principal desafio “conseguir desenvolver

sistemas computacionais que sejam capazes de entender a linguagem humana de acordo com os parâmetros de uma língua formal, sistematizada com propósitos e finalidades bem definidos quanto ao seu uso” (Fonseca; Guelpeli; Souza Netto, 2022, p. 3).

Enquanto a linguística, como “estudo sistemático da linguagem” (Vajjala, 2020, p. 8, tradução nossa) se ocupa em entender, descrever e traduzir os conceitos e estrutura da linguagem natural, a Ciência da Informação se ocupa da estruturação de Ontologias, Tesouros e Vocabulários controlados utilizados para contextualização da informação, descrição de “um domínio do conhecimento de forma estruturada, através de: classes, propriedades, relações, restrições, axiomas e instâncias” (Santarem Segundo; Coneglian, 2015, p. 227) e do tratamento da desambiguação do sentido das palavras que melhora a exatidão evitando a multiplicidade de significados associados a uma única palavra no momento da recuperação da informação (Nhacuongue, 2020, p. 17).

Dessa forma, conforme apresentado por Siqueira e Pereira (1989, p. 46) “a recuperação da informação revela-se como um item prioritário na interação entre a [Ciência da Informação] e a Inteligência Artificial”, pois é através da recuperação da informação e do processamento, pelo computador dos termos em linguagem natural que é possível solucionar problemas informacionais e apresentar soluções através da interação entre o usuário e os agentes inteligentes.

Na busca de soluções para as questões informacionais da sociedade, os estudos em Ciência da Informação têm motivado debates sobre a transformação do grande volume de dados operacionais produzidos diariamente, em informações estratégicas que deem suporte à tomada de decisões e agreguem valor às organizações (Santos Vidotti, 2009).

2.3. Inteligência Artificial

A busca por uma Inteligência Artificial não é algo novo. Em diferentes períodos da história, desde o desenvolvimento da escrita até a segunda guerra mundial, passando pela ascensão do antigo império egípcio, o homem já procurou desenvolver máquinas que simulassem uma inteligência similar a sua e desenvolvessem o seu trabalho (Nakamiti, 2009).

Contudo, o marco inicial da Inteligência Artificial, conforme a estudamos hoje, foi firmado em 1956, quando o termo foi usado oficialmente na Faculdade de

Dartmouth, nos Estados Unidos, em um evento com pesquisadores interessados em redes neurais e no estudo da inteligência (Nakamiti, 2009).

A IA é considerada um ramo da ciência/engenharia da computação que atua no objetivo de desenvolver sistemas computacionais que realizam tarefas que hoje, não têm solução algorítmica (uma sequência finita de ações que resolve um problema específico) ou que ainda sejam melhor executadas por homens do que por máquinas, ou seja, problemas que não têm soluções exatas e sim, com muitas variáveis (Sichman, 2021).

Em 1991, na primeira publicação sobre o tema, chamada de *Artificial Intelligence*, os autores Elaine Rich e Kevin Knight diziam vagamente que essa área “se ocupa do estudo de computadores que fazem coisas que, no momento, as pessoas fazem melhor.” (Cozman; Neri, 2021, p. 21). Por se tratar de um conceito complexo e abrangente, “não tem uma definição acadêmica propriamente dita”, como ressalta Sichman (2021, p.36), mas podemos entender, em síntese, que essa inteligência é um conjunto de técnicas de programação que emulam as principais funções associadas à inteligência humana, como raciocínio, aprendizagem e autoaperfeiçoamento para solucionar problemas em informática através de programas (Nakamiti, 2009).

Esse conjunto de técnicas é chamado de comportamento inteligente e a principal dificuldade está em definir o que é ser inteligente. Isso porque, segundo Cozman e Neri (2021, p. 22), os seres humanos têm muita flexibilidade com relação ao termo inteligência, por exemplo “o que hoje é considerado uma atividade inteligente pode se tornar banal assim que suas regras são codificadas de forma computacional”. Por isso, talvez seja mais produtivo falar de Inteligência Artificial no que tange às suas características e no âmbito de seus eixos temáticos do que sob a luz epistemológica de sua natureza.

Os indivíduos se comunicam e interagem por meio da linguagem natural utilizando um conjunto de símbolos e signos, tanto verbais quanto não verbais (Fonseca *et al.*, 2018). Os autores destacam que a essência da linguagem está profundamente conectada ao contexto em que ocorre, aos indivíduos que a utilizam e à forma como é empregada. Sendo composta por regras, variações e ambiguidades, seu processamento por máquinas exige uma análise criteriosa desses elementos para garantir uma interpretação eficaz.

Objetivo principal da inteligência Artificial é desenvolver soluções e resolver problemas que os humanos fazem melhor que as máquinas, e que ainda não possuem solução computacional viável. Para tanto, é necessário entender como os seres humanos solucionam esses problemas, como são geradas as possíveis soluções, como é feita a escolha da melhor solução e seus efeitos em futuras decisões. (Sichman, 2021).

Neste caminho de resolver problemas humanos ainda insolúveis por máquina, se encontra o principal desafio da IA, que está em desenvolver ferramentas “[...] que sejam capazes de entender a linguagem humana” para atingir um determinado propósito (Fonseca *et al.*, 2018).

A Inteligência Artificial pode ser caracterizada por um conjunto de tecnologias, métodos, modelos e técnicas, que quando combinados entre si proporcionam a “busca, raciocínio e representação de conhecimento, mecanismos de decisão, percepção, planejamento, processamento de linguagem natural, tratamento de incertezas e aprendizado de máquina” (Sichman, 2021, p. 39).

Segundo Russell e Norvig (2004) a maioria das tecnologias de IA podem ser classificadas em seis disciplinas:

Processamento de linguagem natural: possibilita a interação e comunicação entre humanos e máquinas usando linguagem natural, facilitando a compreensão e geração de texto por sistemas computacionais;

Aprendizado de máquina: permite que sistemas identifiquem padrões, aprendam com dados e se adaptem a novas situações sem programação explícita para cada tarefa;

Visão computacional: abrange a interpretação de informações visuais do ambiente, como reconhecimento de imagens e vídeos;

Representação do conhecimento: organiza e armazena informações de maneira estruturada, permitindo que sistemas acessem e utilizem dados adquiridos ou inferidos;

Raciocínio automatizado: aplica lógica e algoritmos para responder a questões complexas e realizar deduções automáticas com base no conhecimento armazenado;

Robótica: envolve o controle e coordenação de dispositivos físicos para tarefas como movimentação, manipulação de objetos e interação com o ambiente físico.

As ferramentas vão além de entender, aprender, reconhecer e produzir conteúdo da linguagem humana. As máquinas realizam o tratamento da linguagem semelhante ao obtido pelos humanos, possibilitando identificar significado e contexto das palavras (Zeroual; Lakhouaja, 2018; Coneglian, 2020).

Segundo Stanford University (2016) Inteligência Artificial está evoluindo para se tornar “uma força central na sociedade” (Stanford University, 2016, p. 9, tradução nossa), o foco está se direcionando para desenvolvimento de sistemas inteligentes que possam colaborar efetivamente com as pessoas, tornando conscientes das necessidades e interações humanas.

Algumas áreas aplicadas têm ganhado destaque como,

- **Aprendizado de máquina em larga escala**, preparado para atuar com grandes volumes de dados;
- **Aprendizado profundo**, tem revolucionado as áreas de percepção computacional, “com o reconhecimento de objetos e rotulagem de vídeos” (Stanford University, 2016, p. 9, tradução nossa);
- **Aprendizado por reforço**, muda o foco do reconhecimento de padrões para tomada de decisões com base nas experiências;
- **Robótica**, voltada hoje para o treinamento da interação com o mundo;
- **Visão computacional**, teve grandes avanços, pela primeira uma máquina executou tarefa de visão melhor que um humano;
- **Processamento de linguagem natural**, focado no aprimoramento do diálogo com pessoas e tradução simultânea;
- **Sistemas colaborativos**, desenvolvimento de sistemas autônomos que colaboram com outros sistemas;
- **Crowdsourcing e computação humana**: combina o que os humanos fazem bem com o que os computadores são bons para obter melhores resultados em tarefas complexas;
- **Internet das Coisas (IoT)**: interconexão de dispositivos para coleta e compartilhamento de dados par uso com funções inteligentes;
- **Computação neuromórfica**: tecnologias que imitam redes neurais para melhorar eficiência.

A evolução da Inteligência Artificial vai além de avanços tecnológicos: ela redefine como organizamos, interpretamos e utilizamos informações. Desde sua concepção, a IA busca não apenas replicar capacidades humanas, mas também

aprimorar processos que envolvem grandes volumes de dados, como classificação, aprendizado e tomada de decisões.

Essa transformação impacta significativamente áreas dedicadas à organização e gestão de informações, permitindo o desenvolvimento de soluções mais inteligentes e eficientes que integram percepção, aprendizado e colaboração, potencializando a capacidade humana em ambientes digitais.

2.4. Ciência de Dados

O termo *Data Science* (Ciência de Dados) apareceu pela primeira vez, segundo Cao (2017), no prefácio do livro “*Concise Survey of Computer Methods*” de Peter Naur publicado em 1974.

A Ciência de Dados pode ser descrita tanto como técnica quanto como Ciência. Enquanto técnica ela visa captar o que os dados representam, sem necessariamente procurar entender o fenômeno que causou este dado. Já enquanto Ciência, o foco está em entender o porquê dos dados, objetivando uma análise mais profunda, por meio da exploração dos dados que resulta em maior compreensão de seu contexto original, o que a torna “[...] uma Ciência com e sem teorias” (Fonseca, 2021, p. 1594, tradução nossa).

As abordagens científicas e técnica diferem no uso de dados e hipóteses. Na ciência convencional, hipóteses são formuladas previamente e testadas para confirmar ou refutar teorias. Já na abordagem técnica, os dados são manipulados primeiro, invertendo a ordem tradicional de geração de hipóteses, com foco em insights sem necessariamente confirmar teorias. (Fonseca, 2021)

Cao (2017) define que a Ciência de dados como estudo dos dados e complementa que é um campo interdisciplinar novo que se vale de conhecimento de áreas como estatística, informática, computação, comunicação, gestão e sociologia para estudar dados. Os autores Rautenberg; Carmo, 2019 reafirmam o caráter interdisciplinar da Ciência de Dados, e acrescenta a Ciência da Informação como uma área relacionada.

Além de interdisciplinar, é uma ciência abrangente, que afeta outras Ciências, provocando mudanças inclusive na “forma como fazemos ciência de um modo geral” (Fonseca, 2021, p. 1596, tradução nossa).

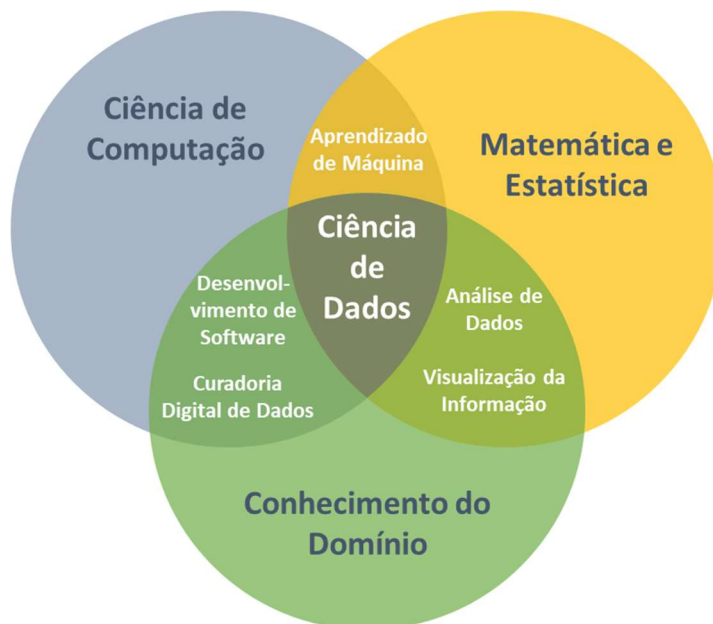
O autor apresenta críticas feitas por diversos autores que apresentam argumentos traz a reflexão se “Ciência de Dados é uma ferramenta para resolver

problemas ou se ela é uma nova forma de fazer Ciência” (Fonseca, 2021, p. 1598, tradução nossa).

Com uma área interdisciplinar, exige habilidades em Ciência da Computação para manipulação e armazenamento de dados, uso de algoritmos de Aprendizado de Máquina e visualização da informação. O domínio de Matemática e Estatística é essencial para interpretar resultados e compreender algoritmos.

Além disso, o conhecimento do domínio do problema é crucial para formular hipóteses relevantes e gerar informações úteis para a tomada de decisão. Essa convergência de áreas torna a Ciência de Dados uma ponte entre tecnologia, análise quantitativa e aplicação prática. (Rautenberg; Carmo, 2019).

Figura 3 – Interdisciplinaridade da Ciência de Dados



Fonte: Adaptada de Rautenberg e Carmo (2019).

A Ciência de Dados está intrinsecamente ligada à aplicação de métodos, onde softwares são utilizados para transformar dados em informações relevantes, através da sua extração e, bases de base de dados extensas, complexas e dinâmicas, contribuindo de forma decisiva para a tomada de decisões (Rautenberg; Carmo, 2019).

É proposta por Fonseca (2021) uma visão da Ciência de Dados em três fases, onde afirma, a partir da discussão de cada fase, que a Ciência de Dados, não está desconectada da teoria. Que esta é apresentada e utilizada em diferentes etapas de diferentes maneiras, conforme segue:

Fase 1 - Planejamento da coleta de dados: esta fase tem como objetivo planejar e estruturar a Seleção, Coleta, Curadoria dos dados que irão auxiliar na etapa

de análise. Os dados serão coletados a partir de diferentes fontes contendo ou não origens científicas;

Fase 2: Análise de dados: esta fase é o ponto principal da Análise de Dados, são reunidas ferramentas e técnicas necessárias para extrair dos dados insights significativos que irão embasar o entendimento do contexto;

Fase 3: interpretação dos resultados: nesta fase os resultados são analisados à luz das teorias das Ciências que originaram os dados.

Fonseca (2021) afirma que a Ciência de Dados recebe o auxílio das chamadas Ciências Auxiliares. Áreas como Ciência da Informação, Estatística e Ciência da Computação dão suporte nas diferentes fases de desenvolvimento da Ciência de Dados. A Ciência da Informação tem papel importante nas etapas de exploração, identificação, armazenamento e preservação dos dados, tendo domínio de todo seu ciclo de vida. As áreas de Estatística e Ciência da Computação, auxiliam nas etapas de processamento, tratamento e análise de resultados.

Com isso, a CD não está limitada à análise estatística ou ao processamento de dados, mas compreende diversas práticas que passam pela engenharia de dados e vão até a visualização e comunicação dos resultados. Seu foco está na geração de insights a partir do processamento e interpretação dos dados (Moutinho, 204, p. 6).

Independente da técnica ou método utilizado, a CD tem um Ciclo de Vida que envolve sete etapas principais: **obtenção de dados**, avaliando e selecionando fontes primárias; **ingestão de dados**, centralizando dados em formatos uniformes; **exploração de dados**, gerando hipóteses iniciais; **definição de parâmetros**, ajustando os algoritmos de aprendizado de máquina; **implementação do modelo**, desenvolvendo modelos por treinamento e testes iterativos; **utilização do modelo**, aplicando-o a cenários reais para verificar sua eficácia; e **tomada de decisão**, combinando os resultados do modelo com insights visuais para decisões baseadas em conhecimento intensivo. (Rautenberg; Carmo, 2019).

Portanto, a Ciência de Dados emerge como uma abordagem multidisciplinar que integra conhecimento de diversas áreas para transformar dados em decisões estratégicas. Seu ciclo de vida estruturado, desde a obtenção até a tomada de decisões, reflete uma combinação complexa de técnicas, ferramentas e abordagens analíticas. A interação com as Ciências Auxiliares, como a Ciência da Informação, Estatística e Ciência da Computação, fortalece sua aplicação prática, oferecendo uma visão holística e integrada para entender e explorar dados no cenário contemporâneo.

3. TRABALHOS RELACIONADOS

Neste capítulo, são apresentadas pesquisas relacionadas ao uso de Inteligência Artificial e Ciência de Dados na Gestão Eletrônica de Documentos. Embora não tenha sido identificado nenhum estudo com foco direto na aplicação de IA em GED, trabalhos relevantes foram encontrados, analisando o uso de IA em áreas correlatas, como Gestão de Documentos e Ciência da Informação.

No artigo apresentado por Martins (2010, p.1), traz uma discussão, através de uma revisão de literatura, sobre técnicas de Inteligência Artificial. O autor “apresenta vários possíveis usos da Inteligência Artificial pela Ciência da Informação na resolução de problemas atualmente difíceis de serem tratados tanto pelo processamento manual quanto automatizado”.

O autor apresenta diversos trabalhos que exploram esse tema. Há um destaque para sistemas de recuperação da informação, que utilizam redes neurais e algoritmos evolutivos para refinar resultados e reduzir a lacuna entre as buscas do usuário e os documentos recuperados. Agentes inteligentes são empregados para localizar informações autonomamente e adaptar-se ao ambiente com base em feedback do usuário. Outras aplicações incluem interfaces inteligentes, que personalizam experiências; classificação automática de conteúdos usando ontologias; processamento de linguagem natural para interfaces mais intuitivas; e mineração de dados, que identifica padrões em grandes volumes de informações.

Ao final, o autor conclui que o “uso da Inteligência Artificial fora da área de exatas enfrenta grande dificuldade de receber adeptos”, e acrescenta que tecnologias futuras podem ajudar a organizar a crescente quantidade de informações, especialmente diante do cenário de "caos informacional" que se aproxima.

Silva e Nathansohn (2018, p. 111) realizou uma “revisão bibliográfica, aplicada e descritiva, com análises bibliométricas e cientométricas da produção científica”, utilizando como fonte de pesquisa a base de dados Brapci. Este trabalho teve como objetivo “é analisar as discussões tecidas pela interdisciplinaridade entre as duas áreas de conhecimento, sob a ótica da IA” (Silva; Nathansohn, 2018, p. 112).

O estudo analisou 43 publicações científicas sobre Inteligência Artificial na Ciência da Informação, resultando em 33 artigos após filtragem. A produção científica revelou baixa frequência, com destaque para 2014, com cinco publicações.

Predominaram artigos individuais e poucos autores recorrentes. Os temas centrais incluíram tecnologias emergentes, aplicabilidade da IA, interdisciplinaridade entre CI e Ciência Cognitiva, e impacto organizacional.

A maioria dos artigos relacionam-se a discussões e a respeito do uso dessas tecnologias, teorias e relações interdisciplinares e implicações na educação. Apenas quatro artigos apresentaram pesquisas aplicadas, mas suas contribuições não são discutidas no artigo. Por fim, o estudo ressalta a necessidade de maior “aprofundamento teórico e aplicado no campo da IA, inclusive, maior produção e comunicação científicas” (Silva; Nathansohn, 2018, p. 125).

No estudo apresentado por Dorneles (2022, p. 1), caracteriza-se como uma Revisão Sistemática de Literatura (RSL), com o objetivo de “[...] identificar aplicações de Inteligência Artificial (IA) na gestão de documentos”. Com uma abordagem exploratória descritiva, focando em uma análise qualitativa-quantitativa, com recorte temporal entre os anos de 2016 e 2021, utiliza-se das bases de dados LISTA, EMERALD e *Scopus* para conduzir suas pesquisas.

O estudo destaca a predominância da área de saúde em pesquisas sobre o uso de IA em sistemas de registros eletrônicos. Dos 79 artigos analisados, 66 tratam dessa área, enfatizando o uso de tecnologias como *machine learning*, processamento de linguagem natural e *deep learning*.

Apenas três estudos abordaram questões ligadas à gestão de documentos, um trata do uso de Processamento de Linguagem Natural (PNL) e Aprendizado de Máquina no contexto do fluxo documental, ou outro apresenta uma ferramenta que poia decisões estratégicas sobre retenção e descarte de arquivos, e o último faz uma “discussão sobre as lições aprendidas, problemas e implicações da IA no arquivo” (Dorneles, 2022, p.16).

Por fim, o autor conclui que com as informações apresentadas seu trabalho foi possível ter uma visão dos estudos realizados, percebe um aumento expressivo estudos nesta área, com visualização de estudos relevantes. Ele considera que foi possível demonstrar “as potencialidades das tecnologias da IA para a potencialização no tratamento de dados e informações” (Dorneles, 2022, p.18).

Alves e Ribeiro (2024, p. 2) realizou uma revisão bibliográfica, que “tem caráter exploratório descritivo, com abordagem mista”, de publicações das bases de dados

Brapci, Google Acadêmico e Periódico Capes, entre os anos 2000 e 2022. Este trabalho teve como objetivo uma análise estatística das publicações que abordam aplicações da Inteligência Artificial no contexto na Ciência da Informação.

O estudo analisou 51 artigos sobre IA no contexto da Ciência da Informação entre 2000 e 2022. O maior número de resultados foi obtido no Google Acadêmico (28), seguido pelo Portal de Periódicos da CAPES (12) e BRAPCI (11). Houve aumento significativo nas publicações brasileiras nos últimos quatro anos. A revista Ciência da Informação (IBICT) concentrou o maior número de artigos. Os dados revelaram tendências e resultados importantes por meio das estratégias de busca nas bases analisadas para atender aos objetivos da pesquisa.

Com esta pesquisa, foi possível observar que houve um aumento das pesquisas sobre IA na CI nos últimos quatro anos, que passou de oito em 2019 para 20 em 2022, no entanto este volume era ainda considerado baixo, sendo considerada pela autora que o uso dessas tecnologias “exclusivo na CI ainda não é uma realidade em sua maioria” (Alves; Ribeiro, 2024, p. 13).

O estudo desenvolvido por Magalhães *et al.* (2024, p.1) apresentou os resultados iniciais de uma Revisão Sistemática de Literatura, com objetivo de “investigar como as tecnologias oriundas da Inteligência Artificial e da Ciência de Dados podem ser utilizadas na Gestão Eletrônica de Documentos”. Considerando o primeiro marco da presente dissertação de mestrado.

O artigo utilizou uma abordagem ampla para mapear as três áreas temáticas e, a partir da leitura do título, palavra-chave e resumo, selecionou 58 estudos, destes, 43% abordavam a área da saúde, evidenciando o uso de IA na automação e previsão de diagnósticos e classificações de doenças.

Estes resultados preliminares destacaram a necessidade de melhoria no protocolo de revisão sistemática, em especial na *string* de busca, levando a um refinamento do método e à continuação da pesquisa, que culminou na atual dissertação.

4. METODOLOGIA DE PESQUISA

Esta seção apresenta as considerações metodológicas que fundamentam a condução desta pesquisa, detalhando as abordagens, técnicas e procedimentos adotados para alcançar os objetivos propostos.

A metodologia da pesquisa é baseada no planejamento e condução de uma Revisão Sistemática da Literatura (RSL), que é uma forma de realizar a identificação, avaliação e interpretação de estudos realizados sobre um determinado problema, assunto ou área específica.

Uma revisão sistemática de literatura pode oferecer contribuições valiosas tanto como processo quanto como produto da pesquisa:

Como processo: se destaca por sua ênfase em um planejamento detalhado e explícito das estratégias e etapas a serem seguidas, demanda a formulação clara de questões de pesquisa, a definição de critérios de inclusão e exclusão de estudos, a seleção criteriosa das fontes de informação e a análise sistemática dos resultados obtidos (Kitchenham; Charters, 2007).

Como produto: fornece um panorama abrangente e organizado do estado da arte, identificando padrões, lacunas de estudos em uma determinada área, tendências e empíricas para discutir ou desenvolver hipóteses teóricas (Kitchenham; Charters, 2007). Isso apoia pesquisadores e profissionais ao consolidar conhecimentos existentes e indicar direções futuras para pesquisa e aplicação prática.

Assim, a pesquisa tem caráter exploratório-descritivo e, para a análise e síntese dos estudos, adota uma abordagem combinando aspectos qualitativos e quantitativos. O foco de análise são artigos científicos, que segundo Gil (2008) são importantes fontes utilizadas para obter informações e embasar um estudo científico.

Esta pesquisa busca identificar informações que irão contribuir para o melhor entendimento da relação entre as áreas de IA e GED, por isso é considerada pesquisa básica. Que, segundo Minayo (2010) tem como direcionamento o avanço na área para ampliar o conhecimento teórico sobre determinado tema através da elaboração de teorias, não tem objetivos práticos, embora podem servir como subsídio a ela.

Para desenvolver esta pesquisa optou-se pela abordagem mista ou qualitativa/quantitativa ou quali/quant, neste tipo de pesquisa ocorre a coleta e análise de dados tanto qualitativos quanto quantitativos.

Richardson; *et al.* (2007, p. 80) afirmam que pesquisas qualitativas “têm como objetivo situações complexas ou estritamente particulares”, já pesquisas quantitativas “podem descrever a complexidade de determinado problema, analisar a interação de certas variáveis, compreender e classificar processos dinâmicos”

O uso da pesquisa quantitativa pode ser selecionado como uma forma de evitar distorções na interpretação e análise de dados, pois ela tem “a intenção de garantir a precisão dos resultados” (Richardson; *et al.*, 2007, p. 70) visto que os dados tabulados e sistematizados podem complementar e corroborar com os resultados da pesquisa qualitativa.

Há uma diversidade de métodos e técnicas de pesquisa, que o pesquisador pode utilizar, em vários casos, inclusive, é possível lançar mão da combinação de diferentes técnicas de forma a amplificar o potencial da pesquisa, garantir legitimidade ao processo e a qualidade do resultado (Braga, 2007).

Para a etapa de análise de dados, a técnica de análise temática e de conteúdo foi escolhida enquanto técnica qualitativa, visto que auxilia no processo de categorização, classificação e identificação das informações importantes para o resultado deste trabalho, ao passo que permite quantificar e comparar sistematicamente seu conteúdo.

O uso da análise temática e de conteúdo, com caráter científico, pode ser percebida desde o início do século, sendo desenvolvida ao longo das primeiras décadas por pesquisadores norte-americanos (Richardson; *et al.* (2007). Segundo Bardin (1977) é um instrumento oriundo das Ciências Sociais que partiu dos estudos e codificação dos símbolos de mensagens em diferentes formatos ou suportes.

A análise temática que consiste na classificação de elementos com mesma “significação constitutiva”, por meio da adoção de certos critérios para estruturação de índices, indicadores e taxonomia (Bardin, 1977). Diversos critérios que podem ser utilizados para elaboração desta análise, dentre eles estão: função, finalidade, estrutura tipo ou modal, associação, equivalência, exclusão.

Para Minayo (2010, p. 304), a análise de conteúdo possui o caráter quantitativo quando usado como técnica de tratamento de dados e caráter qualitativo quando “busca a interpretação cifrada do material”.

A análise de conteúdo fornece um conjunto diverso de técnicas de análise e descrição de comunicações, de forma sistemática e objetiva, para produção de indicadores que permitirão a realização de inferências para geração de novos conhecimentos (Bardin, 1977).

Já as técnicas de bibliometria foram escolhidas como técnicas quantitativas. A bibliometria irá auxiliar na quantificação dos elementos envolvidos na pesquisa, como indicadores de artigos, autores, categorias etc., enquanto a análise de sistemas complexos (ou análise de redes) irá auxiliar na análise da relação entre os elementos e com representações que auxiliam a obter insights para identificação de novos caminhos para a área pesquisada.

Na próxima seção são apresentadas as etapas para desenvolvimento da pesquisa, as técnicas escolhidas para coleta e análise de dados, suas definições e aplicações.

5. DESENVOLVIMENTO DA PESQUISA

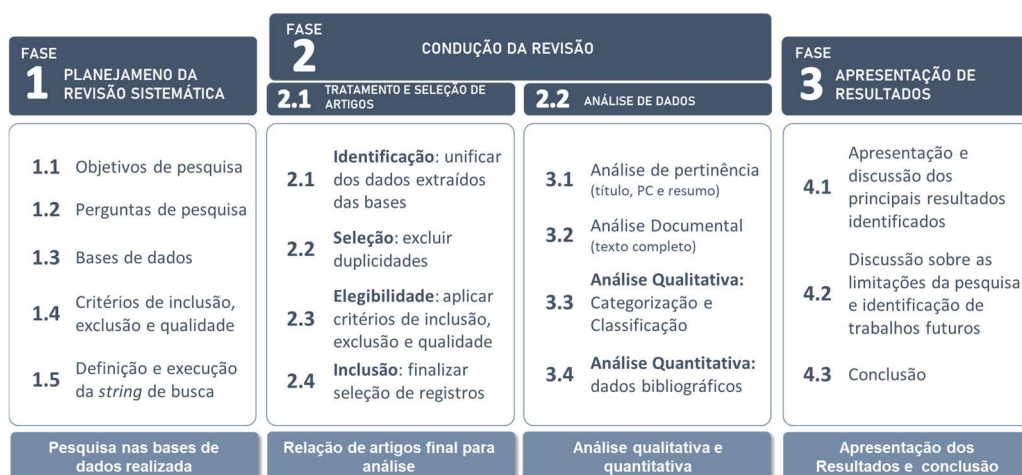
A metodologia de pesquisa para este trabalho é Revisão Sistemática de Literatura (RSL). Que consiste em uma estratégia rigorosa e replicável para identificar, avaliar e interpretar estudos disponíveis e relevantes relacionados a um tópico de pesquisa. Todos os passos e métodos são controlados por meio de um protocolo de revisão, permitindo, portanto, avaliar a completude do conjunto de estudos e a confiabilidade de todo o processo (NHMRC, 2000).

Segundo Kitchenham (2004), uma revisão sistemática é organizada em três fases principais: Fase 1) Planejamento da Revisão; Fase 2) Condução da Revisão, e Fase 3) Publicação dos Resultados. Para melhor adequar o conteúdo deste trabalho a Fase 2 foi subdividida em duas etapas (Tratamento e seleção de artigos e análise de dados), conforme apresentado na **Figura 4**.

Para Kitchenham (2007) a RSL deve ser estruturada em três etapas: Planejamento, que considera a identificação da necessidade, definição do problema de pesquisa e elaboração do protocolo; Realização, execução da pesquisa, seleção dos estudos, aplicação dos critérios de inclusão, exclusão e qualidade, extração e síntese de dados; e, por fim Relato, contempla especificação dos meios de divulgação e preparação do relatório final.

Apesar da estruturação em etapas, não são necessariamente lineares, há muita iteração entre as atividades durante sua realização. Durante a aplicação do resultado de pesquisa é analisado e pode implicar em alterações nas etapas já concluídas (Kitchenham; Charters, 2007).

Figura 4 – Etapas de desenvolvimento da pesquisa



Fonte: Elaborado pela autora (2024)

A seguir são apresentadas cada uma das etapas e subetapas de desenvolvimento da pesquisa.

5.1. Protocolo de Revisão Sistemática de Literatura (Etapa 1.1)

O desenvolvimento da RSL desta pesquisa foi realizado seguindo as etapas descritas por Kitchenham (2007), sendo apresentadas abaixo na sequência: Planejamento, por meio do preenchimento do protocolo da RSL com as informações sobre a pesquisa e estruturação das etapas de execução.

Nesta etapa são apresentados o objetivo da pesquisa, as questões de pesquisa, bases de dados utilizadas, os critérios de seleção, os critérios de exclusão e os critérios de qualidade. Como resultado desta etapa tem-se a estruturação e execução da *string* de busca, determinante para o avanço da pesquisa.

5.1.1. Objetivos de pesquisa

Para o desenvolvimento e preenchimento correto do protocolo de RSL, o objetivo da pesquisa é o primeiro ponto a ser apresentado. Os objetivos descritos no protocolo são exatamente os objetivos definidos para esta pesquisa, conforme apresentado no **Quadro 1**, a seguir:

Quadro 1 – Objetivo de pesquisa

Objetivo de Pesquisa	
Item	Descrição
Objetivo Geral	Investigar a integração de tecnologias, como a Inteligência Artificial e a Ciência de Dados na Gestão Eletrônica de Documentos, com o intuito de identificar padrões, desafios e oportunidades que possam contribuir para a evolução das práticas de gestão documental.
Objetivo Específico 1	Levantar as pesquisas existentes sobre o uso de técnicas de Inteligência Artificial no processo de Gestão Eletrônica de Documentos;
Objetivo Específico 2	Mapear e classificar essas técnicas de Inteligência Artificial utilizadas na Gestão Eletrônica de Documentos a partir da análise de conteúdo;
Objetivo Específico 3	Analisar como essas técnicas podem melhorar a eficiência da Gestão Eletrônica de Documentos no acesso, recuperação, tratamento, análise, uso e segurança de informações.

Fonte: Elaborado pela autora (2024)

5.1.2. Questões de pesquisa

Com os objetivos em mãos foi possível elaborar as perguntas de pesquisa, que irá guiar o processo investigativo. Abaixo são apresentadas as perguntas de pesquisa, alinhadas com o problema de pesquisa deste trabalho, bem como com os objetivos geral e específicos já mencionados, conforme **Quadro 2**.

Quadro 2 – Questões de pesquisa

Questão de pesquisa	
Item	Descrição
P1	Quais são os principais pontos de interseção entre Gestão Eletrônica de Documentos, Inteligência Artificial e a Ciência de Dados?
P2	Quais desafios, tendências e lacunas de pesquisa estão presentes na interseção dessas áreas?
P3	Como as aplicações de Inteligência Artificial e Ciência de Dados têm influenciado as práticas na Gestão Eletrônica de Documentos?

Fonte: Elaborado pela autora (2024)

Cada questão de pesquisa é analisada e deve ser respondida de acordo com diferentes pontos de vista. Para tanto é apresentada por Kitchenham (2007), a metodologia PICOC (*Population, Intervention, Comparison, Outcomes and Context*), traduzido aqui como: População, Intervenção, Comparação, Resultado e Contexto (Kitchenham; Charters, 2007, p. 11, tradução da autora).

Para essa classificação e o mapeamento da pesquisa devem ser respondidas tendo-se em mente todo o ambiente que envolve a pesquisa: População, que pode ser uma função específica, área de pesquisa ou grupo de empresas ou pessoas; Intervenção, é a metodologia, ferramenta, tecnologia ou procedimento utilizada para abordar um problema específico; Comparação, é a metodologia, ferramenta, tecnologia ou procedimento a ser compara com a intervenção; Resultado, refere-se às melhorias esperadas a partir da intervenção realizada; e, por fim, Contexto, definição do contexto em que a comparação ocorre, os participantes do estudo e as tarefas a serem executadas (Kitchenham; Charters, 2007).

Abaixo são apresentados os critérios segundo a metodologia PICOC:

Quadro 3 – Descrição dos elementos PICOC da Pesquisa

PICOC (<i>Population, Intervention, Comparison, Outcomes and Context</i>)	
Critérios	Descrição
População	Estudos que discutem sobre o uso de técnicas de inteligência artificial na gestão eletrônica de documentos enquanto metodologia, processo, ferramenta, técnica ou prática.
Intervenção	Análise e comparação das técnicas de inteligência artificial passíveis de serem aplicadas ao processo de gestão eletrônica de documentos.
Comparação	Uso tradicional de técnicas de gestão eletrônica de documentos.
Resultado	Identificação de técnicas de inteligência artificial que efetivamente aumentem a eficiência na recuperação de informações estratégicas para tomada de decisões dentro das organizações.
Contexto	Estudos desenvolvidos em ambiente acadêmico, pela pesquisadora responsável e em pequena escala.

Fonte: Elaborado pela autora (2024)

5.1.3. Definição de Bases de Dados

Para uma pesquisa bibliográfica, a escolha da base de dados é de suma importância, por esse motivo, optou-se por fazer uma pesquisa ampla contemplando as principais base de dados científicas.

Foram selecionadas as seguintes bases de dados para pesquisa de artigos científicos:

- **Brapci:** base de dados referenciais de artigos de periódicos em Ciência da Informação, reúne e indexa artigos de periódicos da área de Ciência da Informação, que contempla Biblioteconomia e Arquivologia. “Disponibiliza referências e resumos de 19.255 textos publicados em 57 periódicos nacionais impressos e eletrônicos da área de CI. Dos periódicos disponíveis, 40 estão ativos e 17 históricos (descontinuados)”, (Bufrem et al., 2010).
- **Lista:** *Library, Information Science & Technology Abstracts*, base de dados referenciais de artigos de periódicos internacionais em Ciência da Informação,

indexa mais de 500 periódicos científicos, além de livros e relatórios de pesquisas.

- **Scopus:** “Scopus é um banco de dados de resumos e citações de literatura revisada por pares, incluindo revistas científicas, livros e anais de conferências”, além do serviço de pesquisa oferece também estatística do resultado da busca (ELSEVIER, 2015; 2022, tradução nossa).
- **Web Of Science:** disponibiliza material bibliográfico (livros, artigos científicos, anais de congresso, patentes etc.), monitoramento de resultados, estatísticas de resultados, dentre outros. Possui mais de 196 milhões de registros, mais de 34 mil revistas revisadas por pares, 254 categorias de assunto, mais de 134 mil livros e 109 mil patentes. (Clarivate, 2023).

A escolha tanto da base de dados Web of Science quanto da Scopus se deu porque, percebeu-se, a partir de uma pesquisa exploratória, que mesmo existindo periódicos indexados em ambas as bases, existem também periódicos indexados em uma e que não são indexados na outra. E encontrou-se artigos relevantes em ambas as bases que seriam perdidos caso não se realizasse o cruzamento de informações.

A escolha da base de dados Brapci e Lista foi feita por conter um vasto arcabouço de artigos publicados pela área da Ciência da Informação, sendo a Brapci nacional e Lista internacional, o que torna fundamental o conhecimento de sua produção, visto que o trabalho é desenvolvido nesta área.

5.1.4. Definição de critérios de inclusão, exclusão e qualidade

Outro elemento fundamental para o planejamento da revisão sistemática é definir os Critérios de Inclusão (CI) e os Critérios de Exclusão (CE). Esses critérios permitem selecionar estudos relevantes para responder às questões de pesquisa e descartar os estudos fora do escopo. Observa-se 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 também satisfeito. Por esse motivo a ordem de aplicação inicia-se pelos critérios de exclusão, apresentados no **Quadro 4**.

Quadro 4 – Descrição dos Critérios de Exclusão

Critérios de Exclusão	
Item	Descrição
CE1	O estudo está escrito em um idioma diferente de português, inglês e espanhol.
CE2	O texto completo não possui acesso aberto ou não pode ser acessado por outro motivo.
CE3	O estudo é livro, capítulo de livro, nota, carta, <i>proceedings</i> (anais), <i>retracted</i> (retração).
CE4	O estudo aborda apenas o tema GED sem intersecção com IA.
CE5	O estudo apenas cita GED como ferramenta usada na pesquisa, como resultado de pesquisa ou sugestão de uso.
CE6	Artigos anteriores a 2000 (filtro de 2000 a 2024).

Fonte: Elaborado pela autora (2024)

Foram definidos 6 Critérios de Exclusão. O primeiro (CE1) irá desconsiderar artigos, cujo conteúdo não esteja em português, inglês ou espanhol, essa decisão foi tomada para assegurar a inclusão de literatura relevante em diferentes línguas, ampliando a perspectiva da pesquisa. O segundo critério (CE2) desconsiderará os artigos que não apresentam texto integral de acesso aberto ou que não podem ser acessados por outro motivo. O terceiro (CE3) irá eliminar os tipos de documentos como livro, capítulo de livro, editorial, nota, carta, *proceedings* (anais), *retracted* (retração), pois o foco do trabalho é analisar as pesquisas científicas realizadas nas últimas décadas.

O quarto critério (CE4) irá desconsiderar artigos que abordam apenas o tema GED sem relação com IA, o quinto critério (CE5) irá desconsiderar estudos que apenas citem o termo de Gestão Eletrônica de Documentos (GED) como ferramenta usada para desenvolver a pesquisa, como resultado da análise apresentada na própria pesquisa ou como sugestão de uso feita pelos pesquisadores.

Por fim, o sexto e último critério (CE6) refere-se ao ano de publicação das pesquisas, que não deverá ser anterior ao ano de 2000.

Os critérios de inclusão foram definidos tendo em vista os objetivos e questões de pesquisa, conforme apresentado no **Quadro 5**.

Quadro 5 – Descrição dos Critérios de Inclusão

Critérios de Inclusão	
Item	Descrição
CI1	O estudo discute ou propõe técnica, estratégia, ferramenta ou qualquer abordagem de IA ligada à GED.
CI2	O estudo não propõe uma abordagem específica, mas discute possibilidades, tendências, benefícios e implicações na intersecção entre GED e IA.

Fonte: Elaborado pela autora (2024)

Quanto aos Critérios de Inclusão, foram definidos 2 critérios. O primeiro (CI1) irá considerar se os artigos recuperados apresentam técnica, estratégia, ferramenta ou qualquer abordagem de IA ligada à GED. O segundo critério (CI2) irá considerar se os artigos recuperados não propõem uma abordagem específica, mas discute possibilidades, tendências, benefícios e implicações na intersecção entre GED e IA.

Com base na definição dos critérios de inclusão e exclusão, foram especificados os procedimentos para seleção dos estudos e avaliação da qualidade do conjunto final. Conforme sugerido por Kitchenham (2004), o processo de seleção e análise dos artigos foi estruturado em duas etapas. Na primeira etapa, é feita a análise do título, palavras-chave, resumo e, para alguns casos, a introdução e conclusões, após essa primeira análise são aplicados os critérios. Na segunda etapa, serão lidas a versão completa dos artigos, e novamente os critérios de inclusão e exclusão serão aplicados.

Para a avaliação da qualidade dos estudos, foi especificado um conjunto de critérios e adotado o instrumento de avaliação utilizado por Dyba, Dingsoyr e Hanssen (2007). Esse instrumento é baseado em uma escala de três valores numéricos: 1 (um) - o critério é completamente atendido; 0,5 (meio) - o critério é parcialmente atendido; e 0 (zero) - o critério não é atendido.

A pontuação final é dada pela soma das pontuações de cada critério, que poderá variar de 0 (zero) a 5 (cinco). Trabalhos que receberem notas inferiores a 2,5 (dois e meio), ou seja, que atende a menos da metade dos critérios, serão desconsiderados para o *corpus* final. No **Quadro 6** são apresentados os critérios de qualidade.

Quadro 6 – Descrição dos Critérios de Qualidade

Critérios de Qualidade	
Item	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 e credibilidades do estudo são discutidas.
CQ4	São apresentadas perspectivas de trabalhos futuros com base nas contribuições.
CQ5	O trabalho apresenta um referencial teórico sólido e alinhado aos objetivos da pesquisa

Fonte: Elaborado pela autora (2024)

Os Critérios de Qualidade irão identificar se os artigos selecionados possuem as informações necessárias para realização do mapeamento e análise das técnicas, bem como se o nível de detalhamento é suficiente para atingir o objetivo desta pesquisa. Esses critérios são fundamentais para assegurar que apenas trabalhos de alta relevância e qualidade sejam considerados.

O primeiro critério (CQ1), identifica se existe uma razão fundamental pela qual o estudo foi conduzido. O segundo critério (CQ2) irá averiguar se as contribuições do estudo estão claras e de acordo com o objetivo apresentado. O terceiro critério (CQ3) identifica a credibilidade e limitações da técnica que permitirá inferir condições para implantação das técnicas e análise de sua eficiência, o quarto critério (CQ4), identifica se são apresentadas perspectivas de trabalhos futuros com base nas contribuições, o quinto e último critério (CQ5), verifica se o trabalho apresenta um referencial teórico sólido que irá auxiliar no entendimento dos conceitos adotados e caminho percorrido.

Com isso conclui-se a estruturação do protocolo de revisão sistemática de literatura. Na próxima etapa serão definidos os termos para elaboração da *string* de busca.

5.2. Definição e execução da *String* de busca

A metodologia adotada para a elaboração da *string* de busca foi fundamental para assegurar a abrangência e precisão na identificação de termos mais relevantes. O rigor metodológico empregado ao longo de cada etapa do processo confere

confiança às escolhas realizadas, garantindo que a busca incluía termos relevante para a pesquisa.

A *string* de busca foi estruturada a partir de dois conjuntos principais: o primeiro relacionado à área de Gestão Eletrônica de Documentos e suas derivações, e o segundo, às áreas de Inteligência Artificial e Ciência de Dados. A identificação e definição dos termos foram realizadas com base em uma análise dos dados e informações disponíveis em cada um desses conjuntos. Essa abordagem permitiu a criação de uma *string* que reflete as interseções entre essas áreas, assegurando a inclusão de publicações relevantes para o escopo da pesquisa.

As atividades foram estruturadas em três passos essenciais. O Primeiro Passo, denominado **Identificação**, consiste na identificação de termos diretamente da literatura, seguida pela extração de informações pertinentes, que são então organizadas em um banco de dados específico para termos.

No Segundo Passo, **Prototipagem**, cada termo identificado passa por testes de busca nas principais bases de dados, com o objetivo de explorar suas variações e o volume de registros associados. Nessa etapa, os termos são avaliados tanto de forma isolada quanto em combinações progressivas, visando identificar a configuração mais eficiente para a construção da *string* de busca. O Terceiro e último Passo, **Seleção**, é dedicado à escolha dos termos que demonstraram relevância ao tema da pesquisa e consistência nos resultados das bases de dados quando aplicados em conjunto. A **Figura 5** ilustra essa sequência.

Figura 5 – Processo de definição de *string* de busca



Fonte: Elaborado pela autora (2024)

5.2.1. Definição dos termos da área de GED

A seguir será apresentada todas as atividades desenvolvidas para definição dos termos de GD para serem utilizados na *string* de busca.

5.2.1.1. Atividade 1: Identificação (mapeamento dos termos existentes)

Esta é uma etapa com característica exploratória, consiste em identificar os termos utilizados na literatura relacionados área de Gestão Eletrônica de Documentos. Para tanto foram realizadas pesquisas nas bases de dados *Web of Science*, *Scopus*, *Brapi* e em livros especializados que possibilitou conhecer os termos utilizados por especialistas e pesquisadores, bem como suas variações.

Com a leitura dos textos de referência foi possível extrair os termos e suas definições, os quais foram documentados e utilizados para ampliar o entendimento da área e embasamento teórico desta dissertação, conforme apresentado no **Quadro 7**.

Quadro 7 – Lista de termos e definições

MAPEAMENTO DOS TEMOS DE GED		
Palavra-chave	Definição	Referência
<i>Document management</i>	A gestão de documentos é usada para rastrear e gerenciar o processo de criação de documentos, desde a concepção até a conclusão, garantindo transparência e responsabilidade através da automatização de tarefas-chave.	Wilkins (2019)
<i>Electronic Document Management System</i>	Software que utiliza técnicas automatizadas distribuição, utilização, armazenamento, recuperação, proteção e preservação de informações armazenadas em formatos digitais, acordo com políticas e procedimentos estabelecidos.	Pearce-Moses (2005)
	É um software que gerencia a criação, armazenamento e controle de documentos semi-estruturados. Consiste em várias tecnologias, incluindo, mas não se limitando a, gestão de documentos, COLD (Computer Output to Laser Disk), imagem e fluxo de trabalho. Os sistemas EDMS gerenciam um item de conteúdo como uma unidade individual, em vez de preservar sua relação com um grupo maior de documentos que fornecem evidências da mesma função organizacional.	NARA (2000)
<i>Electronic Record</i>	"Dados ou informações que foram capturados e fixados para armazenamento e manipulação em um sistema automatizado e que requerem o uso do sistema para serem compreensíveis por uma pessoa" (tradução nossa).	Pearce-Moses (2005)
<i>Electronic Recordkeeping System</i>	"Aplicação de computador que pode coletar, organizar e categorizar registros para facilitar sua gestão, preservação, recuperação, uso e disposição." (tradução nossa).	Pearce-Moses (2005)
	É um sistema de informação eletrônico (ERKS) onde os registros são coletados, organizados e categorizados para facilitar sua preservação, recuperação, uso e disposição e garante que os registros que mantêm possuam autenticidade e confiabilidade suficientes para atender a todas as necessidades de manutenção de registros.	NARA (2000)
<i>Electronic Records Management</i>	A gestão de registros eletrônicos (ERM) utiliza técnicas automatizadas para gerenciar registros em diversos formatos, sejam eletrônicos, em papel, microformas, etc.	NARA (2000)

MAPEAMENTO DOS TEMOS DE GED		
Palavra-chve	Definição	Referência
Gerenciamento Eletrônico de Documentos	"Trata-se de um sistema de gestão da informação voltado ao registro e recuperação digital dos documentos vitais ao funcionamento das entidades mantenedoras. O GED é um sistema de gerenciamento capaz de acelerar os processos de recuperação da informação, reduzir custos e impressos, e manter a integridade dos documentos"	Véras Neto, Silva e Dias (2010)
Gestão de Documentos	"Considera-se gestão de documentos o conjunto de procedimentos e operações técnicas à sua produção, tramitação, uso, avaliação e arquivamento em fase corrente e intermediária, visando a sua eliminação ou recolhimento para guarda permanente."	Brasil (1991)
	"Conjunto de procedimentos e operações técnicas referentes à produção, tramitação, uso, avaliação e arquivamento [...] de documentos em fase corrente e intermediária, visando sua eliminação ou recolhimento [...]. Também chamado administração de documentos."	Arquivo Nacional (2005)
Gestão Documental	A gestão documental, por meio da classificação e da tabela de temporalidade, permite controlar e recuperar rapidamente informações, contribui para a racionalização da produção e dos fluxos documentais, e assegura a correta avaliação e destinação dos documentos, aumentando a eficácia dos serviços arquivísticos na instituição.	Akaichi e Merlo (2018)
	"[...] a gestão documental é um fazer indispensável, que auxiliará no funcionamento de um arquivo, por suas atividades de controle e tratamento arquivístico."	Santos, Silva e Bari (2021)
	"[...] funções basilares da gestão documental: classificação, tramitação, avaliação de documentos, descrição e arquivamento."	Dorneles (2022)
Gestão Documental e Gestão de Documentos	Para os autores a gestão de documentos envolve uma série de tarefas, incluindo a análise da tipologia documental, a normalização e informatização de processos, a coordenação entre organismos, a regulamentação de transferências, a classificação e ordenação de documentos, o estudo de descartes, a manutenção da informação atualizada, a preparação de informações, e a elaboração de índices e tesouros.	Calderon <i>et al.</i> (2004)
	A gestão documental "é responsável pelo o ciclo de vida dos documentos", desenvolve atividades de análises e classificações, garantindo acesso contínuo a informações relevantes e fornecendo subsídios confiáveis e seguros para a formulação de estratégias organizacionais.	Freitas e Albuquerque (2016)
Gestão documental e Records Management	A Gestão de Documentos refere-se a "técnicas e metodologias" para gerenciar o ciclo de vida documental, focado na sistematização e planejamento das atividades documentais e centrada na informação para sua disseminação e uso.	Anna e Silva (2015)
<i>Recordkeeping</i>	Trata de "criar e manter evidências completas, precisas e confiáveis das transações comerciais na forma de informações registradas." (tradução nossa).	Borglund e Anderson (2010)
<i>Recordkeeping System</i>	"Políticas e procedimentos coordenados que permitem que registros sejam coletados, organizados e categorizados para facilitar sua gestão, incluindo preservação, recuperação, uso e disposição." (tradução nossa).	Pearce-Moses (2005)
<i>Records Management Application</i>	Aplicação de Gestão de Registros (RMA) tem como principais funções categorizar e localizar registros e identificar registros que estão prontos para disposição. O software RMA também armazena, recupera e dispõe dos registros eletrônicos mantidos em seu repositório, são capazes de gerenciar registros independentemente do seu meio.	NARA (2000)
<i>Rrecords Management</i>	A gestão de registros envolve a administração formal de objetos de informação que servem como evidência de transações ou decisões, garantindo que esses registros sejam completos e imutáveis após sua declaração.	Wilkins (2019)

Fonte: Elaborado pela autora (2024)

Com isso a busca por esses termos na literatura tornou-se ainda mais fundamental para garantir a amplitude que se busca nesta pesquisa.

Os termos *Electronic Health Records* (EHR), *Construction Management Information Systems*, *Engineering Data Management System* (EDMS), *Engineering document management systems* (EDMS), foram desconsiderados para esta pesquisa. Isso se justifica pelo fato de se tratarem de sistemas de gestão de documentos com foco muito específico em uma área de conhecimento.

Os sistemas ligados a EHR, em sua maioria foca na gestão dos dados de pacientes adquiridos através de consultas médicas e exames. Os objetivos de muitas pesquisas encontradas estavam em estudar as tecnologias de previsão de diagnósticos a partir dos sintomas apresentados e das bases de dados de códigos e descrição de doenças, típicos desta área. Outro ponto discutido nas pesquisas é a interoperabilidade de dados através dos diferentes sistemas.

Já trabalhos que tratam de sistemas de gestão de dados de engenharia ou documentos de engenharia, focam no processo de gestão de projeto e em documentos típicos desse tipo de processo, onde há uma grande diversidade de documentos ligados ao projeto. Nesta área existe uma grande complexidade ao lidar com os documentos, visto que possuem grandes volumes de revisões de um mesmo documento e eles se relacionam entre si.

5.2.1.2. Atividade 2: Prototipagem

Após a identificação dos termos, iniciou-se a realização de novas buscas nas bases de dados. Agora o objetivo é mapear os termos para analisar a relevância de seus resultados.

Para a análise de relevância foram realizadas pesquisas de cada termo separadamente nas bases de dados da *Web of Science* e *Scopus*. Para essas bases optou-se por utilizar apenas os termos em inglês, visto que esta é a língua de indexação das mesmas. Foram selecionados dos 100 primeiros (“top 100”) registros do resultado de busca, ordenados por ordem de relevância.

Esses dados foram exportados para o Excel onde foram analisadas as informações de título, palavra-chave e resumo. O objetivo desse processo analisar a relevância dos termos a partir dos “top 100” artigos encontrados para a área de Gestão

Eletrônica de Documentos e encontrar a composição ideal da *string*. Os quantitativos estão apresentados no **Quadro 8**.

Quadro 8 – Quantitativos dos termos de GD

Resultado de busca por termos isolados		
TERMO	SCOPUS	WEB OF SCIENCE
<i>Electronic Recordkeeping System*</i>	15	4
<i>Electronic Recordkeeping</i>	50	20
<i>Electronic record* management System*</i>	106	58
<i>Electronic record* management</i>	294	146
<i>Electronic document management System*</i>	316	118
<i>Electronic document management</i>	593	239
<i>document management</i>	3628	1751
<i>Record* management</i>	8430	1254

Fonte: Elaborado pela autora (2024)

O próximo passo teve como objetivo a classificação dos artigos recuperados. Após a leitura do título, palavra-chave e resumo, os artigos foram classificados com Relevante, aborda assuntos relacionados a GED e como Não Relevante, não aborda o assunto diretamente, apesar de possuir o termo pesquisa, o foco das pesquisas não está na GED. O resultado deste passo foi a classificação dos termos mais relevantes, conforme apresentado no **Quadro 9**.

Quadro 9 – Porcentagem da análise de relevância dos termos de GD

Teste de relevância dos termos - TOP 100 artigos mais recentes			
TERMO	SCOPUS	WEB OF SCIENCE	MÉDIA
<i>Electronic document management System*</i>	86%	88%	87%
<i>Electronic record* management System*</i>	93%	92%	93%
<i>Electronic Recordkeeping System*</i>	87%	75%	81%
<i>Electronic document management</i>	82%	82%	82%
<i>Electronic record* management</i>	87%	93%	90%
<i>document management</i>	71%	73%	72%
<i>Electronic Recordkeeping</i>	73%	55%	64%
<i>Record* management</i>	48%	96%	72%

Fonte: Elaborado pela autora (2024)

A análise preliminar dos termos utilizados na busca revelou que termos mais gerais, como "*Records Management*", "*Electronic Recordkeeping*" e "*Document Management*", geram um volume de resultados extremamente alto. Embora esses resultados incluam técnicas de Inteligência Artificial relevantes para o objetivo da pesquisa, a seleção e filtragem de artigos que tratam exclusivamente de Gestão

Eletrônica de Documentos torna-se desafiadora devido ao volume excessivo e à diversidade de temas abordados.

Por outro lado, a busca com termos mais específicos, tais como *electronic recordkeeping*, *electronic recordkeeping system**", *lectronic record* management*", *electronic record* management system**", *electronic document management*" e *electronic document management eystem**", apresenta um foco mais direcionado à GED. Esses termos proporcionam resultados mais pertinentes e específicos ao contexto da gestão eletrônica de documentos, facilitando a identificação de artigos relevantes.

Entretanto, a utilização exclusiva desses termos específicos pode resultar na exclusão de artigos que, apesar de não se enquadrarem perfeitamente na definição restritiva de GED, contêm informações e técnicas que podem contribuir significativamente para a pesquisa.

Na sequência, são definidos os critérios de relevância para incorporação dos termos na *string* de GED.

5.2.1.3. Passo 3: Seleção dos Termos

Neste passo, foi definido um critério para a seleção dos termos, visando a criação de uma *string* de busca eficaz. Termos com classificação inferior a 80% foram desconsiderados, dessa forma, não farão parte da *string* de busca os termos: *document management*, *electronic recordkeeping*, *record management*.

Para refinar ainda mais a *string*, novas buscas foram realizadas, desta vez, os termos foram incorporando à *string* de forma progressiva e registrados seus quantitativos de registros recuperados, conforme **Quadro 9**.

Quadro 10 – Porcentagem da análise de relevância dos termos de GD

Busca progressiva dos termos de GED		
TERMO	SCOPUS	WEB OF SCIENCE
"Electronic Recordkeeping"	50	20
"Electronic Recordkeeping" AND "Electronic Recordkeeping System*"	50	20
"Electronic Recordkeeping" AND "Electronic Recordkeeping System*" AND "Electronic record* management"	343	164
"Electronic Recordkeeping" AND "Electronic Recordkeeping System*" AND "Electronic record* management" AND "Electronic record* management System*"	343	164
"Electronic Recordkeeping" AND "Electronic Recordkeeping System*" AND "Electronic record* management" AND "Electronic record* management System*" AND "Electronic document management"	925	399
"Electronic Recordkeeping" AND "Electronic Recordkeeping System*" AND "Electronic record* management" AND "Electronic record* management System*" AND "Electronic document management" AND "Electronic document management System*"	925	399

Fonte: Elaborado pela autora (2024)

Com essa metodologia é possível identificar termos que não afetam a composição da *string*, alguns por serem genéricos outros por estarem intrinsecamente ligados. Para a análise dos termos de GED, optou-se por manter apenas os termos mais genéricos, que foram: *Electronic Recordkeeping*, *Electronic record* management* e *Electronic document management*.

Com o trabalho de prototipagem, observou-se que bases de dados internacionais como *Web of Science*, *Scopus* e *LISTA* possuem uma estrutura de pesquisa avançada, com utilização de buscadores booleanos muito completa. Para essas bases as buscas foram feitas com os termos em inglês.

Já na base de dados *Brapci*, além de utilização dos termos em inglês, foi necessário utilizar também os termos em português (que possui mais variações). Houve dificuldades de combinação de mais de dois termos em uma única *string*, por esse motivo, na base de dados *Brapci*, a busca foi realizada pelos termos isolados, em inglês e em português.

O resultado final deste passo foi a *string* de busca elaborada e testada, pronta para ser utilizada na pesquisa, como pode ser observado no **Quadro 11**.

Quadro 11 – *String* de busca para GED

String de busca para GED	
Bae de dados	Descrição
Web of Science, Scopus e LISTA	("electronic recordkeeping" OR "electronic record* management" OR "electronic document management")
Brapci	"electronic recordkeeping"; "electronic record management", "electronic records management"; "electronic document management"; "gestão eletrônica de documentos"; "gerenciamento eletrônico de documentos"; "gestão eletrônica de registro"; "gestão eletrônica de registros"; "gerenciamento eletrônico de registro" e "gerenciamento eletrônico de registros"

Fonte: Elaborado pela autora (2024)

5.2.2. Definição dos Termos de IA e CD

Para composição da *string* de busca de termos de Inteligência Artificial e Ciência de dados procedeu-se seguindo procedimentos semelhantes aos descritos na elaboração da string de GED. Este processo garantiu a inclusão dos termos mais utilizados na literatura e necessários para investigar a aplicação dessas tecnologias na Gestão Eletrônica de Documentos (GED).

As fontes de Informação utilizadas para identificar os termos desta área foram:

- Relatório sobre inteligência artificial intitulado *WIPO Technology Trends 2019: Artificial Intelligence*, elaborado pela *World Intellectual Property Organization* (WIPO).
- Relatório sobre inteligência artificial, intitulado *Artificial Intelligence: A worldwide overview of AI patents and patenting by the UK AI sector*, elaborado por *Intellectual Property Office* (IPO);
- Tese de Doutorado intitulada *Inteligência Artificial no campo Ciência, Tecnologia e Sociedade: um estudo comparativo entre indicadores da produção científica*, elaborada por Luciana Castro Groenner da Universidade Federal de São Carlos.

O uso de fontes confiáveis garante maior precisão e credibilidade na escolha dos termos e na construção da *string* de busca, elemento essencial para alcançar resultados relevantes em uma pesquisa, evitando a inclusão de informações irrelevantes ou inadequadas.

Além disso, fundamentar-se em fontes confiáveis fortalece a análise e interpretação dos resultados, assegurando que a pesquisa seja conduzida com rigor e consistência metodológica. No quadro a seguir são apresentadas as *strings* de busca utilizadas para realização das pesquisas, conforme apresentado do **Quadro 12**.

Quadro 12 – Lista de termos d

Listagem de termos – IA e CD		
Termos encontrado	Termo pesquisado	Fonte
active learning	"active learning"	Groenner (2022)
adaboost	adaboost	Groenner (2022)
adversar* network*	"adversar* network**"	IPO (2024)
allocation	allocation	Groenner (2022)
ant colony	"ant colony"	IPO (2024)
artificial intelligence	"artificial intelligence"	Abadi (2020), IPO (2024)
association rule	"association rule"	IPO (2024)
augmented reality	"augmented reality"	WIPO (2019)
auto encod*	"auto encod**"	IPO (2024)
autonom* comput*	"autonom* comput**"	IPO (2024)
back propagat*	"back propagat**"	IPO (2024)
backpropagation	backpropagation	Groenner (2022)
back propogat*	"back propogat**"	IPO (2024)
bee colony	"bee colony"	IPO (2024)
bio inspired aproaches	"bio inspired aproaches"	WIPO (2019)
biometrics	biometrics	WIPO (2019)
character recognition	"character recognition"	Groenner (2022), WIPO (2019)
character* analy* (analysis e analytic)	"character* analy**"	Groenner (2022)
classification	classification	WIPO (2019)
clustering	clustering	Groenner (2022)
cognitiv* comput*	"cognitiv* comput**"	IPO (2024)
cognitive and neuroscience-inspired prcessing	"cognitive and neuroscience-inspired prcessing"	Abadi (2020)
collaborat* filter*	"collaborat* filter**"	IPO (2024)
comput* creativity	"comput* creativity"	Groenner (2022)
computer vision	"computer vision"	WIPO (2019)
connections	connections	Groenner (2022)

Listagem de termos – IA e CD		
Termos encontrado	Termo pesquisado	Fonte
data mining	"data mining"	Abadi (2020)
data science	"data science"	-
decision support system	"decision support system"	Abadi (2020)
decision tree	"decision tree"	Groenner (2022)
deep learning	"deep learning"	Abadi (2020), IPO (2024), WIPO (2019)
deep belief network*	"deep belief network**"	IPO (2024)
description logic*	"description logic**"	WIPO (2019)
descriptive model*	"descriptive model**"	Groenner (2022)
dialogue	dialogue	WIPO (2019)
differential* evol* algorithm*	"differential* evol* algorithm**"	IPO (2024)
dimensional* reduc*	"dimensional* reduc**"	IPO (2024)
distributed ai	"distributed ai"	WIPO (2019)
embedding*	embedding*	Groenner (2022)
ensemble learn*	"ensemble learn**"	IPO (2024)
evolution* algorithm*	"evolution* algorithm**"	IPO (2024)
evolution* comput*	"evolution* comput**"	IPO (2024)
expert system*	"expert system**"	Groenner (2022), WIPO (2019)
face* analy* (analysis e analytic)	"face* analy**"	Groenner (2022)
facial analy* (analysis e analytic)	"facial analy**"	Groenner (2022)
facial recognition	"facial recognition"	Groenner (2022)
factorization machin* (factorisation machin*)	"factorization machin**"	IPO (2024)
feature engineer*	"feature engineer**"	IPO (2024)
feature extraction	"feature extraction"	Abadi (2020), IPO (2024)
feature selection	"feature selection"	Abadi (2020), Groenner (2022), IPO (2024)
fire fly	"fire fly"	IPO (2024)
fuzzy (environment*, logic, number, set*, system, c)	*fuzzy*	Groenner (2022), IPO (2024), WIPO (2019)
gaussian mixture model	"gaussian mixture model"	IPO (2024)
gaussian process*	"gaussian process**"	IPO (2024)
genetic algorithm	"genetic algorithm"	Abadi (2020)
genetic program*	"genetic program**"	IPO (2024)

Listagem de termos – IA e CD		
Termos encontrado	Termo pesquisado	Fonte
gradient tree boosting	"gradient tree boosting"	Groenner (2022)
hand writing analy* (analysis e analytic)	"hand writing analy**"	Groenner (2022)
hand writing recognition	"hand writing recognition"	Groenner (2022)
hidden markov model	"hidden markov model"	Groenner (2022)
high dimensional* model*	"high dimensional* model**"	IPO (2024)
high dimensional* system*	"high dimensional* system**"	IPO (2024)
high dimentional data	"high dimentional data"	Abadi (2020), IPO (2024)
high dimensional feature*	"high dimensional feature**"	IPO (2024)
high dimensional input*	"high dimensional input**"	IPO (2024)
high dimensional space*	"high dimensional space**"	IPO (2024)
image segmentation	"image segmentation"	WIPO (2019)
video segmentation	"video segmentation"	WIPO (2019)
inductive reasoning	"inductive reasoning"	Groenner (2022)
information extraction	"information extraction"	WIPO (2019)
intelligent application	"intelligent application"	Abadi (2020)
intelligent system	"intelligent system"	Abadi (2020)
k means	"k means"	IPO (2024)
kernel learn*	"kernel learn**"	IPO (2024)
knowledge processing	"knowledge processing"	Abadi (2020)
knowledge reasoning	"knowledge reasoning"	WIPO (2019)
knowledge representation	"knowledge representation"	Abadi (2020)
latent dirichlet	"latent dirichlet"	Groenner (2022)
latent representation	"latent representation"	WIPO (2019)
latent semantic analysis	"latent semantic analysis"	Groenner (2022)
latent variable*	"latent variable**"	IPO (2024)
learning algorithm	"learning algorithm"	Groenner (2022)
learning model	"learning model"	Groenner (2022)
link* predict*	"link* predict**"	IPO (2024)
logic programming	"logic programming"	WIPO (2019)
logical learning	"logical learning"	WIPO (2019)
logistic regression	"logistic regression"	Groenner (2022)
machine intelligen*	"machine intelligen**"	IPO (2024)

Listagem de termos – IA e CD		
Termos encontrado	Termo pesquisado	Fonte
machine learn*	"machine learn**"	Abadi (2020), Groenner (2022), IPO (2024), WIPO (2019)
machine perception	"machine perception"	Abadi (2020)
machine translation	"machine translation"	WIPO (2019)
map reduce	"map reduce"	IPO (2024)
memetic algorithm*	"memetic algorithm**"	IPO (2024)
morphology	morphology	WIPO (2019)
multi agent system	"multi agent system"	Groenner (2022)
multi* label* classif*	"multi* label* classif**"	IPO (2024)
multi* objective* optim*	"multi* objective* optim**"	IPO (2024)
multi* objective* algorithm*	"multi* objective* algorithm**"	IPO (2024)
multilayer perceptron	"multilayer perceptron"	Groenner (2022)
multi task learning	"multi task learning"	WIPO (2019)
natural language generation	"natural language generation"	Groenner (2022), WIPO (2019)
natural language processing	"natural language processing"	Abadi (2020), Groenner (2022)
natural gradient	"natural gradient"	IPO (2024)
neural network*	"neural network**"	Abadi (2020), IPO (2024), WIPO (2019)
neural turing	"neural turing"	IPO (2024)
neuro morph* comput*	"neuro morph* comput**"	IPO (2024)
non negative matri* factor*	"non negative matri* factor**"	IPO (2024)
object recognition	"object recognition"	Abadi (2020), IPO (2024)
object tracking	"object tracking"	WIPO (2019)
objective function	"objective function"	Groenner (2022)
ontology	ontology	WIPO (2019)
overfitting	overfitting	Groenner (2022)
particle swarm*	"particle swarm**"	IPO (2024)
pattern recogni*	"pattern recogni**"	Abadi (2020), IPO (2024)
phonology	phonology	WIPO (2019)
policy gradient method	"policy gradient method"	IPO (2024)
predictive analy* (analysis e analytic)	"predictive analy**"	Groenner (2022), WIPO (2019)
predictive model	"predictive model"	Groenner (2022)

Listagem de termos – IA e CD		
Termos encontrado	Termo pesquisado	Fonte
probabilist* algorithm*	"probabilist* algorithm**"	Groenner (2022)
probabilist* approach*	"probabilist* approach**"	Groenner (2022)
probabilist* method*	"probabilist* method**"	Groenner (2022)
probabilist* technique*	"probabilist* technique**"	Groenner (2022)
probabilist* graphical*	"probabilist* graphical**"	WIPO (2019)
probabilist* reason*	"probabilist* reason**"	WIPO (2019)
q-learn*	"q-learn**"	IPO (2024)
random-forest*	"random-forest**"	Groenner (2022), IPO (2024)
rankboost	"rankboost"	Groenner (2022)
recommend* system*	"recommend* system**"	Groenner (2022), IPO (2024)
recurrent neural network	"recurrent neural network"	Abadi (2020)
regression mode*	"regression mode**"	Groenner (2022)
regression tree*	"regression tree**"	WIPO (2019)
reinforcement learning	"reinforcement learning"	Abadi (2020), Groenner (2022), IPO (2024), WIPO (2019)
relational learning	"relational learning"	WIPO (2019)
robotic*	robotic*	WIPO (2019)
rule learning	"rule learning"	WIPO (2019)
scene undertanding	"scene undertanding"	WIPO (2019)
self learning	"self learning"	Groenner (2022)
sentiment analysis	"sentiment analysis"	IPO (2024), WIPO (2019)
sparse represent*	"sparse represent**"	Abadi (2020), IPO (2024)
sparse code*	"sparse code**"	IPO (2024)
spectral cluster*	"spectral cluster**"	IPO (2024)
speech analy* (analysis e analytic)	"speech analy**"	Groenner (2022)
speech processing	"speech processing"	WIPO (2019)
speech recognition	"speech recognition"	Groenner (2022), WIPO (2019)
speech synthesis	"speech synthesis"	WIPO (2019)
speech to speech	"speech to speech"	WIPO (2019)
stochastic algorithm	"stochastic algorithm"	Groenner (2022)
stochastic approach	"stochastic approach"	Groenner (2022)

Listagem de termos – IA e CD		
Termos encontrado	Termo pesquisado	Fonte
stochastic method	"stochastic method"	Groenner (2022)
stochastic technique	"stochastic technique"	Groenner (2022)
stochastic gradient*	"stochastic gradient**"	Groenner (2022), IPO (2024)
*supervised learning	**supervised learning"	Abadi (2020), Groenner (2022), IPO (2024), WIPO (2019)
*supervised training	**supervised training"	Groenner (2022)
support vector machine*	"support vector machine**"	Groenner (2022), IPO (2024), WIPO (2019)
swarm behav*	"swarm behav**"	IPO (2024)
swarm intelligen*	"swarm intelligen**"	Groenner (2022), IPO (2024)
target function	"target function"	Groenner (2022)
test data set	"test data set"	Groenner (2022)
text analy* (analysis e analytic)	"text analy**"	Groenner (2022)
text recognition	"text recognition"	Groenner (2022)
training data set	"training data set"	Groenner (2022)
transfer learn*	"transfer learn**"	Groenner (2022), IPO (2024)
validation data set	"validation data set"	Groenner (2022)
variation* infer*	"variation* infer**"	IPO (2024)
vector machine*	"vector machine**"	IPO (2024)
xgboost	xgboost	Groenner (2022)

Fonte: Elaborado pela autora (2024)

Após a identificação e listagem dos termos, realizou-se a busca pelos termos isolados nas bases de dados *Web of Science* e *Scopus*, com objetivo de identificar sua existência nas bases de dados. Os termos que não tiveram retorno foram excluídos da lista.

Com os termos restantes passou-se para a realização de novas buscas, através do processo de incorporando à *string* de forma progressiva. No entanto, para a identificação a relevância desses termos IA e CD, é preciso considerar o contexto da área de GED. Com isso, as buscas progressivas foram realizadas a partir da

combinação da *string* de GED com os termos de IA e CD. No Apêndice 1 tem o registro de todo esse processo.

Essa metodologia permite identificar termos relevantes que, em combinação com os termos de GED, apresentam resultados significativos para esta pesquisa. O resultado final é a definição de uma *string* de busca robusta e direcionada, construída a partir da combinação estratégica de termos das áreas de Gestão Eletrônica de Documentos, Inteligência Artificial e Ciência de Dados.

Essa abordagem garante maior precisão na recuperação de informações, alinhando a pesquisa aos objetivos estabelecidos e otimizando a análise dos dados coletados, no **Quadro 13** é apresentada a *string* final de busca.

Quadro 13 – *String* final de busca

STRING FINAL - GED - GESTÃO ELETRÔNICA DE DOCUMENTOS: SCOPUS: 168 registros encontrados WEB OF SCIENCE: 59 registros encontrados		Web of Science = Tópico Scopus = Article title, Abstract, Keywords Não foi feito nenhum filtro, considerou-se todo o resultado.
<p><i>("Electronic Recordkeeping" OR "Electronic record* management" OR "Electronic document management") AND</i></p> <p><i>("artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "artificial intelligence" OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR *fuzzy* OR robotic* OR dialogue OR "predictive analy*" OR "recommend* system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "association rule" OR "augmented reality" OR "image segmentation" OR "genetic algorithm" OR "descriptive model*" OR "probabilist* method*" OR "self learning")</i></p>		

Fonte: Elaborado pela autora (2024)

Na sequência será realizada a condução da Fase 2 da Revisão Sistemática da Literatura, com a aplicação da *string* de busca e seleção dos registros que irão compor o *corpus* documental para análise.

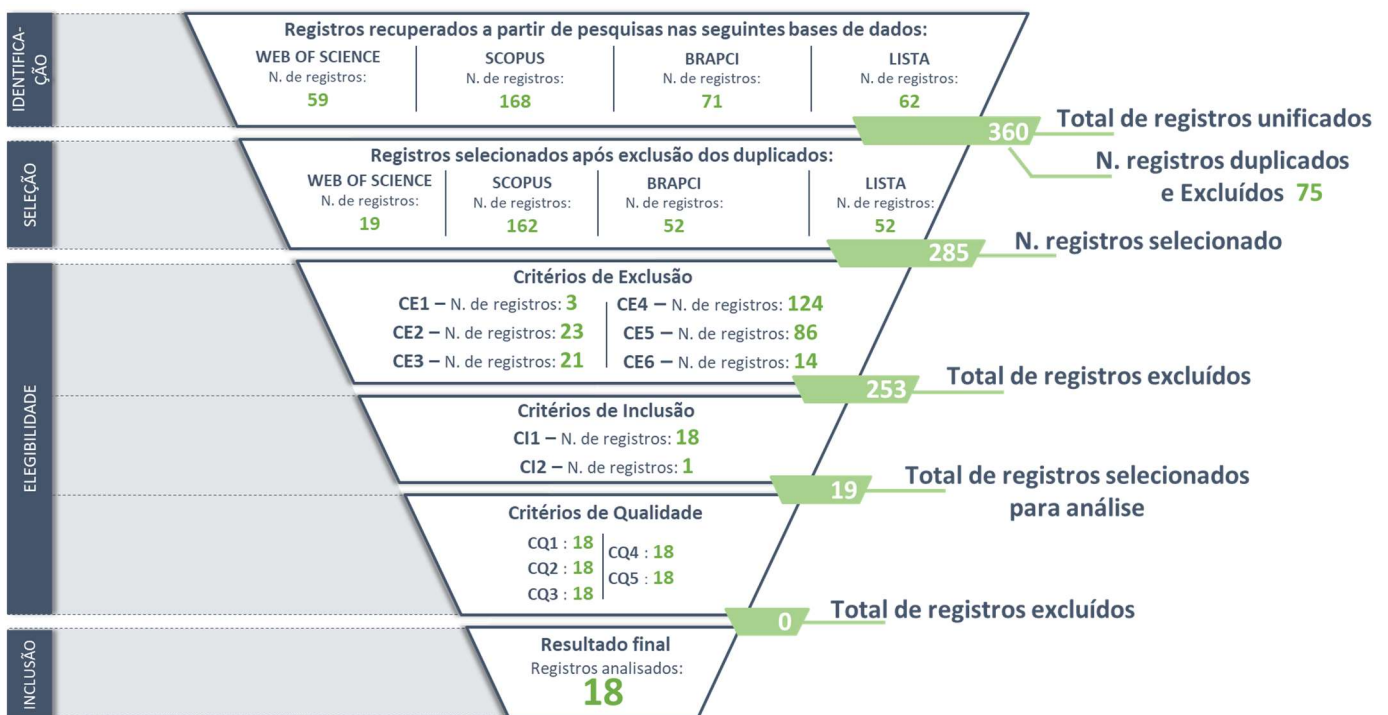
5.3. Tratamento e Seleção de Artigos (Etapa 2)

O tratamento e seleção de artigos corresponde à primeira etapa da Fase 2 da Revisão Sistemática da Literatura, focando na condução da revisão. Abrange todas as atividades necessárias para a preparação e análise das informações, realizadas de forma manual com os recursos dos programas da Microsoft. O processo inclui a **Identificação (2.1)**, que consiste na lista unificada e padronizada dos dados

provenientes das bases de dados científicas. Em seguida ocorre a **Seleção (2.2)**, análise e eliminação de artigos duplicados. A etapa de **Elegibilidade (2.3)**, onde, a partir da leitura do título, palavras-chave e resumo, são aplicados os critérios de inclusão, exclusão e qualidade. Por fim, realiza-se a etapa de **Inclusão (2.4)**, que contempla a preparação dos registros para leitura do texto integral, anotação e análise temática e de conteúdo, detalhadas nas subseções a seguir.

Na **Figura 6** são apresentados todos os passos desenvolvidos para seleção dos artigos que irão compor o *corpus* de análise desta pesquisa.

Figura 6 – Fluxograma Prisma 2020



Fonte: Elaborado pela autora (2024)

A seguir são detalhadas cada uma das etapas do processo de tratamento e seleção de dados para análise.

5.3.1. Identificação (Etapa 2.1)

Após as pesquisas nas bases de dados científicos, são extraídos os relatórios do resultado das buscas, e unificados em uma única planilha. É realizada a padronização dos metadados, garantindo que todos os registros possuam os mesmos tipos de informações e contenham todas as informações necessárias para identificar, analisar e referencia-los.

Para garantir a identificação da origem da base de dado, cada artigo recebeu uma codificação sequencial. Ela identifica a base de origem da qual foi extraído, uma sigla do objeto de pesquisa (GED – Gestão Eletrônica de Documentos), a data da pesquisa (AA: ano, MM: mês e DD: dia) e um sequencial numérico de 4 dígitos. Dessa forma os códigos utilizados seguiram o seguinte padrão:

- **Brapci:** BRA-GED-AAMMDD-NNNN (ex.: BRA-GED-240820-0001)
- **LISTA:** LIS-GED-AAMMDD-NNNN (ex.: LIS-GED-240821-0001)
- **Scopus:** SCO-GED-AAMMDD-NNNN (ex.: SCO-GED-240806-0001)
- **Web of Science:** WOS-GED-AAMMDD-NNNN (ex.: WOS-GED-240806-0001)

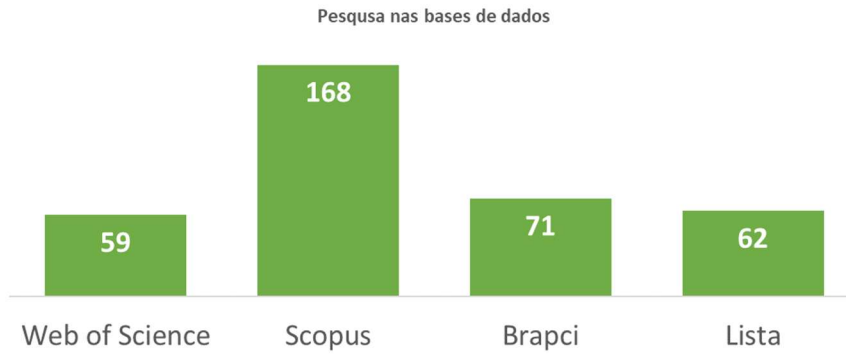
Para armazenar estas informações foi utilizado programa Microsoft Excel. Com ele é possível realizar a manipulação dos dados, como unificação, comparação, anotações etc., como também cálculos matemáticos para a análise quantitativa. Nesta planilha, cada linha representa um artigo e cada coluna representa um metadado do artigo.

As informações registradas na planilha são: Autor nome abreviado, Autor nome completo, Título, Título da fonte, Palavra-Chave do autor, Resumo, Ano, Idioma, Tipo de Documento, Título da Conferência, Data da Conferência e Local da Conferência, ORCID, Volume, Edição, Página inicial, Página final, DOI, DOI link, Número de páginas e Área/Categoria.

Para os artigos que não apresentarem palavra-chave, mas passarem em todos os critérios de avaliação, as palavras-chaves serão incluídas pelo pesquisador para realização de análises quantitativas. Para artigos duplicados, será considerada a informação do registro mais completo, ou selecionado o registro mais completo e complementadas as informações a partir dos registros duplicados.

Com as pesquisas realizadas nas bases de dados foram recuperados um total de 360 registros, sendo: 168 registros na *Scopus*, 71 registros na *Brapci*, 62 registros na *LISTA* e 59 registros na *Web of Science*, conforme demonstrado na **Figura 7**.

Figura 7 – Pesquisa nas bases de dados



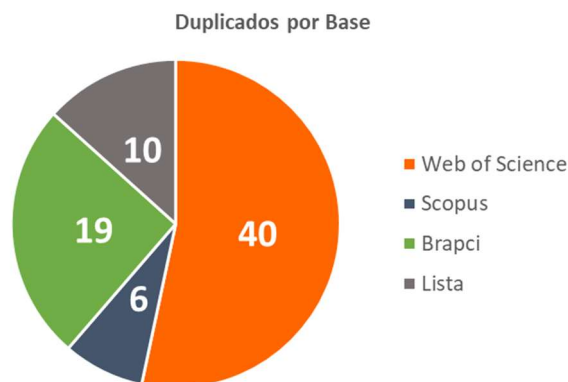
Fonte: Elaborado pela autora (2024)

Como resultado dessa etapa temos a relação de todos os artigos recuperados, identificados por suas respectivas bases de dados e, por fim, campos e metadados padronizados. Essa preparação possibilita o desenvolvimento das próximas etapas.

5.3.2. Seleção (Etapa 2.2)

A etapa de seleção é dedicada à exclusão dos artigos duplicados. A análise de redundância é realizada comparando título, autor, palavra chave e resumo. Dos 360 artigos recuperados, 75 eram duplicados, sendo: 40 da *Web of Science*, 19 da Brapci, 10 da Lista e 6 da *Scopus*, conforme ilustra a **Figura 8**.

Figura 8 – Duplicidade por base



Fonte: Elaborado pela autora (2024)

Todo os 40 artigos duplicados na *Web of Science*, apareciam também na *Scopus*, já as duplicidades presentes nas bases Brapci e LISTA, a grande maioria estão duplicados no próprio resultado de pesquisa.

Para a análise de duplicidade, foi realizada a comparação, em um primeiro momento entre os registros da *Scopus* e *Web of Science*.

A primeira comparação realizada foi pelo título, títulos idênticos que tenham as mesmas palavras-chave são excluídos. Depois a comparação é pelo nome dos autores, autores iguais com título diferente, mas que tratam do mesmo trabalho, para tanto analisa-se palavra-chave e resumo.

Mais frequentemente os registros da Web of Science possuíam menos informações em comparação com suas duplicatas na *Scopus*, dentre as informações que menos aparece são palavra-chave indexada pela base, DOI e link de acesso à publicação.

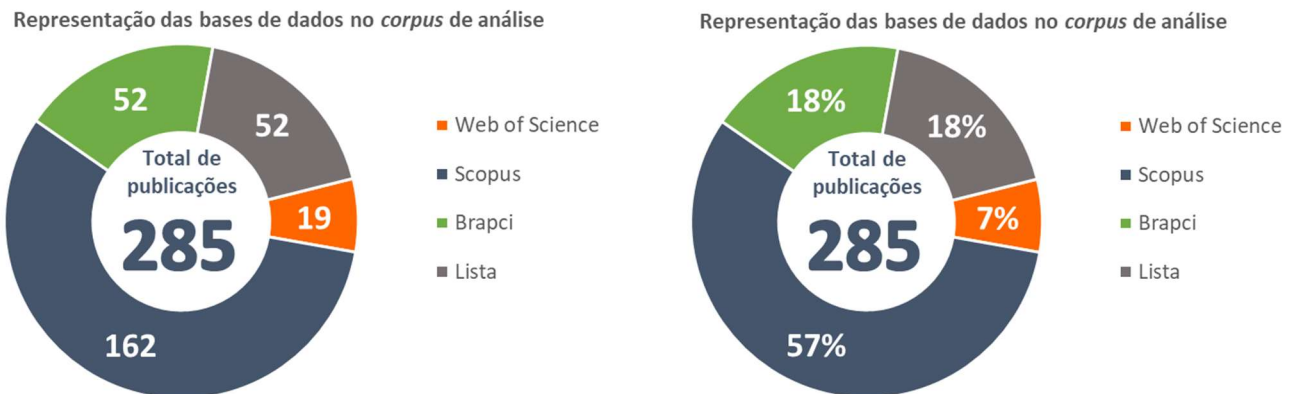
Dos 40 registros duplicados, 37 estão repetidos tanto na base da Scopus quanto na base da Web of Science e três estão repetidos no próprio resultado de busca. Estes últimos possuem são exatamente os mesmos trabalhos, mas com títulos diferentes.

Por fim foi realizada a comparação entre os registros remanescentes da Scopus e Web of Science e os registros da Brapci e LISTA, tendo em vista que alguns periódicos indexados na Brapci e LISTA também estão na *Web of Science*.

Dos registros duplicados é necessário selecionar um registro para manter no banco de dados e um registro para excluir. O registro extraído da *Scopus* foi escolhido para manter, visto que é a base comum de comparação e as informações existentes no banco de dados são mais completos.

Ao final desta etapa foram excluídos 75 registros duplicados, restando na base 285 registros, sendo 162 da *Scopus*, representando 57% da base, 52 da Brapci e da LISTA, o que representa 18% cada e 19 da *Web of Science*, representando 7%, conforme ilustra a **Figura 9**.

Figura 9 – Representação das bases de dados.



Fonte: Elaborado pela autora (2024)

5.3.3. Elegibilidade (Etapa 2.3)

Esta é uma etapa analítica, tem como objetivo selecionar os artigos que formarão o *corpus* final como base para o desenvolvimento desta pesquisa. Primeiro foi realizada a leitura do título, palavras-chaves e resumo dos arquivos recuperados. Depois foram eliminados todos os artigos que não atendem aos critérios de exclusão definidos no protocolo de revisão sistemática apresentados no **Quadro 4**, na sequência, prosseguiu com a análise dos artigos que se enquadram nos critérios de inclusão, apresentados no **Quadro 5**.

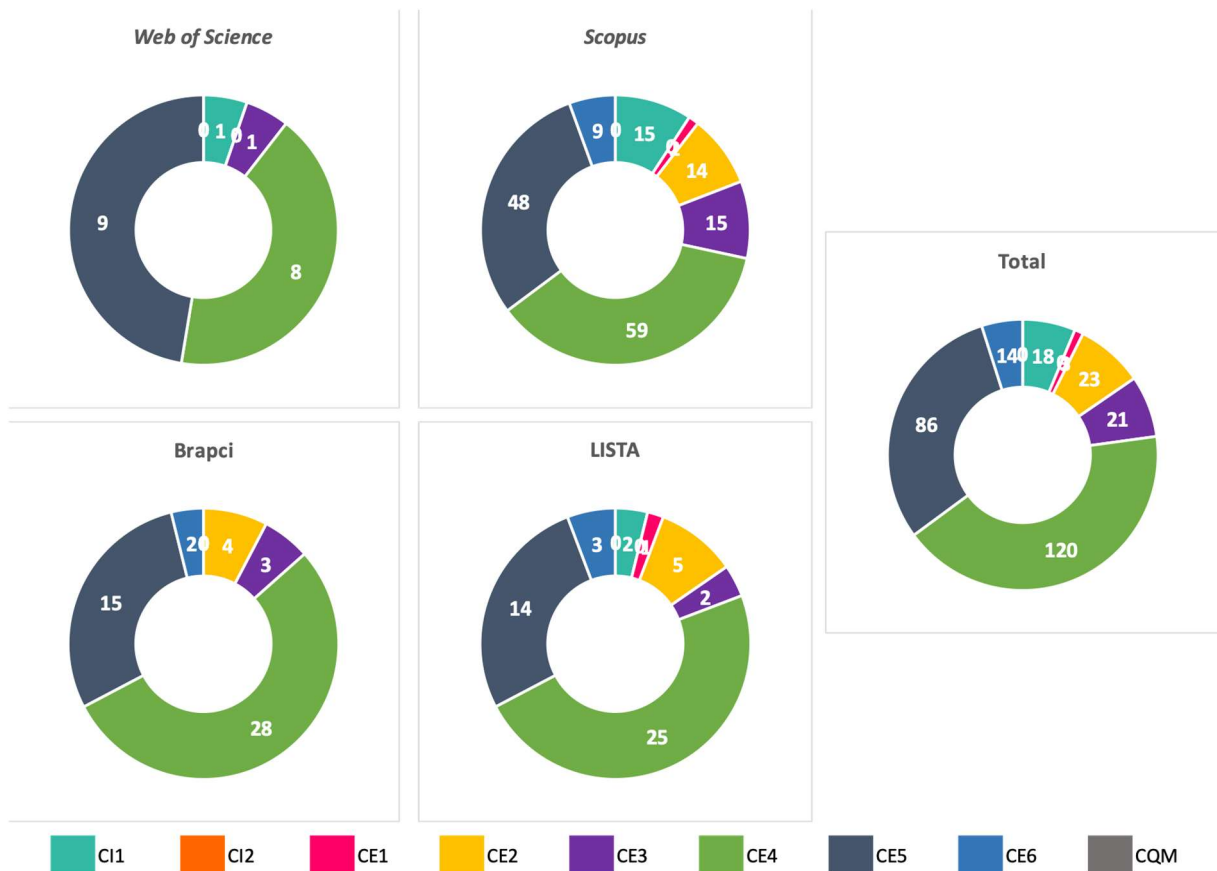
Escolheu-se aplicar primeiramente os critérios de exclusão, pois os artigos que atendiam a pelo menos um dos critérios seriam eliminados das análises seguintes. Para os artigos restantes foram aplicados os critérios de inclusão, na sequência os artigos que atendiam a um dos dois critérios de inclusão seriam selecionados para a terceira análise, que contemplou a leitura do texto integral.

Por fim são aplicados os critérios de qualidade, onde foi analisado se os artigos escolhidos possuem todas as informações necessárias para atingir ao objetivo de pesquisa deste trabalho.

A avaliação dos critérios de exclusão e de inclusão foram aplicados seguindo os seguintes passos: primeiro, uma análise do título, a verificação das palavras-chave do autor e da base de dados, na sequência, uma avaliação do resumo e por fim, a leitura dinâmica de partes do texto, como introdução, desenvolvimento e conclusão. Contudo, para esta etapa, não era necessário executar todos os passos obrigatoriamente, assim que o critério era identificado, registrava-se na planilha e

passava-se para a análise do artigo seguinte. Na **Figura 10** são apresentados os quantitativos para cada base de dados.

Figura 10 – Distribuição dos critérios aplicados em cada base de dados



Fonte: Elaborado pela autora (2024)

Foram identificados três registros que se enquadram no CE1 (o estudo está escrito em um idioma diferente de português, inglês e espanhol), 23 registros que se enquadram no CE2 (o texto completo não possui acesso aberto ou não pode ser acessado por outro motivo), 21 registros que se enquadram no CE3 (o estudo é livro, capítulo de livro, editorial, nota, carta, *proceedings* (anais), conference review (revisão de conferência) e *retracted* (retração)), 120 registros que se enquadram no CE4 (o estudo aborda apenas o tema GED sem intersecção com IA e CD), 86 registros que se enquadram no CE5 (o estudo apenas cita GED como ferramenta usada na pesquisa, como resultado de pesquisa ou sugestão de uso), 14 registros que se enquadram no CE6 (artigos anteriores a 2000). Com isso somam-se 267 artigos a serem excluídos, restando 18 para análise quanto ao critério de inclusão.

Foram identificados 18 registros que atendem ao CI1 (o estudo discute ou propõe técnica, estratégia, ferramenta ou qualquer abordagem de IA ou de CD ligada à GED), zero registros que atendem ao CI2 (o estudo não propõe uma abordagem

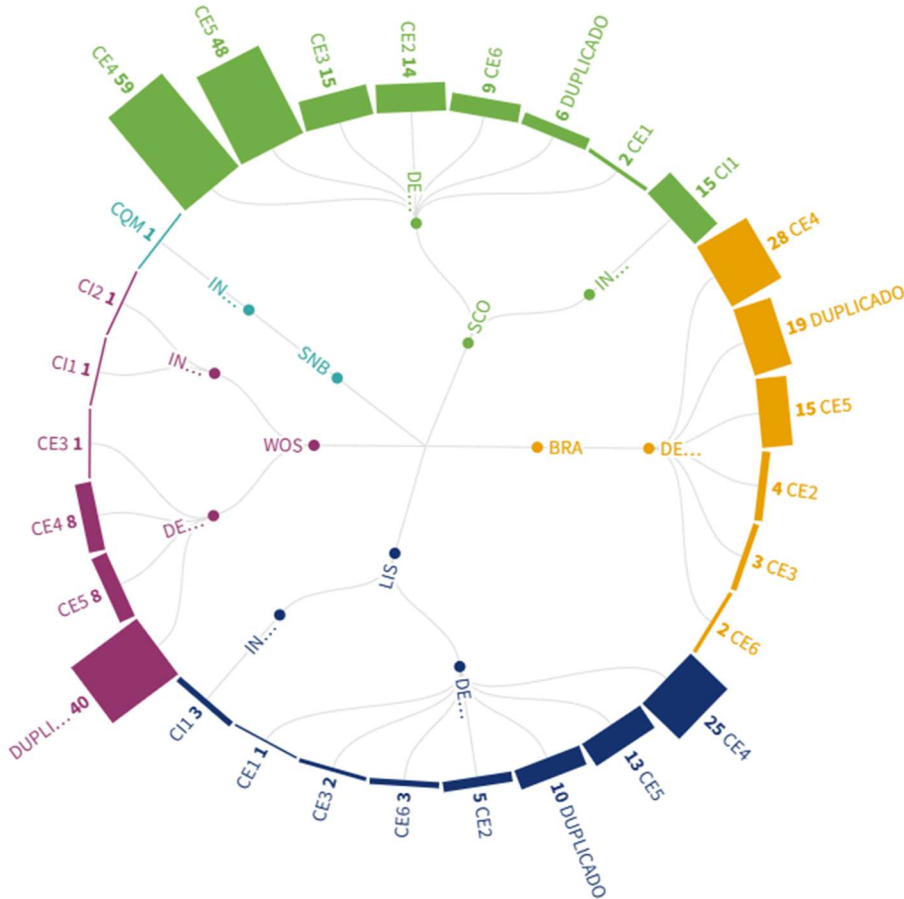
específica, mas discute possibilidades, tendências, benefícios e implicações na intersecção entre GED, IA e CD). Com isso somam-se 18 artigos selecionados para leitura do texto integral e análise quanto ao critério de qualidade.

Para cada critério de qualidade foram atribuídas três notas: 1 (um) - o critério é completamente atendido; 0,5 (meio) - o critério é parcialmente atendido; e 0 (zero) - o critério não é atendido. Para atender a este critério de qualidade, a soma de todos os critérios não podem ser inferiores a 2,5. Para atribuir as notas para cada critério de qualidade, fez-se necessária a leitura do texto integral e como resultado, nenhum dos 18 artigos receberam a soma das notas inferior a 2,5 pontos, sendo todos mantidos para a análise de conteúdo.

5.3.4. Inclusão (Etapa 2.4)

Ao final deste processo de análise de pertinência, foram selecionados 18 artigos para realização da etapa de análise temática e de conteúdo. O resumo da aplicação dos critérios e o resultado final da etapa de tratamento e seleção de artigos pode ser observado na **Figura 11**.

Figura 11 – Aplicação dos critérios



Fonte: Ferramenta Flourish, elaborado pela autora (2024)

Todos os artigos selecionados foram baixados e salvos em uma única pasta. Optou-se por adotar duas formas de anotação das informações extraídas dos artigos, a primeira em formato de planilha para melhor quantificar os dados coletados e a segunda em formato de texto, que permite descrever melhor as informações adquiridas.

Para o formato planilha foi utilizado o Excel extraída da planilha de tratamento dos dados, foram registradas informações como: Código, Autor, Título, Objetivo, Natureza da pesquisa, Tipo de Amostra utilizada, Tipo de Instituição impactada no estudo, onde pode ser aplicado, País de Origem dos Autores, Universidade/Instituição de origem dos autores, Área/Setor de estudo, Classificação da Solução Apresentada, tipo de abordagem.

Para o formato texto foi criada de uma “Ficha de Anotação” (APÊNDICE 2 – Ficha de Anotações), para cada artigo, com modelo padronizado, onde foram sendo realizadas anotações durante a leitura, que compreendiam os termos e significados fornecidos pelos autores, etapas de desenvolvimentos, descrição de informações pertinentes para o entendimento do artigo, dúvidas e descobertas.

Com a definição do *corpus* de análise, deu-se início à leitura de texto integral foi possível esboçar as primeiras estruturas e conceitos para classificação, o tipo de informação necessária para extrair desses artigos. Foram feitos alguns paralelos entre os dados já mapeados para entender as possíveis relações existentes. Com o entendimento mais amplo dos conteúdos, foi feita uma nova análise dos artigos excluídos classificados como CE4 e CE5 para garantir a qualidade dessa análise.

6. ANÁLISE DE DADOS

Esta seção corresponde à Fase 2 da Revisão Sistemática da Literatura, focando na condução da revisão. O objetivo central é apresentar os achados desta pesquisa, buscando responder às questões que guiaram todo o desenvolvimento do estudo.

A análise de conteúdo tem como objetivo categorizar e classificar qualquer o conteúdo dos artigos levantados a partir da aplicação de um protocolo revisão sistemática de literatura. A análise de conteúdo tem como objetivo identificar características e conceitos de qualquer conteúdo e utiliza-los para descrever, comparar, identificar relações entre os diferentes elementos e quantifica-los (Carlomagno; Rocha, 2016).

Para essa análise foi realizada a leitura do texto integral em momentos e com objetivos diferentes. Num primeiro momento foi realizada uma leitura “transversal” que buscou conhecer os assuntos tratados nos artigos. Iniciou-se um primeiro levantamento de dados e informações e novas pesquisas foram necessárias para entender os conceitos apresentados.

Como resultado, definiu-se os critérios de análise e o tipo de informação que seria extraída dos artigos. Passou se então para nova rodada de leitura de texto integral. Para esse processo foi elaborada uma ficha de leitura contendo um *template* (apresentada no APÊNDICE 2 – Ficha de Anotações) para anotação das informações coletadas.

No **Quadro 14** é apresentada a relação dos 18 artigos selecionados. Inclui o código de identificação, título, um resumo do assunto tratado no artigo e a citação dos respectivos autores.

PESQUISAS ANALISADAS				
	Código	Título	Assunto	Citação
01	LIS-GED-240821-0027	CADWeb -- Categorização automática de documentos digitais.	A pesquisa apresenta uma ferramenta intitulada CADWeb, que foi desenvolvida para realizar a categorização (classificação) automática de documentos em PDF. Utiliza-se de categorias pré-definidas e cada documento pode ser enquadrado em uma ou mais categorias. O sistema faz a leitura do documento, extrai todo o texto, realiza o processamento do texto para permitir realizar os cálculos matemáticos que irá realizar a categorização a partir da frequência e proximidade das palavras. Como resultado, a aplicação apresentou um percentual de 84,6% de acerto.	Gomes e Moraes FI (2011)
02	LIS-GED-240821-0032	Clustering tagged documents with labeled and unlabeled documents	A pesquisa apresenta um algoritmo, desenvolvido pelos autores, de classificação automática de documentos, chamado <i>Constrained-PLSA</i> , utilizando a técnica de agrupamento (<i>clustering</i>), baseado em aprendizagem de máquina não supervisionado (<i>machine learning semi-supervised</i>). O objetivo é identificar a eficiência na classificação de documentos por agrupamento, para documentos rotulados (<i>labeled</i>) e não rotulados (<i>unlabeled</i>). Além desse algoritmo, os autores realizaram testes em outros 3 algoritmos de classificação com diferentes características. Apesar de aplicarem os testes em um corpus com apenas resumos de artigos, os autores consideram os resultados positivos, afirmam que quase todos os métodos podem se beneficiar do uso de rótulos, e que o <i>Constrained-PLSA</i> supera os outros métodos em diversos aspectos.	Liu, Hsaio e Chen (2011)
03	SCO-GED-240806-0006	Mechanisms for optimization of detection and correction of text errors based on combining multilevel morphological analysis with n-gram models	A pesquisa apresenta uma solução para tornar a correção automática mais eficiente, segura e precisa, e faz a aplicação em um Sistema de Gerenciamento Eletrônico de Documentos (SGED), testando com 100 documentos. Os autores apresentam e comprovam a hipótese de que ao analisar padrões de distribuição de erros e calcular suas probabilidades, é possível desenvolver um sistema mais eficaz de detecção e correção de erros ortográficos e semânticos em documentos eletrônicos. A inovação neste trabalho está em tratar os erros de forma probabilística, abordando não apenas erros isolados, mas também através de padrões múltiplos num conjunto de documentos.	Jumanov e Karshiev (2020)
04	SCO-GED-240806-0008	Implementation of Decision Support Subsystem in Electronic Document Systems Using Machine Learning Techniques	O trabalho apresenta uma metodologia para desenvolver um subsistema de suporte à tomada de decisão integrado a um sistema de gestão eletrônica de documentos (SGED). Utilizando técnicas de aprendizado de máquina, o subsistema compara a similaridade de desenhos técnicos de projeto, eliminando duplicidades no banco de dados. Além disso, automatiza a tomada de decisões ao sugerir as próximas operações com base no histórico de ações do usuário, promovendo eficiência, consistência e otimização nos processos de gestão documental.	Obukhov, Krasnyansky e Nikolyukin (2019)
05	SCO-GED-240806-0013	The algorithm of document classification of research and education institution using machine learning methods	Os autores analisaram seis diferentes algoritmos de classificação de documentos, utilizando a abordagem de aprendizado de máquina, para categorizar 3.000 (três mil) documentos de uma instituição científica e educacional. A comparação destacou que não há uma única abordagem perfeita, no entanto, os autores consideraram os resultados positivos, tanto em relação ao tempo de treinamento do classificador, quanto à precisão de classificação na análise dos documentos.	Krasnyansky, Obukhov e Solomatina (2019)

PESQUISAS ANALISADAS				
	Código	Título	Assunto	Citação
06	SCO-GED-240806-0018	Machine Learning Approach for Automatic Categorization of Service Support Requests on University Information Management System	Os autores aplicaram cinco abordagens de aprendizado de máquina para categorização automática de solicitações de suporte de serviço no sistema de gerenciamento de informações da universidade. Foram processados 17.831 relatórios de bugs e solicitações de suporte de serviço, que deveriam ser categorizados de acordo com quatro categorias e pré-definidas (categorias incluem sistema de informação estudantil, sistema de informação de pessoal, sistema de gerenciamento eletrônico de documentos e sistema de projetos de pesquisa científica). A precisão na classificação variou entre 78,2 a 92,26, resultado considerados promissores, pelos autores.	Onan, Atik Yalçın (20
07	SCO-GED-240806-0044	Development of a Methodology for Protecting a Software Package for Electronic Document Management	O artigo apresenta diversas técnicas de autenticação para uso em sistema de gestão eletrônica de documentos e propõe o uso em conjunto de duas técnicas de autenticação biométrica com a técnica de geração de sequência de chaves criptográficas. Realizam a comparação de duas técnicas (fuzzy extractor e redes neurais) e identificam que o uso de redes neurais é mais eficiente e reduz a quase zero o índice de falsificação.	Fayziev et al (2021)
08	SCO-GED-240806-0048	uAIS: An experience of increasing performance of NLP information extraction tasks from legal documents in an electronic document management system	O artigo apresentou uma abordagem de Processamento de Linguagem Natural (PLN), que permite diminuir a velocidade e a demanda por recurso da CPU para o processo de extração de informações de documentos, de um Sistema de Gestão Eletrônica de Documentos (SGED), já em operação que realiza a extração de aproximadamente 55 dados de documentos legais. A proposta é distribuir, através da implementação de microsserviços, o processamento de dado. Os autores consideram resultados são promissores e mostram uma melhoria de 77% de desempenho, que reflete em aumento significativo da produtividade de sistemas baseados em outras abordagens.	Ruiz, et al (2020)
09	SCO-GED-240806-0050	Localization of scores and average in Algerian baccalaureate transcripts	O artigo propõe um método para a extração automática de notas e médias dos históricos escolares de alunos do bacharelado argelino. Foram analisadas 43 imagens digitalizadas de históricos, abrangendo o período de 1997 a 2015. O processo inclui um detalhado tratamento automatizado das imagens, removendo ruídos para garantir a extração precisa das informações. Como resultado, o método alcançou uma alta taxa de acurácia, com 95,96% de precisão na identificação e extração das informações.	Kefali e Drabsia (20
10	SCO-GED-240806-0056	Fuzzy Model for Estimating the Probability of User Error in the Electronic Document Management System	O artigo apresenta o desenvolvimento de um modelo para estimar a probabilidade de erros cometidos por usuários ao utilizarem um sistema de Gestão Eletrônica de Documentos (SGED). Essa abordagem permite acumular conhecimento sobre os usuários, os erros mais comuns, identificar fraquezas e necessidades de melhoria. No conteúdo do trabalho é possível identificar que houve o processamento do modelo proposto, no entanto, não foi possível identificar a acurácia do resultado de pesquisa.	Fil, Nefed Binkovska (2020)

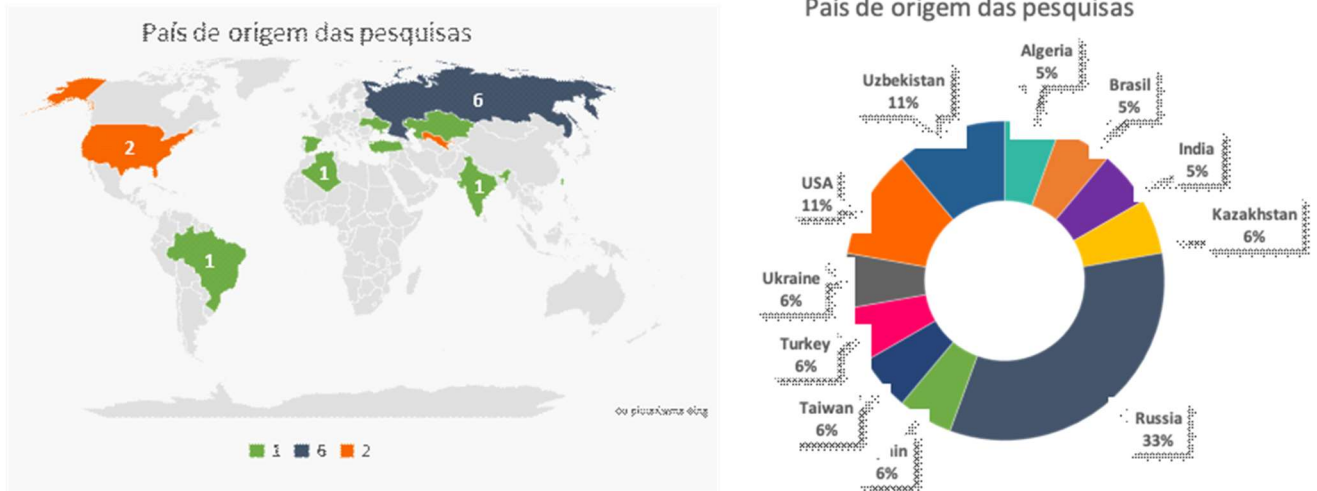
PESQUISAS ANALISADAS				
	Código	Título	Assunto	Citação
1 1	SCO- GED- 240806 -0060	Diagram recognition: Domain knowledge based approach	O artigo apresenta um método para reconhecimento de diagramas de documentos eletrônicos, chamado Recognition Domain Knowledge Kernel (RDKK). Com a aplicação de um algoritmo foi possível identificar símbolos e conexões de diagramas e registrar essas informações em um banco de dados. Apesar do artigo não especificar qual foi a amostragem nem o índice de acerto da aplicação, os autores consideram o resultado promissor e preveem a ampliação de seu uso para uma variedade maior de símbolos e diagramas, para reconhecimento de imagens de baixa qualidade ou reconhecimento de símbolos de diversas línguas.	Bashir e G (2013)
1 2	SCO- GED- 240806 -0075	Development of intelligent electronic document management system model based on machine learning methods	Num primeiro momento os autores realizam uma análise softwares de Gestão Eletrônica de Documentos (SGED), e identifica que não possuem características e funcionalidades de um sistema inteligente de gestão de documentos. Na sequência, um modelo teórico de sistema inteligente de GED é proposto. Para este modelo o SGED é dividido em subsistemas e são apresentados métodos para otimização dos subsistemas com uso de inteligência artificial, aprendizagem de máquina e algoritmos multiagentes. Os subsistemas devem realizar tarefas que envolvem automatização no cadastro de documentos, dos fluxos de trabalho, extração e análise de dados e recuperação inteligente de informações. Este modelo apresenta uma visão bem ampla e inovadora das diferentes funções que um SGED pode desempenhar, no entanto, é uma proposição teórica e exigirá pesquisas experimentais avançar em seu desenvolvimento.	Sambetba , <i>et al.</i> (20
1 3	SCO- GED- 240806 -0079	Development of the documents comparison module for an electronic document management system	O artigo tem como objetivo apresentar uma abordagem para comparação de documentos em um sistema de gestão eletrônica de documentos (SGED). O sistema realiza a comparação automática de diferentes versões de um documento para identificar alterações entre elas, como diferenças de conteúdo textual, alterações na formatação ou estrutura e atualizações em seções específicas (tabelas, parágrafos, cabeçalhos). Os autores consideram o resultado positivo, pois demonstram eficiência 98% de caracteres reconhecido, além de apresentar um bom desempenho no processamento dos algoritmos.	Mikheev e Yakimov (2019)
1 4	SCO- GED- 240806 -0088	Algorithm of adaptation of electronic document management system based on machine learning technology	O artigo tem como objetivo a automatização do processo de adaptação e personalização de um sistema de gestão eletrônica de documentos de acordo com as características individuais dos usuários. Para tanto mapeia um conjunto de critérios para calcular a adaptabilidade e desenvolve um algoritmo para implantação. O artigo apresenta as etapas de desenvolvimento e aplicação do algoritmo. Os autores consideram o resultado, alcançado por meio de uma simulação, positivo, mesmo não sendo possível visualizar a adaptação diretamente em um software.	Krasnyans <i>et al.</i> (202
1 5	SCO- GED- 240806 -0100	Designing an information system for the electronic document management of a university: Automatic classification of documents	O trabalho apresenta técnicas de classificação automática de documentos acadêmicos utilizando abordagem de Machine Learning (ML). Quadro métodos diferentes foram testados e comparados entre. Como conclusão são identificados os métodos mais adequados de extração de características do texto e o algoritmo de classificação, para resolver o problema de classificação automática de documentos universitários.	Tkachenko Denisova (2022)

PESQUISAS ANALISADAS				
	Código	Título	Assunto	Citação
1 6	SCO- GED- 240806 -0115	Document discourse for managing construction project documents	O artigo apresenta uma abordagem de classificação automática de documentos que utiliza as relações semânticas entre documentos do projeto como estrutura para classificar, organizar e gerenciar os documentos. Além das funções automatizadas, o sistema também oferece aos usuários a funcionalidade de definir, classificar e visualizar documentos relacionados através de seus vínculos semânticos e dependências existentes entre eles. Os autores concluem que o desempenho do classificador é afetado negativamente quando a informação é escassa, mas melhora significativamente quando mais informação é disponibilizada.	Al Qady e Kandil (20
1 7	SCO- GED- 240806 -0128	Concept relation extraction from construction documents using natural language processing	O artigo apresenta uma ferramenta para extração automática/semiautomática de conjunto de conceitos, por meio da análise semântica de documentos contratuais usando uma ferramenta de processamento de linguagem natural (PLN). Para análise da eficiência, foi realizado teste para extração de conceitos com humanos avaliadores, utilizando uma amostragem do texto e realizada uma comparação dos resultados médios dos resultados dos humanos avaliadores com os resultados da ferramenta. Concluiu-se que, em termos de precisão e recall, a ferramenta alcançou aproximadamente 90% do desempenho médio dos humanos avaliadores.	Al Qady e Kandil (20
1 8	WOS- GED- 240806 -0028	Decision-Making Support System for tasks distribution in Personal Data Operators Register Maintaining based on a Fuzzy- Production Model	O trabalho apresenta uma solução para automação da tomada de decisão na distribuição de tarefas para operadores em um Sistema de Gestão Eletrônica de Documentos (SGED). Utilizando a lógica fuzzy como método de inferência para estimar probabilidades, os autores demonstraram que o tempo necessário para a tomada de decisão por especialistas pode ser reduzido em cinco vezes. Essa abordagem aumenta a rapidez de execução em 80% do tempo gasto por especialistas, que permite otimização dos processos e redução da carga intelectual dos especialistas, que poderá dedicar-se a ações mais estratégicas.	Ismagilov, al. (2019)

Fonte: Elaborado pela autora (2024)

Esta relação oferece uma visão geral sobre o conteúdo de cada estudo, permitindo identificar os principais assuntos tratados, compreender suas abordagens e acessar informações relevantes sobre seus resultados.

Estas pesquisas se originam de 11 países, sendo eles: Argélia, Brasil, Índia, Cazaquistão, Rússia, Espanha, Taiwan, Turquia, Ucrânia, Estados Unidos e Uzbequistão. Como pode ser observado na **Figura 12**, o país com maior número de artigos publicados nesta área é a Rússia, com 6 publicações, representando 33%, seguida por Estados Unidos e Uzbequistão, ambos com 2 trabalhos, representando 11% das publicações. Os outros países, incluindo o Brasil apresentam apenas uma publicação cada, representando 6% da fatia.

Figura 12 – Países de Origem das pesquisas

Fonte: Elaborado pela autora (2024)

Como era de se esperar, a universidade que mais produziu trabalhos foi a *Tambov State Technical University*, na Rússia, com três publicações, seguida pela *Purdue University*, nos Estados Unidos. As publicações destacaram colaborações tanto entre universidades quanto entre instituições públicas e privadas.

Um exemplo significativo é a parceria entre a *University of Zaragoza* e a empresa *InSynergy Consulting S.A.*, especializada em tecnologia e inovação. Essa colaboração resultou no desenvolvimento de um subsistema baseado em Processamento de Linguagem Natural, que aprimorou a extração de dados em contratos, aumentando o desempenho em 77% e reduzindo significativamente o uso de CPU e memória do servidor. Com essa colaboração foi possível ter acesso a um *corpus* documental que não teria se ela não existisse.

Quadro 15 – Universidades produtoras das pesquisas

Universidades		
Universidade/Instituição de origem dos autores	Qtd. Artigos	Referências
<i>Badji Mokhtar University</i>	1	Kefali e Drabsia (2018)
<i>Image Processing Systems Institute of RAS</i>	1	Mikheev e Yakimov (2019)
<i>InSynergy Consulting S.A</i>	1	Ruiz, Román, Garrido e Mena (2020)
<i>Islamic University of Science and technology</i>	1	Bashir e Giri (2013)

Universidades		
Universidade/Instituição de origem dos autores	Qtd. Artigos	Referência
<i>Izmir Katip Çelebi University</i>	1	Onan, Atik e Yalçın (2021)
<i>Kazan Federal University</i>	1	Ismagilov, <i>et al.</i> (2019)
<i>Kharkiv National Automobile and Highway University</i>	1	Fil, Nefedov e Binkovskaya (2020)
<i>L. N. Gumilyov Eurasian National University</i>	1	Sambetbayeva, <i>et al.</i> (2022)
<i>Ministry of Internal affairs of the Republic of Uzbekistan</i>	1	Fayziev <i>et al.</i> (2021)
<i>National Chiao Tung University</i>	1	Liu, Hsaio, Lee e Chen (2013)
<i>Omsk State Technical University</i>	1	Tkachenko e Denisova (2022)
<i>Pontifícia Universidade Católica do Rio de Janeiro</i>	1	Gomes e Moraes Filho (2011)
<i>Purdue University</i>	2	Al Qady e Kandil (2013); Al Qady e Kandil (2010)
<i>Samara National Research University</i>	1	Mikheev e Yakimov (2019)
<i>Samarkand State University</i>	1	Jumanov e Karshiev (2020)
<i>Satbayev University</i>	1	Sambetbayeva, <i>et al.</i> (2022)
<i>Siberian State Automobile and Highway University</i>	1	Tkachenko e Denisova (2022)
<i>Tambov State Technical University</i>	3	Obukhov, Krasnyanskiy e Nikolyukin (2019); Krasnyanskiy, Obukhov e Solomatina (2019); Krasnyanskiy; <i>et al.</i> (2020)
<i>Tashkent State Technical University</i>	1	Fayziev <i>et al.</i> (2021)
<i>Universidade Cândido Mendes</i>	1	Gomes e Moraes Filho (2011)
<i>University of Zaragoza</i>	1	Ruiz, <i>et al.</i> (2020)

Fonte: Elaborado pela autora (2024)

A colaboração entre universidades, bem como entre universidades e organizações públicas e privadas, é fundamental para o avanço das pesquisas e para o compartilhamento de conhecimento. Essas parcerias promovem a integração de diferentes perspectivas, recursos e expertises, ampliando o impacto das descobertas

científicas e tecnológicas. Além disso, elas viabilizam a aplicação prática dos resultados de pesquisa, fomentando a inovação, o desenvolvimento de soluções customizadas e a criação de tecnologias mais eficientes e adaptadas às necessidades do mercado e da sociedade.

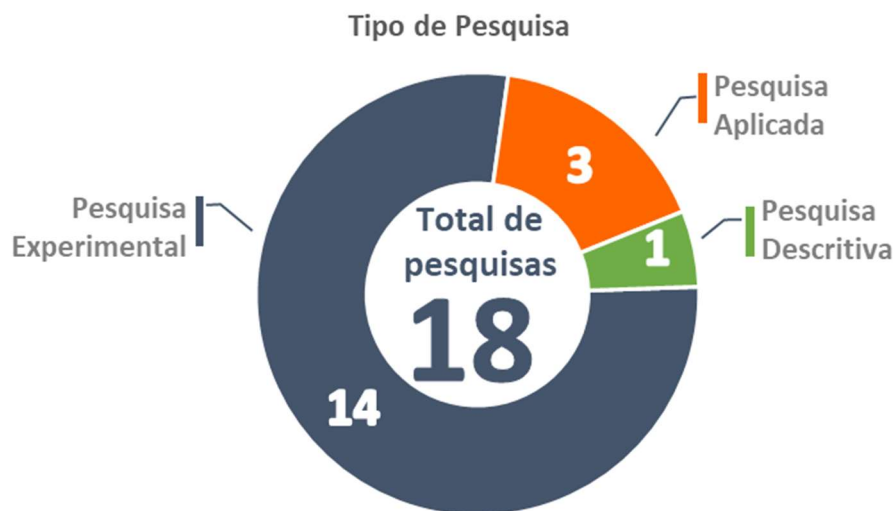
As pesquisas puderam ser classificadas também quanto à abordagem adotada para alcançar seus objetivos. As abordagens identificadas foram:

Pesquisa Experimental: Esse tipo de pesquisa envolve a realização de experimentos controlados para testar hipóteses sobre o comportamento de um sistema, função ou algoritmo. O pesquisador submete seus objetos à influência de algumas variáveis e observa os resultados, muitas vezes comparando o desempenho entre diferentes condições ou abordagens (Gil, 2008; Lakatos; Musgrave, 1970).

Pesquisa bibliográfica: Baseia-se em fontes como livros e artigos científicos utilizados para obter informações e embasar o estudo (Gil, 2008). Segundo Minayo (2010, p. 184) a pesquisa bibliográfica traz clareza e compreensão sobre a realidade estudada, e deve apresentar minimamente as principais pesquisas clássicas e pesquisas mais atuais mais relevantes, para que seja abrangente suficiente para embasar as discussões pretendidas.

Pesquisa Aplicada: tem como foco problemas concretos e apresentar soluções práticas ou criar inovações. Segundo Minayo (2010), tem como finalidade a ação e geralmente, conecta conhecimentos teóricos a contextos específicos, como o desenvolvimento de tecnologias.

A partir da análise representada na **Figura 13**, percebe-se uma predominância de pesquisas experimentais na área de Gestão Eletrônica de Documentos (GED).

Figura 13 – Abordagem de pesquisa

Fonte: Elaborado pela autora (2024)

Dos 18 artigos analisados 14 são classificadas como Pesquisa Experimental e três como aplicada. Dos trabalhos analisados 10 deles apresentaram ferramentas, softwares ou algoritmos desenvolvidos pelos autores. Alguns desses trabalhos são apenas simulações e não tiveram como resultado o desenvolvimento de uma ferramenta em si, mas apresentam resultados satisfatórios que justifica desenvolver uma ferramenta. A seguir são listas algumas dessas ferramentas e suas funções.

Os autores Gomes e Moraes Filho (2011) apresentaram o CADWeb, um software, em caráter experimental para categorização (classificação) automática de documentos. O software foi utilizado para classificar 32 documentos em PDF de uma biblioteca digital universitária e obteve um percentual de 84,6% de acerto.

Os autores Liu, Hsaio, Lee e Chen (2013) apresentaram o *Constrained-PLSA*, um algoritmo para classificação de documentos com e sem rótulos, foi aplicado a resumos de artigos, cujos rótulos foram fornecidos pelos próprios usuários e os resultados alcançados superam outras abordagens utilizadas como comparação.

Os autores Bashir e Giri (2013) desenvolveram um algoritmo para reconhecimento de diagramas de documentos eletrônicos, chamado *Recognition Domain Knowledge Kernel* (RDKK), com isso foi possível identificar símbolos e conexões de diagramas (documentos técnicos de engenharia) e registrar essas informações em um banco de dados. Esse trabalho pode auxiliar na compreensão, por máquina, de diagramas que hoje só são compreendidos por humanos.

Vale ressaltar que muitas dessas pesquisas utilizaram em suas amostras um volume relativamente pequeno de documentos e poucos deles apresentaram o uso diretamente em um GED. Esses dados nos fazem questionar a maturidade dessas técnicas e seu potencial de aplicação em sistemas de gestão eletrônica de documentos.

O maior volume de documentos processados foi apresentado pelos autores Gomes e Moraes Filho (2011). Eles processaram 3.394 documentos online, no entanto, todos eles eram resumos de artigos com anotações realizadas pelos usuários. Apesar do volume não ser pequeno, no entanto o “resumo do artigo é de tamanho limitado” (Liu; Hsaio; Lee; Chen, 2013, p. 597, tradução nossa), conforme esclarecido pelos próprios autores.

7. DISCUSSÃO DOS RESULTADOS

As discussões incluem uma análise crítica e detalhada dos dados levantados, comparando-os com as evidências da literatura existente.

Além disso, procura-se destacar padrões identificados, desafios, tendências, lacunas e oportunidades relacionadas ao uso da Inteligência Artificial e Ciência de Dados no contexto da Gestão Eletrônica de Documentos. Por meio desta abordagem, busca-se fornecer interpretações claras e embasadas, oferecendo contribuições teóricas e práticas para o avanço da área.

Um primeiro ponto que despertou curiosidade é que na análise de 18 artigos científicos, não há indicação de técnicas de Ciência de Dados. Mesmo o trabalho apresentado por Mikheev e Yakimov (2019), que possui o termo *Data Science* descrita no campo de palavra-chave, não há indicativo de associação com esta área.

haja menção explícita às abordagens de CD nas aplicações desenvolvidas para a GED, todos os trabalhos experimentais utilizaram suas técnicas na etapa de pré-processamento de dados.

A etapa de pré-processamento é essencial para garantia da qualidade dos resultados das abordagens de inteligência artificial e aprendizagem de máquina, segundo Krasnyanskiy, Obukhov e Solomatina (2019, p.4, tradução nossa), esta etapa “permite melhorar significativamente a precisão dos resultados obtidos na aplicação do aprendizado de máquina.”

Embora não haja menção explícita às abordagens de CD nas aplicações desenvolvidas para GED, todas as técnicas de pré-processamento de dados utilizadas refletem os fundamentos dessa área. Isso pode ser interpretado de duas formas:

Por um lado, sugere uma possível subvalorização ou desconhecimento do papel central que essas técnicas desempenham na preparação e manipulação de dados, reforçando a necessidade de maior integração terminológica e conceitual entre as áreas. Isso aponta para uma oportunidade de reconhecer formalmente a contribuição da Ciência de Dados nos avanços da GED, além de promover maior interdisciplinaridade.

Por outro lado, poderia indicar que os pesquisadores já internalizaram os conceitos e métodos da CD a tal ponto que os aplicam naturalmente, sem a

necessidade de destacá-los explicitamente. Essa integração implícita sugere maturidade no uso das técnicas e reforça como a CD é fundamental para o avanço da GED, mesmo que sua contribuição não seja diretamente reconhecida no discurso acadêmico.

Para organizar a análise dos artigos selecionados sobre o uso de Inteligência Artificial e Ciência de Dados na Gestão Eletrônica de Documentos, e em resposta às questões de pesquisa, as informações extraídas dos artigos foram categorizadas e classificadas como resultado da análise temática e de conteúdo.

Esta abordagem busca compreender de forma abrangente como essas tecnologias estão sendo aplicadas, as etapas envolvidas na execução dos projetos, os benefícios observados, e os desafios enfrentados.

Essa categorização fornece um panorama estruturado, permitindo identificar tendências e lacunas no uso dessas tecnologias na GED, ampliando o conhecimento da Ciência da Informação (CI) em relação às novas tecnologias.

7.1. P1 - Pontos de interseção entre GED e IA

Na primeira questão de pesquisa procurou-se entender quais são os principais pontos de interseção entre Gestão Eletrônica de Documentos, Inteligência Artificial e Ciência de Dados.

Em resposta a esta pergunta, identificou-se que os principais pontos de interseção entre Gestão Eletrônica de Documentos e Inteligência Artificial são justamente as tecnologias, desenvolvidas para automatizar processos, melhorar a agilidade e eficiência das ações e proporcionar maior compreensão das informações relevantes nos documentos. Isso resulta em uma visão mais estratégica para gestores, profissionais da informação e usuários finais.

Além disso, a integração de IA no GED permite análises preditivas, identificação de padrões, aprimoramento da segurança da informação, otimização no acesso e recuperação de dados, uso de OCR e mineração de texto, e uma tomada de decisão mais fundamentada e rápida.

com base nos temas centrais dos trabalhos analisados, realizou-se a categorização das tecnologias enquanto aplicação prática. Os trabalhos podem ser enquadrados em uma ou mais categorias, dependendo da diversidade de suas

abordagens. Foram identificadas 16 categorias que estão distribuídas conforme apresentado na **Figura 14**.

Classificação de documentos é o tópico com mais pesquisas, com 7 artigos, representa 24% das pesquisas e Identificação de similaridade entre documento está em segundo lugar com 3 artigos.

Figura 14 – Aplicações Práticas



Fonte: Elaborado pela autora (2024)

O principal objetivo dessa classificação é transformar os dados coletados em descrições claras e objetivas de aplicações viáveis, promovendo um entendimento acessível e direcionado sobre as contribuições práticas e o potencial impacto das pesquisas analisadas.

A seguir, são apresentadas as categorias identificadas, acompanhadas de uma descrição detalhada para facilitar a compreensão, bem como a citação dos autores relacionados a cada categoria.

Classificação de documentos: Os documentos são categorizados automaticamente com base em suas características, como tipo de conteúdo, formato, termos-chave e metadados. Esse processo elimina a necessidade de classificação manual, reduzindo o tempo gasto e promovendo maior consistência e precisão na

organização de grandes volumes de informações. **Referências:** Gomes e Moraes Filho (2011); Liu, Hsaio, Lee e Chen (2013); Krasnyanskiy, Obukhov e Solomatina (2019); Onan, Atik e Yalçın (2021); Sambetbayeva, *et al.* (2022); Tkachenko e Denisova (2022); Al Qady e Kandil (2013).

Identificação de similaridade entre documentos: Análise e comparação de desenhos técnicos e outros tipos de documentação, identificando automaticamente documentos similares. Isso é especialmente útil para evitar a duplicação de arquivos e agilizar a recuperação de informações, permitindo que o usuário encontre versões anteriores ou relacionadas a um projeto, otimizando o processo de busca de documentos. **Referências:** Obukhov, Krasnyanskiy e Nikolyyukin (2019); Sambetbayeva, *et al.* (2022); Mikheev e Yakimov (2019).

Roteamento Inteligente de Documentos: O roteamento inteligente utiliza algoritmos que identificam automaticamente a melhor rota para a aprovação, revisão ou execução de um documento dentro de um fluxo de trabalho. Esse sistema analisa as características do documento, como o tipo, conteúdo, e destinatários esperados, e direciona o documento para os responsáveis apropriados, otimizando o processo de aprovação. **Referências:** Sambetbayeva, *et al.* (2022); Ismagilov, *et al.* (2019).

Classificação semântica de documentos: Organização de documentos por meio de análise semântica de conteúdo, considerando as conexões e relações entre eles. Esse método permite classificar documentos com base no contexto e nas dependências dentro do conjunto de arquivos. **Referências:** Al Qady e Kandil (2013); Al Qady e Kandil (2010).

Extração de informações de documentos: Esta aplicação torna-se vantajosa especialmente para empresas que lidam com grandes volumes de documentos em formatos físico e digital. Para lidar com os desafios de acessar e utilizar informações contidas em documentos arquivados, como faturas, contratos etc., esta aplicação permite a identificação e extração automatizada de informações essenciais de documentos com formatos padronizados ou similares. **Referências:** Kefali e Drabsia (2018); Bashir e Giri (2013).

Identificação de Documentos Relacionados: Ao interpretar relações semânticas, o sistema melhora a busca de documentos além das palavras-chave, identificando conteúdos por temas ou contexto e relações semânticas, como contratos, anexos, e comunicações associadas. O sistema identifica automaticamente

documentos que precisam ser vistos em conjunto. **Referências:** Sambetbayeva, *et al.* (2022); Al Qady e Kandil (2010).

Indexação automática de documentos e fluxos de trabalho: Automatiza a extração de informações visuais, como símbolos e conexões, agilizando a análise e organização de documentos com diagramas complexos, permitindo que a compreensão por máquina. **Referências:** Bashir e Giri (2013); Sambetbayeva, *et al.* (2022).

Adaptação e personalização da interface: Essa funcionalidade permite que o sistema se adapte ao usuário segundo seu histórico, características e padrões de atividades. Beneficiando diretamente o usuário final com um sistema mais adaptado às suas necessidades, aumentando sua aceitação na utilização do sistema. **Referências:** Krasnyanskiy; *et al.* (2020).

Aprimorar a Experiência do Usuário: Este modelo permite acumular informações sobre os erros mais frequentes dos usuários, identificar fragilidades e propor melhorias nos Sistemas de Gestão Eletrônica de Documentos (SGED), beneficiando diretamente o usuário final com um sistema mais adaptado às suas necessidades. **Referências:** Fil, Nefedov e Binkovskaya (2020).

Autenticação biométrica: Garantia de autenticidade e integridade na realização de acessos a sistemas de gestão eletrônica de documentos (GED), bem como nas ações realizadas ao longo de todo o ciclo de vida documental e dos fluxos de trabalho. A autenticação biométrica combinada com a geração de chaves criptográficas protege contra acessos não autorizados, evita alterações indevidas, e assegura que todas as operações são executadas apenas por usuários devidamente autenticados. **Referências:** Fayziev *et al.* (2021).

Comparação de versões documentos: Análise e comparação automática de diferentes versões de um documento para identificar alterações entre elas, como diferenças de conteúdo textual (adição, exclusão ou modificação de frases, palavras ou caracteres); alterações na formatação ou estrutura e atualizações em seções específicas (tabelas, parágrafos, cabeçalhos). **Referências:** Mikheev e Yakimov (2019).

Construção de Ontologias: A partir de análises semânticas é possível realizar a criação de ontologias, representando relações e temas principais de documentos, útil para organizar conteúdos complexos ou treinamentos. **Referências:** Al Qady e Kandil (2010).

Correção Ortográfica e Semântica: Correção Ortográfica e Semântica: A detecção e correção de erros em documentos eletrônicos tornam-se mais eficazes e precisas com a aplicação de correções automáticas. Essa funcionalidade melhora a qualidade das informações em Sistemas de Gestão Eletrônica de Documentos (SGED) nos textos, metadados, fluxos de trabalho e processos documentais. **Referências:** Jumanov e Karshiev (2020).

Otimização de Desempenho: A utilização de uma arquitetura baseada em microsserviços e message brokers permite uma melhoria significativa do desempenho de subsistemas de PLN. A arquitetura permite comunicação assíncrona e paralela entre máquinas, reduzindo a demanda por recursos de CPU e memória e tornando os sistemas mais leves e escaláveis. **Referências:** Ruiz, *et al.* (2020).

Registro Automático de Documentos: O sistema realiza o registro de novos documentos de forma automática, preenchendo campos de metadados como data, autor, tipo de documento e contexto, sem a necessidade de intervenção manual. **Autores relacionados:** Sambetbayeva, *et al.* (2022).

Suporte à tomada de decisões: Com base em padrões de uso e histórico de operações do usuário, o subsistema pode sugerir as próximas etapas do fluxo de trabalho. Por exemplo, ao analisar um desenho técnico, o sistema pode recomendar as ações mais prováveis, como revisões, aprovações ou ajustes, baseadas em operações similares já realizadas. Isso não só acelera o processo, mas também reduz a chance de erros humanos ao automatizar etapas críticas da decisão. **Referências:** Obukhov, Krasnyanskiy e Nikolyukin (2019).

Com essa classificação é possível ter um panorama dos principais desenvolvimentos de tecnologias de IA e CD na área de Gestão Eletrônica de Documentos.

Para cada aplicação prática, foram identificadas as abordagens empregadas, abrangendo ferramentas, técnicas, métodos, modelos, algoritmos, arquiteturas, frameworks e bibliotecas, entre outros recursos. Essa identificação permite compreender não apenas as tecnologias utilizadas, mas também como elas se inter-relacionam para resolver problemas específicos. Além disso, esse mapeamento facilita a identificação de lacunas e oportunidades para novas contribuições no campo da Gestão Eletrônica de Documentos e áreas correlatas.

Quadro 16 – Classificação quanto aos benefícios

Abordagens técnicas mapeadas	
Abordagem	Referência
Adaptação e personalização da interface	
<i>Artificial Neural Networks (ANN)</i>	Krasnyanskiy; <i>et al.</i> (2020); Fayziev <i>et al.</i> (2021); Krasnyanskiy, Obukhov e Solomatina (2019); Obukhov, Krasnyanskiy e Nikolyyukin (2019)
Aprimorar a Experiência do Usuário	
<i>Fuzzy Logic</i>	Fil, Nefedov e Binkovskaya (2020); Ismagilov, <i>et al.</i> (2019); Obukhov, Krasnyanskiy e Nikolyyukin (2019);
<i>Fuzzy Logic Toolbox</i>	Fil, Nefedov e Binkovskaya (2020)
<i>MathLab</i>	Fil, Nefedov e Binkovskaya (2020)
<i>Multiple Input Single Output (MISO)</i>	Fil, Nefedov e Binkovskaya (2020)
Autenticação biométrica	
<i>Artificial Neural Networks (ANN)</i>	Krasnyanskiy; <i>et al.</i> (2020); Fayziev <i>et al.</i> (2021); Krasnyanskiy, Obukhov e Solomatina (2019); Obukhov, Krasnyanskiy e Nikolyyukin (2019)
<i>Bayes-Hemming network</i>	Fayziev <i>et al.</i> (2021)
<i>Bayes-Pearson</i>	Fayziev <i>et al.</i> (2021)
<i>Biometric Authentication</i>	Fayziev <i>et al.</i> (2021)
<i>Bose–Chaudhuri–Hocquenghem code (BCH code)</i>	Fayziev <i>et al.</i> (2021)
<i>Converters Biometrics-Code (CBC)</i>	Fayziev <i>et al.</i> (2021)
<i>Error-correcting codes</i>	Fayziev <i>et al.</i> (2021)
<i>Face recognition</i>	Fayziev <i>et al.</i> (2021)
<i>Fingerprint Recognition</i>	Fayziev <i>et al.</i> (2021)
<i>Fuzzy extractors</i>	Fayziev <i>et al.</i> (2021)
<i>Geometry hand form Recognition</i>	Fayziev <i>et al.</i> (2021)
<i>Graphological recognition</i>	Fayziev <i>et al.</i> (2021)
<i>Hadamard</i>	Fayziev <i>et al.</i> (2021)
<i>Iris recognition</i>	Fayziev <i>et al.</i> (2021)
<i>Multidimensional Bayesian functionals</i>	Fayziev <i>et al.</i> (2021)
<i>Pearson metrics</i>	Fayziev <i>et al.</i> (2021)
<i>Perceptrons</i>	Fayziev <i>et al.</i> (2021)

Abordagens técnicas mapeadas	
Abordagem	Referência
<i>Quadratic forms</i>	Fayziev <i>et al.</i> (2021)
<i>Self-correcting code</i>	Fayziev <i>et al.</i> (2021)
<i>Voice Recognition</i>	Fayziev <i>et al.</i> (2021)
Classificação de documentos	
<i>1-gram model (1-g model)</i>	Onan, Atik e Yalçın (2021)
<i>Artificial Neural Networks (ANN)</i>	Krasnyanskiy; <i>et al.</i> (2020); Fayziev <i>et al.</i> (2021); Krasnyanskiy, Obukhov e Solomatina (2019); Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Bag of Words (BoW)</i>	Gomes e Moraes Filho (2011); Liu, Hsaio, Lee e Chen (2013); Tkachenko e Denisova (2022)
<i>Bernoulli Naive Bayes</i>	Krasnyanskiy, Obukhov e Solomatina (2019)
<i>Case Folding</i>	Gomes e Moraes Filho (2011); Tkachenko e Denisova (2022)
<i>Constrained Optimization Problem KMeans (COP-KMeans)</i>	Liu, Hsaio, Lee e Chen (2013)
<i>Constrained-Kmeans</i>	Liu, Hsaio, Lee e Chen (2013)
<i>Data Extraction</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Kefali e Drabsia (2018)
<i>Decision Tree Algorithm (DTA)</i>	Onan, Atik e Yalçın (2021)
<i>Decision Trees method</i>	Krasnyanskiy, Obukhov e Solomatina (2019)
<i>HashMap</i>	Gomes e Moraes Filho (2011)
<i>Indexing os documents</i>	Krasnyanskiy, Obukhov e Solomatina (2019)
<i>Information extraction</i>	Sambetbayeva, <i>et al.</i> (2022)
<i>K nearest neighbors</i>	Krasnyanskiy, Obukhov e Solomatina (2019)
<i>Kernel functions</i>	Tkachenko e Denisova (2022)
<i>K-fold cross validation</i>	Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022)
<i>Kmeans</i>	Liu, Hsaio, Lee e Chen (2013)
<i>K-Nearest Neighbour Algorithm (KNN)</i>	Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022)
<i>Latent Dirichlet Allocation (LDA)</i>	Liu, Hsaio, Lee e Chen (2013)
<i>Latent Semantic Analysis (LSA)</i>	Al Qady e Kandil (2013)
<i>Lemmatization</i>	Krasnyanskiy, Obukhov e Solomatina (2019)

Abordagens técnicas mapeadas	
Abordagem	Referência
<i>Logistic regression</i>	Krasnyanskiy, Obukhov e Solomatina (2019)
<i>Machine Learning (LM)</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Tkachenko e Denisova (2022); Sambetbayeva, et al. (2022); Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>machine learning algorithms</i>	Sambetbayeva, et al. (2022)
<i>multi-agent approach</i>	Sambetbayeva, et al. (2022)
<i>Multinomial Mixture Latent Dirichlet Allocation (MM-LDA)</i>	Liu, Hsaio, Lee e Chen (2013)
<i>Naive Bayes (BN)</i>	Onan, Atik e Yalçın (2021)
<i>Noise removal</i>	Tkachenko e Denisova (2022); Mikheev e Yakimov (2019)
<i>PLSA Clustering (Probabilistic Latent Semantic Analysis)</i>	Liu, Hsaio, Lee e Chen (2013)
<i>Polynomial Naive Bayes</i>	Krasnyanskiy, Obukhov e Solomatina (2019)
<i>Probabilistic Latent Semantic Analysis (PLSA)</i>	Liu, Hsaio, Lee e Chen (2013)
<i>radial basis function</i>	Tkachenko e Denisova (2022)
<i>Random Forest Algorithm (RF)</i>	Onan, Atik e Yalçın (2021)
<i>Recuction of signs os space size</i>	Krasnyanskiy, Obukhov e Solomatina (2019)
<i>Resolving ambiguity</i>	Sambetbayeva, et al. (2022)
<i>Semi-Supervised Clustering</i>	Liu, Hsaio, Lee e Chen (2013)
<i>Stemming</i>	Gomes e Moraes Filho (2011); Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022)
<i>Stopwords elimination</i>	Gomes e Moraes Filho (2011); Krasnyanskiy, Obukhov e Solomatina (2019); Onan, Atik e Yalçın (2021)
<i>Support Vector Machines (SVM)</i>	Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022)
<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022); Al Qady e Kandil (2013)
<i>Text analysis</i>	Gomes e Moraes Filho (2011); Sambetbayeva, et al. (2022)
<i>Text Classification</i>	Gomes e Moraes Filho (2011)
<i>Text Mining</i>	Gomes e Moraes Filho (2011)

Abordagens técnicas mapeadas	
Abordagem	Referência
<i>Thresholding</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Mikheev e Yakimov (2019); Kefali e Drabsia (2018); Bashir e Giri (2013); Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022)
<i>Tokenization</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Mikheev e Yakimov (2019); Kefali e Drabsia (2018); Bashir e Giri (2013); Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022)
<i>Vector Space Model</i>	Al Qady e Kandil (2013)
<i>Waikato Environment for Knowledge Analysis (WEKA)</i>	Onan, Atik e Yalçın (2021)
<i>Zemberek library</i>	Onan, Atik e Yalçın (2021)
Classificação semântica de documentos	
<i>Latent Semantic Analysis (LSA)</i>	Al Qady e Kandil (2013)
<i>Shallow parsing ou partial parsers</i>	Al Qady e Kandil (2010)
<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022); Al Qady e Kandil (2013)
<i>Vector Space Model</i>	Al Qady e Kandil (2013)
Comparação de versões documentos	
<i>Long Short-Term Memory (LSTM)</i>	Mikheev e Yakimov (2019); Obukhov, Krasnyanskiy e Nikoluykin (2019)
<i>Noise removal</i>	Tkachenko e Denisova (2022); Mikheev e Yakimov (2019)
<i>Optical Character Recognition (OCR)</i>	Mikheev e Yakimov (2019)
<i>Recurrent neural network (RNN)</i>	Mikheev e Yakimov (2019)
<i>Tesseract engine</i>	Mikheev e Yakimov (2019)
<i>Thresholding</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Mikheev e Yakimov (2019); Kefali e Drabsia (2018); Bashir e Giri (2013); Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022)
Correção Ortográfica e Semântica	
<i>multilevel morphological analysis</i>	Jumanov e Karshiev (2020)
<i>N-gram</i>	Jumanov e Karshiev (2020)
<i>Vector Representation</i>	Jumanov e Karshiev (2020)
Extração de informações de documentos	

Abordagens técnicas mapeadas	
Abordagem	Referência
<i>Automatic Diagram Recognition</i>	Bashir e Giri (2013)
<i>Connected Components Labeling</i>	Kefali e Drabsia (2018)
<i>Data Extraction</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Kefali e Drabsia (2018)
<i>Frame Detection</i>	Kefali e Drabsia (2018)
<i>Grayscale Transformation</i>	Kefali e Drabsia (2018)
<i>Linguagem de Programação C</i>	Bashir e Giri (2013)
<i>logical smoothing approach</i>	Bashir e Giri (2013)
<i>Marginal Noise Elimination</i>	Kefali e Drabsia (2018)
<i>Recognition Domain Knowledge Kernel (RDKK)</i>	Bashir e Giri (2013)
<i>Region Code Technique</i>	Bashir e Giri (2013)
<i>Run Length Smoothing Algorithm (RLSA)</i>	Kefali e Drabsia (2018)
<i>Segmentation Technique</i>	Kefali e Drabsia (2018)
<i>size, skew and slant normalization</i>	Bashir e Giri (2013)
<i>Skew Correction</i>	Kefali e Drabsia (2018)
<i>Smoothing</i>	Kefali e Drabsia (2018)
<i>Tables Localization</i>	Kefali e Drabsia (2018)
<i>Tables Segmentation Into Columns</i>	Kefali e Drabsia (2018)
<i>Text Localization</i>	Kefali e Drabsia (2018)
<i>Text Preprocessing</i>	Kefali e Drabsia (2018)
<i>Text Summarization</i>	Kefali e Drabsia (2018)
<i>Thresholding</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Mikheev e Yakimov (2019); Kefali e Drabsia (2018); Bashir e Giri (2013); Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022)
Identificação de Documentos Relacionados	
<i>Information extraction</i>	Sambetbayeva, et al. (2022)
<i>Machine Learning (LM)</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Tkachenko e Denisova (2022); Sambetbayeva, et al. (2022); Obukhov, Krasnyanskiy e Nikol'yukin (2019)
<i>machine learning algorithms</i>	Sambetbayeva, et al. (2022)

Abordagens técnicas mapeadas	
Abordagem	Referência
<i>multi-agent approach</i>	Sambetbayeva, <i>et al.</i> (2022)
<i>Resolving ambiguity</i>	Sambetbayeva, <i>et al.</i> (2022)
<i>Shallow parsing ou partial parsers</i>	Al Qady e Kandil (2010)
<i>Text analysis</i>	Gomes e Moraes Filho (2011); Sambetbayeva, <i>et al.</i> (2022)
Identificação de similaridade entre documentos	
<i>Artificial Neural Networks (ANN)</i>	Krasnyanskiy; <i>et al.</i> (2020); Fayziev <i>et al.</i> (2021); Krasnyanskiy, Obukhov e Solomatina (2019); Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Convolutional Neural Network (CNN)</i>	Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Decision Support Subsystem (DSS)</i>	Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Expert Systems (ES)</i>	Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Fuzzy Logic</i>	Fil, Nefedov e Binkovskaya (2020); Ismagilov, <i>et al.</i> (2019); Obukhov, Krasnyanskiy e Nikolyukin (2019);
<i>Gated Recurrent Unit (GRU)</i>	Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Inductive and Deductive Methods</i>	Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Information extraction</i>	Sambetbayeva, <i>et al.</i> (2022)
<i>Knowledge Processing Algorithm</i>	Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Long Short-Term Memory (LSTM)</i>	Mikheev e Yakimov (2019); Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Machine Learning (LM)</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Tkachenko e Denisova (2022); Sambetbayeva, <i>et al.</i> (2022); Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>machine learning algorithms</i>	Sambetbayeva, <i>et al.</i> (2022)
<i>multi-agent approach</i>	Sambetbayeva, <i>et al.</i> (2022)
<i>Noise removal</i>	Tkachenko e Denisova (2022); Mikheev e Yakimov (2019)
<i>Optical Character Recognition (OCR)</i>	Mikheev e Yakimov (2019)
<i>Recurrent neural network (RNN)</i>	Mikheev e Yakimov (2019)
<i>Resolving ambiguity</i>	Sambetbayeva, <i>et al.</i> (2022)
<i>Semantic Networks</i>	Obukhov, Krasnyanskiy e Nikolyukin (2019)

Abordagens técnicas mapeadas	
Abordagem	Referência
<i>Tesseract engine</i>	Mikheev e Yakimov (2019)
<i>Text analysis</i>	Gomes e Moraes Filho (2011); Sambetbayeva, et al. (2022)
<i>Thresholding</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Mikheev e Yakimov (2019); Kefali e Drabsia (2018); Bashir e Giri (2013); Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022)
Indexação automática de documentos e fluxos de trabalho	
<i>Automatic Diagram Recognition</i>	Bashir e Giri (2013)
<i>Information extraction</i>	Sambetbayeva, et al. (2022)
<i>Linguagem de Programação C</i>	Bashir e Giri (2013)
<i>logical smoothing approach</i>	Bashir e Giri (2013)
<i>Machine Learning (LM)</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Tkachenko e Denisova (2022); Sambetbayeva, et al. (2022); Obukhov, Krasnyanskiy e Nikolyyukin (2019)
<i>machine learning algorithms</i>	Sambetbayeva, et al. (2022)
<i>multi-agent approach</i>	Sambetbayeva, et al. (2022)
<i>Recognition Domain Knowledge Kernel (RDKK)</i>	Bashir e Giri (2013)
<i>Region Code Technique</i>	Bashir e Giri (2013)
<i>Resolving ambiguity</i>	Sambetbayeva, et al. (2022)
<i>size, skew and slant normalization</i>	Bashir e Giri (2013)
<i>Text analysis</i>	Gomes e Moraes Filho (2011); Sambetbayeva, et al. (2022)
<i>Thresholding</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Mikheev e Yakimov (2019); Kefali e Drabsia (2018); Bashir e Giri (2013); Onan, Atik e Yalçın (2021); Tkachenko e Denisova (2022)
Otimização de Desempenho	
<i>Message brokers</i>	Ruiz, et al. (2020)
<i>Message Oriented Middleware (MOM)</i>	Ruiz, et al. (2020)
<i>Round-Robin method</i>	Ruiz, et al. (2020)
Registro Automático de Documentos	
<i>Information extraction</i>	Sambetbayeva, et al. (2022)

Abordagens técnicas mapeadas	
Abordagem	Referência
<i>Machine Learning (LM)</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Tkachenko e Denisova (2022); Sambetbayeva, et al. (2022); Obukhov, Krasnyanskiy e Nikolyyukin (2019)
<i>machine learning algorithms</i>	Sambetbayeva, et al. (2022)
<i>multi-agent approach</i>	Sambetbayeva, et al. (2022)
<i>Resolving ambiguity</i>	Sambetbayeva, et al. (2022)
<i>Text analysis</i>	Gomes e Moraes Filho (2011); Sambetbayeva, et al. (2022)
Roteamento Inteligente de Documentos	
<i>Fuzzy Logic</i>	Fil, Nefedov e Binkovskaya (2020); Ismagilov, et al. (2019); Obukhov, Krasnyanskiy e Nikolyyukin (2019);
<i>Information extraction</i>	Sambetbayeva, et al. (2022)
<i>Machine Learning (LM)</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Tkachenko e Denisova (2022); Sambetbayeva, et al. (2022); Obukhov, Krasnyanskiy e Nikolyyukin (2019)
<i>machine learning algorithms</i>	Sambetbayeva, et al. (2022)
<i>multi-agent approach</i>	Sambetbayeva, et al. (2022)
<i>Resolving ambiguity</i>	Sambetbayeva, et al. (2022)
<i>Text analysis</i>	Gomes e Moraes Filho (2011); Sambetbayeva, et al. (2022)
Suporte à tomada de decisões	
<i>Artificial Neural Networks (ANN)</i>	Krasnyanskiy; et al. (2020); Fayziev et al. (2021); Krasnyanskiy, Obukhov e Solomatina (2019); Obukhov, Krasnyanskiy e Nikolyyukin (2019)
<i>Convolutional Neural Network (CNN)</i>	Obukhov, Krasnyanskiy e Nikolyyukin (2019)
<i>Decision Support Subsystem (DSS)</i>	Obukhov, Krasnyanskiy e Nikolyyukin (2019)
<i>Expert Systems (ES)</i>	Obukhov, Krasnyanskiy e Nikolyyukin (2019)
<i>Fuzzy Logic</i>	Fil, Nefedov e Binkovskaya (2020); Ismagilov, et al. (2019); Obukhov, Krasnyanskiy e Nikolyyukin (2019);
<i>Gated Recurrent Unit (GRU)</i>	Obukhov, Krasnyanskiy e Nikolyyukin (2019)
<i>Inductive and Deductive Methods</i>	Obukhov, Krasnyanskiy e Nikolyyukin (2019)

Abordagens técnicas mapeadas	
Abordagem	Referência
<i>Knowledge Processing Algorithm</i>	Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Long Short-Term Memory (LSTM)</i>	Mikheev e Yakimov (2019); Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Machine Learning (LM)</i>	Krasnyanskiy, Obukhov e Solomatina (2019); Tkachenko e Denisova (2022); Sambetbayeva, et al. (2022); Obukhov, Krasnyanskiy e Nikolyukin (2019)
<i>Semantic Networks</i>	Obukhov, Krasnyanskiy e Nikolyukin (2019)

Fonte: Elaborado pela autora (2024)

Esta classificação foi definida com base no tipo de aplicação, mas não foi possível categorizá-la segundo a abordagem por dois motivos principais:

- Nem todos os autores descrevem claramente o tipo de abordagem. Alguns mencionam apenas o nome da abordagem, outros fornecem definições limitadas, e poucos explicam detalhadamente sua função;
- A mesma abordagem pode ser classificada de forma distinta. Por exemplo, o "Aprendizado Supervisionado" pode ser referenciado como técnica, método ou algoritmo, dependendo do contexto. Da mesma forma, frameworks como *TensorFlow* podem ser considerados ferramenta ou arquitetura.

Essas inconsistências criaram barreiras para o entendimento pleno da função e do tipo de uso das abordagens, dificultando uma análise padronizada e comparativa. Portanto, uma metodologia mais padronizada para descrição e categorização seria essencial para aumentar a clareza e a aplicabilidade prática no campo da Gestão Eletrônica de Documentos.

O desenvolvimento de um tesouro seria uma solução eficiente para padronizar terminologias e classificações, promovendo um entendimento mais claro e consistente entre pesquisadores e profissionais da área. Esse trabalho deveria contar com o envolvimento de especialistas em Ciência da Computação, Ciência de Dados e profissionais da Ciência da Informação, garantindo que o vocabulário desenvolvido seja representativo das práticas tecnológicas e documentais. O tesouro também possibilitaria facilitar a interoperabilidade entre sistemas, a recuperação de

informações e o desenvolvimento de ferramentas que promovam maior precisão no uso das abordagens.

Para concluir a resposta à primeira questão de pesquisa (Quais são os principais pontos de interseção entre Gestão Eletrônica de Documentos, Inteligência Artificial e Ciência de Dados?), é necessário explorar os objetivos em comum dessas áreas:

- **Melhoria da eficiência e precisão:** A Ciência da Informação busca eficiência por meio de padrões, métodos e técnicas para organizar e gerenciar informações de forma estruturada. A IA e CD complementam essa busca ao automatizar processos, aplicar algoritmos avançados para identificar padrões e aumentar a precisão no tratamento da informação.
- **Otimização do acesso à informação:** A CI desenvolve sistemas e práticas para garantir a acessibilidade, organização e preservação dos dados. IA e CD, por sua vez, ampliam essa capacidade com o uso de tecnologias como aprendizado de máquina e análise preditiva, que personalizam e agilizam o acesso às informações relevantes, mesmo em grandes volumes de dados.
- **Suporte à tomada de decisão:** A CI contribui ao oferecer informações organizadas e confiáveis como base para decisões estratégicas. IA e CD aumentam o impacto ao fornecer análises avançadas, insights gerados por modelos preditivos e sistemas que oferecem recomendações automatizadas e baseadas em dados.
- **Melhoria da segurança da informação:** Enquanto a CI define as políticas e práticas para proteção dos dados, IA e CD trazem inovações como detecção de anomalias, sistemas de autenticação avançados e análises de riscos para fortalecer a segurança.
- **Aprimoramento da experiência do usuário:** A CI foca em projetar sistemas acessíveis e fáceis de usar, enquanto IA e CD introduzem recursos como interfaces personalizadas e interativas que tornam a experiência mais intuitiva e eficaz.

7.2. P2 - Desafios, tendências e lacunas de pesquisa

Na segunda questão de pesquisa procurou-se entender quais desafios, tendências e lacunas de pesquisa estão presentes na interseção dessas áreas de IA, CD e GED.

As principais dificuldades enfrentadas pelos autores nos estudos foram reunidas e analisadas, evidenciando barreiras técnicas e conceituais. Embora nem todos tenham descrito explicitamente esses desafios, foi possível identificá-los indiretamente por meio das metodologias apresentadas, detalhamento dos passos e etapas das atividades, e, em alguns casos, pelos próprios resultados das pesquisas.

Esses elementos indicam questões como limitações no acesso a dados, complexidade no desenvolvimento de tecnologias, e desafios na replicação ou validação de soluções, destacando áreas que exigem maior atenção e aprimoramento. Os desafios identificados foram:

Forma de mensurar a eficiência dos algoritmos de classificação: Um dos desafios enfrentados na aplicação de algoritmos de classificação automática de documentos é a ausência de critérios padronizados para avaliar sua eficiência. A literatura não define valores mínimos ou limites objetivos para considerar se a taxa de precisão de um sistema é alta ou baixa. Assim, a avaliação depende da subjetividade e da expertise de especialistas, que determinam o que constitui um resultado satisfatório com base no contexto e nos objetivos do sistema. (Gomes; Moraes Filho, 2011)

Imprevisibilidade no tempo de análise: Um dos desafios na aplicação de algoritmos de classificação automática de documentos é a variabilidade no tempo necessário para processar cada documento. Esse tempo é diretamente influenciado pelo tamanho do arquivo e pelo número de palavras que ele contém, tornando difícil prever com precisão a duração da análise. Essa imprevisibilidade pode impactar a eficiência operacional e a alocação de recursos em sistemas que dependem de processamento em larga escala. (Gomes; Moraes Filho, 2011).

Qualidade na rotulagem de documentos: A precisão dos rótulos é crucial para identificar corretamente o conteúdo dos documentos e otimizar os resultados da classificação automática. Cada documento pode receber um ou mais rótulos, que atuam como palavras-chave representativas. Uma rotulagem inadequada compromete a eficiência do algoritmo de classificação, enquanto rótulos bem

definidos garantem maior precisão e relevância nos resultados obtidos. (Liu; Hsaio; Lee; Chen, 2013).

Falta de um grande conjunto de dados para treinamento e teste: A eficácia de modelos de inteligência artificial depende diretamente da quantidade e qualidade dos dados disponíveis para treinamento e validação. Quanto maior o volume de documentos utilizados, mais robusto e preciso será o modelo desenvolvido. (Obukhov; Krasnyanskiy; Nikolyukin, 2019; Kefali; Drabsia, 2018).

Dificuldade no gerenciamento de classes: Muitas vezes há um número elevado de classes, frequentemente ultrapassando centenas por mil desenhos técnicos, que exige que o modelo lide com uma grande diversidade de categorias, dificultando a organização e classificação precisa. (Obukhov; Krasnyanskiy; Nikolyukin, 2019).

Complexidade de Desenhos Técnicos: A estrutura complexa das imagens em documentos técnicos de engenharia apresenta desafios significativos. Essas imagens frequentemente incluem grandes áreas vazias, quadros variados em tamanho e estilo, além de desenhos com diferentes escalas e ângulos de rotação. Essa diversidade estrutural aumenta a dificuldade de processamento e reconhecimento, exigindo algoritmos avançados e técnicas específicas para garantir que modelos de inteligência artificial possam identificar padrões e similaridades de forma eficiente e precisa. (Obukhov; Krasnyanskiy; Nikolyukin, 2019).

Grande especificidade das áreas do conhecimento: Cada área possui um conjunto de características específicas e limitações que deve ser analisada e estruturada para garantir eficiência e precisão na classificação de documentos. (Krasnyanskiy; Obukhov; Solomatina, 2019).

Falta de um Padrão Universal de Classificação: Por seu caráter diversificado, os documentos possuem grande variedade de características, tipos, tamanhos e estruturas, por isso muitos documentos não podem ser classificados de forma inequívoca e consistente, o que pode gerar inconsistência na classificação. É preciso de basear em determinadas características para garantir essa qualidade. (Krasnyanskiy; Obukhov; Solomatina, 2019).

Expertise na Áreas de Conhecimento: Uma das dificuldades na aplicação de abordagens de Inteligência Artificial está diretamente relacionada ao conhecimento

profundo da área de atuação. Essa lacuna de conhecimento impacta etapas essenciais, como o levantamento das necessidades específicas da aplicação, a definição de parâmetros e critérios de classificação para a configuração do modelo. (Gomes; Moraes Filho, 2011; Fil; Nefedov; Binkovskaya, 2020; Krasnyanskiy; *et al.*, 2020).

Qualidade na digitalização dos documentos: A digitalização de imagens com baixa qualidade apresenta desafios significativos no pré-processamento, dificultando a remoção eficaz de ruídos. Essa limitação impacta diretamente a precisão e a qualidade dos resultados obtidos nas etapas subsequentes de análise e extração de informações. (Kefali; Drabsia, 2018).

Alterações no padrão de documentos: A padronização dos formatos de documentos é um fator crucial para a eficiência na automatização dos processos de tratamento de imagens. No entanto, ao longo dos anos, os documentos frequentemente passam por alterações em seus layouts e estruturas, o que representa um desafio significativo para a automação do pré-processamento, exigindo maior esforço para lidar com essas variações. (Kefali; Drabsia, 2018).

Elementos não textuais em documentos: A presença de elementos não textuais como multicolumna, figuras e em áreas com carimbos e assinaturas, induzem a erros de reconhecimento pelo algoritmo. Estes elementos devem ser mapeados previamente e tratados na etapa de pré-processamento. (Mikheev; Yakimov, 2019).

Qualidade da Classificação de termos: A dificuldade na classificação de termos é um desafio recorrente em pesquisas envolvendo análise textual e processamento de linguagem natural. Tkachenko e Denisova (2022) propuseram uma solução inovadora ao atualizar o método TF-DF (*Term Frequency - Document Frequency*), atribuindo pesos diferenciados às palavras com base em sua classe gramatical (como adjetivos, advérbios e verbos). Essa abordagem reduz a influência de "palavras de ruído", que recebem menor peso no documento, evitando que impactem negativamente os resultados da classificação. Essa solução melhora a precisão e a relevância das classificações automatizadas. (Tkachenko; Denisova, 2022).

Necessidade de grandes volumes de dados: A necessidade de grandes volumes de dados é uma das principais dificuldades na aplicação de técnicas de

análise e classificação em sistemas GED. O sucesso dessas iniciativas depende de um corpus de documentos textualmente intensivo. Essa exigência reflete a importância de dados variados e representativos para treinar algoritmos e garantir que as soluções propostas sejam robustas e generalizáveis. Sem um volume adequado de dados, os resultados podem ser limitados ou enviesados, comprometendo a aplicabilidade e a precisão das tecnologias desenvolvidas. (Al Qady; Kandil, 2013).

Expansão contínua dos documentos: em projetos de engenharia, documentos são gerados continuamente, e a classificação depende de entender e organizar todos os conteúdos. Isso cria dificuldades porque o sistema precisa lidar com um volume crescente e informações cada vez mais complexas, o que pode causar confusão ao classificar documentos. (Al Qady; Kandil, 2013).

Aparecimento de novos temas: A expansão contínua de documentos, especialmente em projetos de engenharia, apresenta desafios significativos. A geração constante de novos documentos, aliada à necessidade de compreender e organizar conteúdos heterogêneos e complexos, exige sistemas capazes de escalar de forma eficiente. Isso pode sobrecarregar os algoritmos de classificação, dificultando a manutenção da precisão e consistência, além de aumentar o risco de confusões ou redundâncias. (Al Qady; Kandil, 2013).

Dificuldade na preparação dos documentos para análise semântica: A preparação dos documentos para análise semântica (*parsing*) é, de fato, uma das maiores dificuldades, pois o algoritmo depende de um formato padronizado para compreender e processar os dados de forma eficaz. Além disso, a análise de IA baseia-se em padrões textuais e não em um entendimento completo das nuances da área de conhecimento e não há percepção contextual ou cognitiva para lidar com ambiguidade ou inconsistências como um humano faria, o que pode limitar a precisão em situações complexas. (Al Qady; Kandil, 2010).

Dados reais para treinamentos e testes: Para o treinamento eficaz de algoritmos, é fundamental dispor de grandes volumes de dados reais e representativos, permitindo análises precisas e garantindo que a aplicação seja replicável em diferentes áreas de domínio. No entanto, o acesso a esses dados pode ser extremamente desafiador devido ao caráter confidencial de muitos documentos e às restrições impostas pela Lei Geral de Proteção de Dados (LGPD), que exige o manejo responsável e seguro das informações pessoais. (Ruiz; *et al.*, 2020)

A integração de IA em sistemas de GED enfrenta diversas lacunas que limitam seu avanço e adoção. A baixa produção científica da área de CI, a falta de padrões específicos para a aplicação de IA em GED resulta em abordagens fragmentadas, dificultando a interoperabilidade e replicabilidade das soluções.

A falta de integração com fluxos de trabalho preexistentes gera uma desconexão entre os sistemas tradicionais de GED e as novas tecnologias baseadas em IA dificultando seu uso de maneira eficaz. Os estudos analisados indicam que as aplicações propostas são majoritariamente pontuais, com apenas um estudo teórico oferecendo uma visão holística do SGED.

Mesmo quando focando na melhoria de processos, automação e agilidade, as soluções abordam apenas funcionalidades e aspectos técnicos, sem considerar estudos sobre uso e usuário ou análises sobre a aceitação e relevância prática entre bibliotecários e usuários finais.

Além disso, os impactos éticos e sociais dessas tecnologias, como questões de privacidade, transparência e impacto no emprego, são ainda pouco explorados, apesar de sua relevância crescente. Outra barreira significativa é a baixa adoção por pequenas e médias empresas, frequentemente devido a custos elevados, falta de conhecimento técnico e percepção de complexidade.

Outra lacuna identificada é a falta de envolvimento de profissionais da informação nas diferentes etapas do desenvolvimento dessas tecnologias. Esses profissionais, possuem conhecimento especializado essencial que poderia colocá-los na vanguarda do desenvolvimento tecnológico.

Seu conhecimento em gestão da informação, organização de dados, análise crítica e compreensão de fluxos de trabalho pode contribuir significativamente para o design, implementação e avaliação de tecnologias inovadoras, especialmente em processos que envolvem Inteligência Artificial, Ciência de Dados e Gestão Eletrônica de Documentos (GED).

Com esta pesquisa identificou-se importantes oportunidades e desafios para o desenvolvimento e aprimoramento do Profissional da Informação no contexto da integração entre Inteligência Artificial, Ciência de Dados e Gestão Eletrônica de Documentos:

Desafios: Envolvem o aprofundamento no domínio das abordagens de IA, especialmente no trabalho colaborativo com profissionais de Tecnologia da Informação (TI). Isso exige a compreensão da lógica das tecnologias, da linguagem técnica utilizada e das possibilidades de aplicação prática. Além disso, é necessário lidar com o ritmo acelerado das inovações e o volume crescente de avanços tecnológicos, que demandam constante atualização.

Oportunidades: Todas as abordagens de IA e aprendizado de máquina precisam de treinamento, o que torna indispensável a participação de profissionais com conhecimento aprofundado na área de atuação. Esses especialistas são responsáveis por tarefas essenciais, como análise e mapeamento de áreas de conhecimento, definição de critérios de classificação, criação de categorias e classes de documentos, além da identificação, análise e categorização de tipos, formatos e estruturas documentais.

Essas atividades são intrinsecamente humanas e não podem ser delegadas a sistemas automatizados, destacando a importância estratégica do Profissional da Informação neste contexto. Dessa forma, a pesquisa identifica que a integração entre IA e GED não apenas reforça a relevância dos profissionais da área, mas também amplia o escopo de suas competências, exigindo uma atuação interdisciplinar, proativa e alinhada às inovações tecnológicas.

Por fim, durante a análise da literatura, ficou evidente que as pesquisas frequentemente apresentam uma abordagem técnica, focada no desenvolvimento de tecnologias e inovação tecnológica. Poucos estudos exploram o uso estratégico da informação, um elemento essencial para impulsionar a sociedade e fomentar inovações em múltiplos níveis.

Ao investigar o cenário da inovação, nota-se que o foco se limita quase exclusivamente às inovações tecnológicas, deixando de lado abordagens importantes como inovações de processos. Este aspecto inclui mudanças nos fluxos de trabalho, formas de produção, recebimento, armazenamento e distribuição de informações e documentos.

As inovações apresentadas estão, em sua maioria, restritas aos sistemas de gestão, sem explorar modificações nos processos fundamentais que os sustentam. Portanto, torna-se fundamental inovar nos processos, desconstruindo as concepções

tradicionais sobre documentos e repensando como as informações circulam nos fluxos organizacionais. Uma abordagem mais integrada e interconectada possibilita uma visão dinâmica e estratégica da informação, promovendo novas práticas e formas de gestão alinhadas às demandas contemporâneas.

7.3. P3 - Influência da IA nas práticas de GED

Na terceira questão de pesquisa procurou-se entender como as aplicações de IA e CD têm influenciado as práticas na GED.

A aplicação de IA e CD está promovendo mudanças significativas nas práticas organizacionais de Gestão Eletrônica de Documentos (GED), transformando atividades manuais e repetitivas em processos automatizados e inteligentes. As pesquisas mostram muitos esforços na descoberta de informações a partir dos documentos produzidos, transformando a visão de meros arquivadores de documentos para geradores de conhecimento através da análise de documentos e processos.

Essas mudanças refletem em melhorias perceptíveis, dentre elas destacam-se:

Automação e Eficiência: abordagens de IA automatizam tarefas como classificação de documentos, extração de informações e geração de relatórios. Isso reduz erros humanos, economiza tempo e melhora a produtividade organizacional.

Na busca por eficiência, a pesquisa desenvolvida Ruiz; *et al.*, 2020, apresenta a busca por melhorias no processamento dos dados, como resultado, desenvolveu uma arquitetura baseada em microsserviços que melhora o desempenho, diminui o uso de recursos da CPU e memória.

Quanto à automação de processos de classificação e categorização de textos e documentos, Al Qady; Kandil, 2010 desenvolve o CRISP um algoritmo que tenta emular os humanos nas habilidades de identificar conceitos existentes nos documentos a partir de texto apresentado em linguagem natural.

Para este trabalho a avaliação dos resultados é realizada a partir da análise das notas e médias registradas. Identificou-se que das 43 imagens processadas, 31 foram perfeitamente identificadas, apenas uma não realizou a extração e as outras tiveram falhas no registro por problemas ligadas à forma como as imagens foram geradas. Kefali e Drabsia (2018)

Análise Preditiva: A análise de dados em grande escala permite prever demandas, otimizar fluxos de trabalho e tomar decisões baseadas em insights gerados por IA.

Quanto à automação de fluxos e distribuição de tarefas, duas pesquisas desenvolveram abordagens promissoras. Os autores Ismagilov; *et al.* e Sambetbayeva, *et al.*, criaram sistemas que identifica e distribui tarefas para especialistas de forma automática a partir da análise das características do documento, como o tipo, conteúdo, e destinatários esperados. Ainda neste sentido os autores Onan; Atik; Yalçın, (2021) desenvolveram uma abordagem para classificar e direcionar corretamente as solicitações de suporte de serviços, através da análise de relatórios de bug e histórico de solicitações.

Personalização: Sistemas baseados em IA adaptam-se às necessidades dos usuários, oferecendo interfaces e funcionalidades personalizadas, melhorando a experiência do usuário e aumentando a adesão.

No tema da personalização os autores Krasnyanskiy; *et al.* (2020) desenvolveram um algoritmo que possibilite a adaptação do Sistema de GED às características assimiladas do usuário a partir de seu histórico, características e padrões de atividades.

Suporte à tomada de decisões: Sistemas de IA fornecem subsídios valiosos ao extrair, organizar e analisar informações relevantes, transformando dados brutos em insights acionáveis.

Os autores Obukhov, Krasnyanskiy e Nikolyukin (2019) desenvolveram um subsistema que se baseia nos padrões de uso do sistema e no histórico de operações do usuário, para sugerir próximas ações associadas à sequência de suas atividades. Isso faz com que o usuário tome ações mais assertivas, evita erros e “cliques” desnecessários e agiliza o processo de trabalho.

Segurança da Informação: Sistemas de IA fortalecem a segurança da informação ao detectar ameaças em tempo real, prevenir vazamentos de dados e monitorar padrões de acesso.

Os autores Fayziev *et al.* (2021) identificaram que a combinação de chaves criptográficas com a autenticação biométrica combinada garante mais segurança e

evita acessos não autorizados, alterações indevidas, e assegura que todas as operações são executadas apenas por usuários devidamente autenticados.

As pesquisas mostram diversos benefícios decorrentes das aplicações levantadas neste trabalho. Os benefícios foram destacados, evidenciando os impactos positivos para os sistemas e seus usuários. Esses benefícios incluem maior eficiência nos processos de gestão documental, redução no tempo necessário para recuperação de informações, aprimoramento da precisão na classificação e organização de documentos, facilidade no acesso aos dados e suporte aprimorado para tomada de decisões estratégicas.

Além disso, observou-se uma contribuição significativa para a experiência do usuário e para as competências de profissionais da informação, proporcionando ferramentas tecnológicas mais intuitivas e eficientes. No **Quadro 17** é apresentada a relação dos benefícios percebidos organizados por tipo de aplicação prática.

Quadro 17 – Relação dos principais benefícios percebidos

Principais Benefícios percebidos	
Benefícios	Descrição
Classificação de documentos	
Agilidade no processamento de documentos	a categorização automática elimina a necessidade de leitura manual, compreensão e definição dos documentos, permitindo que o trabalho seja realizado pela máquina de forma mais rápida e precisa.
Maior uniformidade na classificação dos documentos	com a classificação automática, o sistema aplica critérios consistentes, evitando variações e erros comuns na classificação manual feita por humanos.
Correção Ortográfica e Semântica	
Redução de custos em desenvolvimento e aplicação	O uso de abordagens baseadas em métodos estatísticos, lógicos, semânticos e estruturais, oferece uma alternativa mais econômica e versátil em comparação aos métodos tradicionais. Essas técnicas simplificam a detecção e correção de erros textuais, eliminando a necessidade de ferramentas de alta especialização e custo elevado, com isso é possível reduzir significativamente os custos associados ao desenvolvimento e aplicação de soluções.
Identificação de similaridade entre documentos	
Eliminação de trabalho manual e ganho eficiência	A substituição do trabalho manual por processos automatizados na verificação de similaridade traz maior eficiência e agilidade ao processamento documental. Isso reduz significativamente o tempo necessário para identificar redundâncias e evita a duplicidade de registros no banco de dados. Além disso, a automação melhora a precisão, minimizando erros humanos e otimizando o gerenciamento de informações.
Suporte à tomada de decisões	
Velocidade e precisão na escolha das próximas operações	O fluxo de trabalho é otimizado por meio de sugestões automáticas para os próximos passos, baseadas no histórico de ações do usuário. Esse recurso

Principais Benefícios percebidos	
Benefícios	Descrição
	proporciona maior agilidade, reduzindo o tempo de decisão, e garante consistência nos dados ao alinhar as operações com padrões previamente definidos.
Diminuição de erros nas decisões do usuário	Auxilia na seleção correta de operações, direcionando os usuários para ações mais prováveis e evitando escolhas equivocadas que possam comprometer o projeto.
Autenticação biométrica	
Redução do risco de acesso não autorizado e roubo de identidade	A autenticação biométrica é significativamente mais segura do que métodos tradicionais, como login e senha, pois se baseia em características únicas e difíceis de falsificar de cada usuário.
Redução de erros de autenticação e falsificações	Com a autenticação biométrica, o índice de erro é praticamente zero, garantindo que apenas usuários autorizados possam acessar o sistema. Além disso, a combinação de múltiplos fatores de autenticação (biometria facial e escrita, por exemplo) aumenta a dificuldade para que um invasor consiga enganar o sistema.
Apoio ao Compliance	Com métodos mais eficientes de autenticação as instituições atendem a regulamentações de proteção de dados, como GDPR, LGPD, e outras normas que exigem controle rigoroso de acesso a informações. Além disso permite maior rastreabilidade das ações realizadas pelos usuários, facilitando auditorias e investigações de segurança.
Otimização de Desempenho	
Redução de Recursos	A diminuição da necessidade de CPU e memória reduz custos operacionais e melhora a eficiência do sistema.
Maior Agilidade de resposta	Respostas mais rápidas, com melhoria de até 77% no desempenho, oferecem uma experiência mais fluida e eficaz ao usuário.
Extração de informações de documentos	
Melhoria nos processos	A automatização de extração de informações proporciona otimização dos processos, redução de custos operacionais e melhora a eficiência no gerenciamento de informações corporativas.
Suporte à tomada de decisões	A extração automatizada de informações contidas em documentos permite que as organizações acessem dados estruturados de forma rápida e precisa. Isso oferece maior clareza e embasamento para tomadas de decisão estratégicas, eliminando a dependência de processos manuais demorados e reduzindo erros.
Indexação automática de documentos e fluxos de trabalho	
Facilidade na Indexação e Pesquisa	Diagramas identificados e categorizados automaticamente permitem que os usuários busquem com precisão, não apenas em documentos textuais, como também por tópicos, símbolos específicos ou fluxos apresentados nos diagramas.
Análise de Fluxos e Processos	Facilita a identificação e comparação de fluxos e processos de trabalho, tornando possível mapear e otimizar operações diretamente pelo GED.
Aprimorar a Experiência do Usuário	
Ambiente mais intuitivo e eficiente	O usuário perceberá este benefício por meio de melhorias práticas na sua interação com o sistema. Ele notará uma redução na frequência de erros e mensagens de falha, além de encontrar um ambiente mais intuitivo e eficiente. O sistema poderá sugerir ações corretivas ou preventivas, oferecer tutoriais direcionados e ajustar funções para atender às suas necessidades específicas, criando uma experiência mais personalizada e produtiva. Isso resultará em maior confiança no uso do SGED e na sensação de que o sistema facilita, em vez de complicar, seu trabalho diário.
Roteamento Inteligente de Documentos	

Principais Benefícios percebidos	
Benefícios	Descrição
Redução de Tempo e Esforço Manual	Automatiza o envio dos documentos, reduzindo o tempo que eles passam aguardando por ação e minimiza o tempo gasto com tarefas administrativas repetitivas.
Minimização de Erros Humanos	Evita falhas no envio para pessoas erradas, garantindo que o documento chegue ao destino correto.
Identificação de Documentos Relacionados	
Eficiência Operacional	Acelera processos internos, melhorando a produtividade e a resposta aos clientes. Reduz a necessidade de intervenção humana em processos que podem ser automatizados, permitindo que a equipe foque em atividades mais estratégicas
Registro Automático de Documentos	
Agilidade e Precisão das Informações	Há maior agilidade e eficiência no automático de documentos, diminuindo a incidência de erros e padronização no cadastro.
Comparação de versões documentos	
Aumento de Produtividade	Identificação automática de documentos duplicados e diferentes versões de um mesmo documento acelera fluxos de trabalho, melhorando a eficiência organizacional.
Redução de Erros	Ao automatizar a comparação de documentos, garante mais precisão na identificação de revisões obsoletas, que minimiza os erros humanos comuns no compartilhamento de documentos entre diferentes setores e empresas
Economia de Tempo e Custos	Ao eliminar a necessidade de conferência manual, aumenta-se o volume de processamento e as instituições economizam tempo e recursos operacionais.
Classificação semântica de documentos	
Recuperação de documentos relacionados	Com a classificação semântica, é possível identificar a etapa do processo a que um documento pertence e localizar automaticamente todos os documentos interligados, facilitando a recuperação e análise de informações relacionadas.
Construção de Ontologias	
Organização do conhecimento	Esse benefício permite identificar e mapear conceitos e suas relações de maneira sistemática, tornando mais eficiente a recuperação de informações, o gerenciamento de conteúdo e o compartilhamento de conhecimento. Organização pode otimizar fluxos de trabalho, facilitar treinamentos e melhorar a comunicação entre áreas, assegurando que todos os envolvidos acessem informações de forma clara e consistente.
Adaptação e personalização da interface	
Ambiente mais intuitivo e eficiente	O usuário perceberá este benefício por meio de melhorias e facilidade de encontrar informações, além de encontrar um ambiente mais intuitivo e eficiente. O sistema poderá sugerir próximas ações, oferecer tutoriais direcionados e ajustar funções para atender às suas necessidades específicas, criando uma experiência mais personalizada e produtiva. Isso resultará em maior confiança no uso do SGED.

Fonte: Elaborado pela autora (2024)

Por fim, no que se refere às mudanças impulsionadas pelas tecnologias, todas as pesquisas analisadas destacaram apenas os aspectos positivos relacionados aos seus avanços e implementações. Não foram identificadas discussões ou abordagens

sobre possíveis impactos negativos ou desafios associados a esses desenvolvimentos.

As mudanças nas práticas organizacionais de Gestão Eletrônica de Documentos, além de transformarem processos e atividades, podem provocar resistência significativa de operadores e gestores. Entre os motivos, destacam-se o medo da substituição de empregos por tecnologias avançadas e a dificuldade em aprender novos sistemas e se adaptar às mudanças no fluxo de trabalho. Esses fatores podem criar barreiras importantes para a adoção de inovações, exigindo estratégias claras de treinamento, comunicação e envolvimento das equipes para superar as resistências e maximizar os benefícios das novas tecnologias.

É importante observar que os benefícios proporcionados pela automação, como a eliminação do trabalho manual, impactam diretamente as funções tradicionais de bibliotecários e arquivistas. Atividades antes exclusivas desses profissionais estão sendo transferidas para máquinas, exigindo uma reorientação na formação e atuação dessas categorias. Em vez de executar tarefas repetitivas, é necessário focar no desenvolvimento e aprimoramento das ferramentas de Inteligência Artificial.

O conhecimento técnico e prático desses profissionais é crucial para o mapeamento de critérios, definição de regras e estruturação de processos, atividades essenciais para o desenvolvimento de sistemas eficientes. Essa colaboração reforça o papel estratégico dos bibliotecários e arquivistas nas equipes multidisciplinares, enquanto programadores e desenvolvedores lidam com a codificação e validação técnica.

Assim, repensar a carreira e adotar uma postura de constante adaptação tornou-se fundamental para evitar a obsolescência, garantindo a relevância e a contribuição ativa desses profissionais em um cenário digital em rápida evolução.

Outro aspecto impactado pelas mudanças realizadas por essas tecnologias emergentes é a confiança na Inteligência Artificial e as questões éticas associadas. A democratização do acesso a essas tecnologias levanta preocupações sobre a transparência dos algoritmos, a confiabilidade das decisões automatizadas e a mitigação de vieses. Esses vieses podem afetar diretamente a equidade nas operações e decisões organizacionais, destacando a necessidade de desenvolver

soluções justas, explicáveis e alinhadas com princípios éticos, garantindo confiança e inclusão nos processos mediados por IA.

7.4. Trabalhos futuros

O desenvolvimento desta pesquisa abriu diversas possibilidades para investigações futuras, oferecendo novas abordagens e temas relevantes para dar continuidade ao estudo. A delimitação de escopo é uma decisão importante para o desenvolvimento da pesquisa. Com tudo, essa delimitação também exclui a exploração de outras áreas e temas potenciais que poderiam ser abordados em estudos posteriores.

Uma linha de pesquisa promissora seria explorar áreas relacionadas à gestão de registros de saúde. Essa área foca na automatização de classificação de sintomas e na previsão de doenças a partir de sintomas apresentados pelos pacientes e de seu histórico escolar. Comparar esses resultados com os achados da pesquisa atual poderia oferecer insights sobre o nível de desenvolvimento das duas áreas e suas atuações em diferentes contextos.

Outra área de interesse seria a gestão de documentos técnicos de engenharia. Esses documentos têm grande complexidade devido ao elevado volume de revisões, interações entre múltiplos atores e seu papel direto no acompanhamento físico e financeiro de projetos. Uma análise mais aprofundada poderia contribuir para compreender essas particularidades e mapear os avanços tecnológicos alcançados pelos pesquisadores nessa área.

Além disso, uma abordagem exploratória poderia ser realizada para investigar o desenvolvimento de tecnologias para GED disponíveis no GitHub. Essa análise poderia identificar ferramentas e soluções já testadas, com potencial para serem incorporadas aos sistemas de Gestão Eletrônica de Documentos (GED), promovendo inovação e aprimoramento tecnológico no campo

Outra possibilidade é uma análise de patentes para compreender as inovações no setor privado e compará-las com as pesquisas científicas. Isso poderia revelar tendências no desenvolvimento tecnológico, destacando o papel das universidades como centros geradores de conhecimento e sua influência na inovação social e tecnológica.

Por fim, esta pesquisa identificou uma lacuna importante: grande variedade de terminologias e classificações de abordagens de IA e CD. O desenvolvimento de um tesouro padronizado seria um recurso importante e poderia facilitar a interoperabilidade entre sistemas, melhorar a recuperação de informações e garantir uma abordagem mais precisa na implementação de tecnologias relacionadas à GED e IA. O desenvolvimento do tesouro exigiria a colaboração interdisciplinar entre especialistas em Ciência da Computação, Ciência de Dados e Ciência da Informação, garantindo que as terminologias sejam representativas e alinhadas às práticas tecnológicas e documentais contemporâneas.

Essas abordagens futuras têm o potencial de expandir o conhecimento na área da GED, contribuindo tanto para o desenvolvimento científico quanto para a aplicação prática em contextos organizacionais e tecnológicos.

8. CONSIDERAÇÕES FINAIS

O presente trabalho apresentou o processo e os resultados de uma revisão sistemática da literatura, cujo objetivo principal foi investigar o papel de tecnologias oriundas das áreas de Inteligência Artificial (IA) e Ciência de Dados (CD) na Gestão Eletrônica de Documentos (GED).

O protocolo da revisão sistemática foi elaborado seguindo diretrizes consolidadas na literatura, conforme descrito. A *string* de busca passou por sucessivos refinamentos a partir de uma metodologia de prototipagem que analisa a relevância dos termos isoladamente enquanto explora suas intersecções no contexto deste trabalho para obter maior precisão na busca por informações e a reprodutibilidade deste trabalho.

Para garantir a amplitude desse trabalho, optou-se pelas bases de dados de pesquisa *Scopus*, *Web of Science*, *Brapci* e *LISTA*. As duas primeiras são bases que abrangem diversas áreas do conhecimento, enquanto que as outras duas são da área da Ciência da Informação. Com os resultados em mãos, as duplicatas foram retiradas e os critérios de inclusão, exclusão e qualidade foram aplicados.

Na etapa que se seguiu, deu-se início à Fase 2 da RSL, referente à condução da revisão, as duplicatas foram retiradas e os critérios de inclusão, exclusão e qualidade foram aplicados. Com a formação do *corpus* final de análise procedeu-se, então para a leitura do texto integral e análise subsequente de todos esses trabalhos.

Quanto à qualidade dos estudos analisados, todos atenderam à pontuação mínima exigida. Na maioria dos casos tratava-se de artigos de congresso, que possui um número limitado de páginas e seu conteúdo precisa ser condensado, em contrapartida, apresentam as inovações que estão sendo discutidas nos maiores eventos da área.

A abordagem utilizada para análise dos dados foi a análise temática e de conteúdo, que se mostrou eficaz ao classificar e organizar elementos significativos em categorias que descreveram e interpretaram os dados coletados. Essa metodologia possibilitou a identificação de padrões, conexões e insights relevantes para compreender as contribuições da Inteligência Artificial e da Ciência de Dados na Gestão Eletrônica de Documentos e, conseqüentemente, responder às questões de pesquisa.

A resposta à primeira questão de pesquisa evidencia que a interseção entre Gestão Eletrônica de Documentos, Inteligência Artificial e Ciência de Dados

concentra-se no desenvolvimento de tecnologias que aprimoram a eficiência e a agilidade dos processos organizacionais, ao mesmo tempo em que oferecem uma visão estratégica para usuários e gestores.

As principais contribuições incluem automação de tarefas, análises preditivas, identificação de padrões, segurança aprimorada e maior acessibilidade à informação. A categorização das aplicações práticas, com 16 categorias distintas, destaca áreas como a classificação e similaridade de documentos como as mais exploradas, evidenciando a relevância do tema.

Apesar de avanços notáveis, foram identificadas lacunas na padronização de terminologias e descrições das abordagens utilizadas. Assim, propõe-se o desenvolvimento de um tesouro multidisciplinar como solução para padronizar e enriquecer as práticas no campo, potencializando a colaboração entre as áreas e o impacto das inovações tecnológicas.

A resposta à segunda questão de pesquisa foca a análise dos desafios, tendências e lacunas na interseção entre Inteligência Artificial, Ciência de Dados e Gestão Eletrônica de Documentos, que revela barreiras técnicas e conceituais significativas. Entre os desafios, destacam-se a falta de critérios padronizados para avaliar algoritmos, diferentes denominações para as mesmas abordagens, necessidade de grande volume de dados para treinamentos e testes, dificuldade de acesso a dados reais para simulações, complexidade de documentos técnicos e limitação na integração com fluxos de trabalho existentes.

O avanço dessas tecnologias na GED caminha para a automação dos processos manuais, como catalogação, classificação, indexação, apoio a pesquisas, até mesmo certas tarefas de apoio ao usuário. Com isso essa mudança impacta diretamente as funções dos profissionais da informação como bibliotecários e arquivistas demandando um realinhamento profissional.

Esses especialistas são essenciais no mapeamento de critérios e estruturação dos processos de automação, colaborando em equipes multidisciplinares no desenvolvimento de ferramentas de IA, enquanto programadores se concentram na codificação. Repensar a carreira e adaptar-se é crucial para evitar a obsolescência e garantir a relevância desses profissionais em um ambiente digital em constante evolução.

Outras tendências incluem a necessidade de maior padronização com o envolvimento de profissionais da informação no desenvolvimento de soluções. Os impactos éticos representam lacunas que demandam maior atenção. Essa análise sugere que a colaboração interdisciplinar e a abordagem centrada no usuário são essenciais para superar esses desafios e promover avanços efetivos no campo.

A análise da literatura revela ainda, que as pesquisas tendem a focar o desenvolvimento de tecnologias e inovações tecnológicas, deixando de lado o uso estratégico da informação e abordagens como inovações de processos. Essas inovações geralmente se restringem aos sistemas de gestão, sem refletirem mudanças nos processos fundamentais e na dinâmica de circulação de informações. É essencial desconstruir concepções tradicionais sobre documentos e repensar os fluxos organizacionais, promovendo abordagens mais integradas e estratégicas alinhadas às demandas contemporâneas.

A análise da terceira questão de pesquisa evidencia que as aplicações de Inteligência Artificial e Ciência de Dados têm transformado significativamente as práticas na Gestão Eletrônica de Documentos. Essas tecnologias estão promovendo automação, eficiência e suporte estratégico por meio da análise preditiva, personalização, tomada de decisões aprimoradas e maior segurança da informação.

No entanto, apesar dos benefícios, como redução de erros, agilidade nos processos e melhorias na experiência do usuário, pode haver resistência organizacional, especialmente relacionada ao medo de substituição de empregos e desafios no aprendizado de novos sistemas. Além disso, preocupações éticas como transparência, viés nos algoritmos e confiança na IA também se destacaram como fatores críticos para uma implementação bem-sucedida e inclusiva dessas inovações.

Esta pesquisa abriu novas possibilidades para estudos futuros, abordando temas relevantes e novas abordagens no campo da GED. Entre as perspectivas destacam-se: a análise da gestão de registros de saúde, focando nas tecnologias de classificação de sintomas e previsões de doenças; o estudo de documentos técnicos de engenharia devido à sua complexidade e impacto em projetos; e uma exploração das tecnologias para GED no GitHub para identificar ferramentas testadas que são passíveis de inovação.

Outra abordagem seria a análise comparativa entre patentes e pesquisas científicas para mapear tendências tecnológicas. Além disso, identificou-se a necessidade de desenvolvimento de um tesouro padronizado, essencial para promover interoperabilidade, recuperação de informações e precisão no uso dessas tecnologias, exigindo colaboração interdisciplinar, conforme descrito. Essas abordagens têm potencial para avançar tanto no conhecimento acadêmico quanto na aplicação prática da GED.

Diante do exposto, considera-se que as questões de pesquisa foram plenamente respondidas e o objetivo deste trabalho foi atendido.

REFERÊNCIAS

ABADI, H. H. N.; HE, Z.; PECHT, M. Artificial intelligence-related research funding by the U.S. National Science Foundation and the National Natural Science Foundation of China”. **IEEE Access**, vol. 8, p.183448–183459, 19 out. 2020. Disponível em: <https://ieeexplore.ieee.org/document/9214828>. Acesso em 11 jun. 2024. DOI: [10.1109/ACCESS.2020.3029231](https://doi.org/10.1109/ACCESS.2020.3029231)

AKAICHI, T.; MERLO, F. Gestão de documentos de uma universidade pública no oeste do estado do Paraná. **Informação & Informação**, Londrina, v. 23, n. 2, p.619-632, maio/ago. 2018. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/26850>. Acesso em: 30 abr. 2024.

ALMEIDA JÚNIOR, O. F. Formação, Formatação: profissionais da informação produzidos em série. In: VALENTIM, Marta Lígia (Org.). **Formação do profissional da informação**. São Paulo: Polis, 2002. 152 p. Disponível em: <https://abecin.org.br/wp-content/uploads/2021/03/Formacao-do-profissional.pdf>. Acesso em: 28 maio 2022.

AL QADY, M.; KANDIL, A. Concept relation extraction from construction documents using natural language processing. **Journal of Construction Engineering and Management**, [s. l.], v. 136, n. 3, p. 294-302, 2010. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-77149152182&doi=10.1061%2f%28ASCE%29CO.1943-7862.0000131&origin=inward&txGid=77e8ee8edd23c205b3dd4e52a2784896>. Acesso em: 06 ago.2024.

AL QADY, M.; KANDIL, A. Document discourse for managing construction project documents. **Journal of Computing in Civil Engineering**, [s. l.], v. 27, n. 5, p.466-475 1 set. 2013. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-84882345017&doi=10.1061%2f%28ASCE%29CP.1943-5487.0000201&origin=inward&txGid=b15598cf7b63358deae500d6188c3d01>. Acesso em: 06 ago.2024.

ALVES, F. M. M.; RIBEIRO, M. C. O. Inteligência artificial no contexto da Ciência da Informação. **Revista Brasileira de Educação em Ciência da Informação (REBECIN)**, São Paulo, edição especial, p.1-15. 2024. Trabalho apresentado no 5º Encontro Regional Norte-Nordeste de Educação em Ciência da Informação (), 2023, [Salvador, BA].

ANDRADE, M. V. M. Gerenciamento Eletrônico da Informação: ferramenta para a gerência eficiente dos processos de trabalho. In: **Seminário Nacional de Bibliotecas Universitárias**, Recife, 2002. Disponível em:

http://repositorio.febab.org.br/files/original/29/4093/SNBU2002_083.pdf. Acesso em: 20 jun. 2023.

ANNA, J. S.; SILVA, L. C. Novos rumos aos fazeres arquivísticos: práticas de gestão de documentos em um arquivo privado. **Informação Arquivística**, Rio de Janeiro, v. 4, n. 1, p. 119-143, jan./jun., 2015. Disponível em: <https://www.aaerj.org.br/ojs/index.php/informacaoarquivistica/article/view/61>. Acesso em: 30 abr. 2024.

ARQUIVO NACIONAL (Brasil). **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005. (Publicações Técnicas; n. 51).

ARAÚJO, C. A. V. Correntes teóricas da ciência da informação. **Ciência da Informação**, v. 38, n. 3, 2009. DOI: <https://doi.org/10.18225/ci.inf..v38i3.1240>. Disponível em: <https://www.scielo.br/j/ci/a/qhsrgPL7T6RbKKVbMwrPMNb/?format=pdf&lang=pt>. Acesso em: 28 mar. 2022.

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

BARDIN, L. **Análise de Conteúdo**. São Paulo: Edições 70, 1977.

BASHIR, R.; GIRI, K. J. Diagram recognition: Domain knowledge based approach. In: **2013 International Conference on Machine Intelligence Research and Advancement - ICMIRA 2013**, Shri Mata Vaishno Devi University (SMVDU), Katra, India, p. 445-449 21-23 dez. 2013. DOI: 10.1109/ICMIRA.2013.94. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-84910010844&doi=10.1109%2fICMIRA.2013.94&origin=inward&txGid=3d088e97cdcfa2d71dbf59c2944f6bbe>. Acesso em: 06 ago.2024

BELLUZZO, R. C. B. As competências do profissional da informação nas organizações contemporâneas. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v.7, n.1, p.58-73, jan./jun. 2011.

BORGLUND, E.; ANDERSON, K. LOST Records: The Consequence of Inadequate Recordkeeping Strategies. In: Ninth International Baltic Conference - Databases and Information Systems VI - DB&IS 2010. Latvia, 5-7 jun. 2010. DOI: 10.3233/978-1-60750-687-4-271

BORKO, H. Information science: what is it? **American Documentation**, Washington, v. 19, n. 1, p.3-5, jan. 1968. DOI: <https://doi.org/10.1002/asi.5090190103>. Acesso em: 28 jun 2022.

BRAGA, K. S. Aspectos relevantes para a seleção de metodologia adequada à pesquisa social em Ciência da Informação. In: MUELLER, S. P. M. (Org.). **Métodos para a pesquisa em Ciência da Informação**. Brasília: Thesaurus, 2007. p.17-38.

BRASIL. **Lei n. 8.159**, de 8 de janeiro de 1991. Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l8159.htm. Acesso em: 30 abr. 2024.

BUFREM, L. S.; COSTA, F. D. O.; GABRIEL JUNIOR, R. F.; PINTO, J. S. P. Modelizando práticas para a socialização de informações: a construção de saberes no ensino superior. **Perspectivas em Ciência da Informação**, v. 15, n. 2, 2010.

CALDERON, W. R.; CORNELSEN, J. M.; PAVEZI, N.; LOPES, M. A. O processo de gestão de documentos e da informação arquivística no ambiente universitário. **Ciência da Informação**, Brasília, v. 33, n. 3, p.97-104, mar. 2004. Disponível em: <http://revista.ibict.br/ciinf/article/view/1037>. Acesso em: 30 abr. 2024.

CAO, L. Data Science. **ACM Computing Surveys**, [s. l.], v. 50, n. 3, p.1-42, 29 jun. 2017. Association for Computing Machinery (ACM). Disponível em: <http://dx.doi.org/10.1145/3076253>. Acesso em: 07 ago. 2023.

CARLOMAGNO, M. C.; ROCHA, L. C. Como criar e classificar categorias para fazer análise de conteúdo: uma questão metodológica. **Revista Eletrônica de Ciência Política**, [s. l.] v. 7, n. 1, p. 173-188, 2016.

CONEGLIAN, C. S. **Recuperação da Informação com abordagem semântica utilizando Linguagem Natural**: a Inteligência Artificial na Ciência da Informação. 2020. Tese (Doutorado em Informação da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2020.

COZMAN, F. G.; NERI, H. O que, afinal, é Inteligência Artificial? In: COZMAN, Fabio G.; PLONSKI, Guilherme Ary; NERI, Hugo. (Org.). **Inteligência Artificial: Avanços e Tendências**. São Paulo: Instituto de Estudos Avançados, 2021. cap. 1. p.21-29, PDF. <https://doi.org/10.11606/9786587773131>. Disponível em: <http://www.livrosabertos.sibi.usp.br/portaldelivrosUSP/catalog/book/650>. Acesso em: 30 maio 2022.

DAVENPORT, T. H.; PRUSAK, L. **Conhecimento Empresarial**: como as organizações gerenciam o seu capital intelectual. 4. ed. Rio de Janeiro: Campus, 1998.

DORNELES, S. L. Aplicações de inteligência artificial na gestão documental: revisão sistemática na LISTA, EMERALD e SCOPUS. **Revista Fontes Documentais**. Edição Especial: II ABM, Aracaju. v. 5, 2022.

DORNELES, S. L. Aplicações de inteligência artificial na gestão documental: Revisão sistemática na LISTA, EMERALD e SCOPUS. In: **E. Cerveira & Z. Duarte (Eds.), Integração e convergência das instituições de informação e cultura**, CITCEM – Centro de Investigação Transdisciplinar Cultura, Espaço e Memória. Porto, Portugal, p.85-104, 2024. DOI: 10.21747/978-989-8970-70-1/int.

ELIAS, E. D. Gerenciamento eletrônico de documentos (GED): aplicação na Universidade Federal de Santa Catarina. **ÁGORA**, Florianópolis, v. 22, n. 45, p.15-30, jul. 2012. Disponível em: <https://brapci.inf.br/index.php/res/download/48043>. Acesso em: 30 abr. 2024.

ELSEVIER. Scopus: Guia de referência rápida. **ELSEVIER**, 2015. Disponível em: https://www.periodicos.capes.gov.br/images/documents/Scopus_Guia%20de%20refer%C3%AAncia%20r%C3%A1pida_10.08.2016.pdf. Acesso em: 14 jun. 2023.

FAYZIEV, S. I.; KHAMRAKULOV U. S.; SADIKOV, S. B.; MAMADJANOV, B. N.; Mamadoliyev, S. X. Development of a Methodology for Protecting a Software Package for Electronic Document Management. In: **2021 International Conference on Information Science and Communications Technologies – ICISCT 2021**. Tashkent, Uzbekistan, 3-5 nov. 2021. DOI: 10.1109/ICISCT52966.2021.9670416. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85125054039&doi=10.1109%2fICISCT52966.2021.9670416&origin=inward&txGid=9f673b2b8d59cd43573ceb653331673b>. Acesso em: 06 ago.2024.

FERNEDA, E.; DIAS, G. A. OntoSmart: proposta de um modelo de recuperação de informação baseado em ontologia. In: **XII Congreso ISKO España y II Congreso ISKO España-Portugal: organización del conocimiento para sistemas de información abiertos**. Murcia: Universidad de Murcia, 2015. Disponível em: http://www.iskoiberico.org/wp-content/uploads/2015/11/80_Ferneda.pdf. Acesso em: 28 maio 2022.

FIL, N.; NEFEDOV, L.; BINKOVSKAYA, A. Fuzzy Model for Estimating the Probability of User Error in the Electronic Document Management System. In: **2020 IEEE International Conference on Problems of Infocommunications Science and Technology - PIC S and T 2020**, Kharkiv, Ukraine, p. 259-262, 6-9 out. 2020. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85114404836&doi=10.1109%2fPICST51311.2020.9467950&origin=inward&txGid=2804e3d47fc9752cfc12b4ba1d4723c5>. Acesso em: 06 ago.2024.

FONSECA, C. A.; SOUZA NETTO, R. S. de; GUELPELI, M. V. C.; BODOLAY, A. N. AnoTex: rotina de filtragem de dados estruturados do gênero artigo científico como contribuição para o pln. **Texto Livre: Linguagem e Tecnologia**, [s. l.], v. 11, n. 3, p.40-64, 26 dez. 2018. Universidade Federal de Minas Gerais - Pro-Reitoria de

Pesquisa. <http://dx.doi.org/10.17851/1983-3652.11.3.40-64>. Acesso em: 28 mai. 2022.

FONSECA, C. A.; GUELPELI, M. V. C.; SOUZA NETTO, R. S. de. Representação dos dados estruturados do gênero textual como técnica para o processamento automático de texto. **Texto Livre**, Belo Horizonte-MG, v. 15, p.e35445, 2022. DOI: 10.35699/1983-3652.2022.35445. Disponível em: <https://periodicos.ufmg.br/index.php/textolivre/article/view/35445>. Acesso em: 29 maio 2022.

FREIRE, G. H. de A. Ciência da informação: temática, histórias e fundamentos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 1, p.6-19, jan./abr. 2006. Disponível em: <https://doi.org/10.1590/S1413-99362006000100002>.

FREITAS, L. M.; ALBUQUERQUE, A. C. de. Organização e recuperação da informação arquivística no contexto do planejamento organizacional. In: **Seminário em Ciência da Informação**, 6., Londrina, 2016. Anais [...]. Londrina: UEL, 2016. Disponível em: <http://www.uel.br/eventos/cinf/index.php/secin2016/secin2016/paper/view/343/206>. Acesso em: 30 abr. 2024.

GIL, A. C. **Métodos e Técnicas de pesquisa social**. 6. Ed. São Paulo: Atlas, 2008.

GOMES, R. R. G.; MORAES FILHO, R. O. CADWeb – Categorização automática de documentos digitais. **Ciência da Informação**, Brasília, DF, v. 40, n. 1, p.68-76, jan./abr., 2011. Disponível em: <https://research.ebsco.com/linkprocessor/plink?id=8b77781e-40a8-3b17-ae64-bb666e42a8f8>. Acesso em 21 ago. 2024.

GROENNER, L. C. **Inteligência Artificial no campo Ciência, Tecnologia e Sociedade**: um estudo comparativo entre indicadores da produção científica. Tese (Doutorado em Ciência, Tecnologia e Sociedade) - Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos. São Carlos, p. 170. 2024

GUIMARÃES, L. M. S.; MEIRELES, M. R. G.; ALMEIDA, P. E. M. de. Avaliação das etapas de pré-processamento e de treinamento em algoritmos de classificação de textos no contexto da recuperação da informação. **Perspectivas em Ciência da Informação**, [s. l.], v. 24, n. 1, p.169-190, mar. 2019. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/1981-5344/3505>. Acesso em: 27 jun. 2022.

HEY, T.; HEY, J. e-Science and its implications for the library community, **Library Hi Tech**, [S. l.], v. 24, n. 4, 2016, p.515-528. Disponível em: <https://www-emerald.ez31.periodicos.capes.gov.br/insight/content/doi/10.1108/073788306107153>

[83/full/pdf?title=escience-and-its-implications-for-the-library-community](#). Acesso em: 04 ago. 2023.

HOFFMANN, W. A. M. **Gestão do Conhecimento**: desafios de aprender. São Carlos: Compacta, v. 1, 2009.

IKEBE, N. T., OLIVEIRA JR., M. M. A inovação pela administração do conhecimento. **XXI Simpósio de Gestão da Inovação Tecnológica**, São Paulo, p.7-10, nov. 2000.

INTELLECTUAL PROPERTY OFFICE (IPO). Artificial Intelligence: A worldwide overview of AI patents and patenting by the UK AI sector. Intellectual Property Office, Londres, p.46. 2024. Disponível em: [https://assets.publishing.service.gov.uk/media/5d2dc787ed915d2fe1abfabe/Artificial Intelligence - A worldwide overview of AI patents.pdf](https://assets.publishing.service.gov.uk/media/5d2dc787ed915d2fe1abfabe/Artificial%20Intelligence%20-%20A%20worldwide%20overview%20of%20AI%20patents.pdf). Acesso em 11 jun. 2024.

ISMAGILOV, I. I.; NUGAEV, F. S.; KATASEV, A. S.; TALIPOV, N. G.; KATASEVA, D. V. Decision-Making Support System for tasks distribution in Personal Data Operators Register Maintaining based on a Fuzzy-Production Model. **Revista Dilemas Contemporâneos**: Educación, Política y Valores, [s. l.], v. 4, n. 64, p. 17, mar. 2019. Disponível em: <https://www.dilemascontemporaneoseducacionpoliticayvalores.com/index.php/dilemas/article/view/557>. Acesso em: 06 ago.2024.

JUMANOV I. I.; KARSHIEV Kh. B. Mechanisms for optimization of detection and correction of text errors based on combining multilevel morphological analysis with n-gram models. In: **Journal of Physics**: Conference Series, Volume 1546, IV International Scientific and Technical Conference "Mechanical Science and Technology Update" (MSTU-2020), Omsk, Russia, 17-19 mar. 2020. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1546/1/012082>. Acesso em: Acesso em: 06 ago. 2022.

KEFALI, A.; DRABSIA, S. Localization of scores and average in Algerian baccalaureate transcripts. In: **22nd International Conference on Enterprise Information Systems - ICEIS 2020**; Virtual, Online, 5-7 maio 2020. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85090796518&origin=inward&txGid=51e6dc5575bc30f0815a7b966ec72e13>. Acesso em 06 ago.2024.

KITCHENHAM, B. **Procedures for Performing Systematic Reviews**. Joint Technical Report Software Engineering Group, Keele University (TR/SE-0401), United Kingdom and Empirical Software Engineering, National ICT Australia Ltd, Australia (0400011T.1), 2004.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing systematic literature reviews in software engineering**. Technical Report. EBSE, Keele

University and Durham University Joint Report, 2007. Disponível em:
https://www.elsevier.com/data/promis_misc/525444systematicreviewsguide.pdf.
Acesso em: 16 jun. 2023.

KRASNYANSKIY, M. N.; OBUKHOV, A. D.; SOLOMATINA, E. M. The algorithm of document classification of research and education institution using machine learning methods. In: **2019 International Science and Technology Conference – EastConf 2019**, Vladivostok, Russia; 1-2 mar. 2019. Disponível em:
<https://www.scopus.com/record/display.uri?eid=2-s2.0-85067504184&doi=10.1109%2fEastonf.2019.8725319&origin=inward&txGid=c1a2253c86382e933d550273593979b3>. Acesso em: 06 ago.2024.

KRASNYANSKIY, M. N.; OBUKHOV, A. D.; VOYAKINA, A. A.; SKVORTSOV, V. I.; KHVOROV, V. A. The algorithm of document routing in the electronic document management system using machine learning methods. **Progress in Artificial Intelligence**, [s. l.] v. 9, n. 4, p.287-303, 1 dez. 2020.

LAKATOS, I.; MUSGRAVE, A. **Criticism and the growth of knowledge**. Cambridge: Cambridge University Press, 1970.

LIMA, E. B.; FERREIRA, E. G. A.; ABREU, F. F.; LIMA, G. M. C de; ARAÚJO, C. A. Á. Profissionais da informação: conceitos, competências e mercado de trabalho. **Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação**, In. XVIII ENANCIB, 2017. Disponível em:
<http://hdl.handle.net/20.500.11959/brapci/105233>. Acesso em: 27 jun. 2022.

LIU, C. L.; HSAIO, W. H.; LEE, C. H.; CHEN, C. H. Clustering tagged documents with labeled and unlabeled documents. **Information Processing and Management**, [s. l.], v. 49, p.596–606, maio, 2013. Disponível em:
<https://research.ebsco.com/linkprocessor/plink?id=caa0d78a-603f-3634-91b8-0f846f21c729>. Acesso em 21 ago. 2024.

LOPES, I. L. Estratégia de busca na recuperação da informação: revisão de literatura. **Ciência da Informação**, [s. l.], v. 31, n. 2, 2002. DOI:
<https://doi.org/10.18225/ci.inf.v31i2.961>. Acesso em: 08 fev. 2024.

MAGALHÃES, E. V. S.; ZANIRO, D. L.; SOARES, V. R. M. AMARAL, R. M.; NEVES, B. C. Inteligência Artificial e Ciência de Dados na Gestão de Documentos: uma Revisão Sistemática de Literatura. In: **XXIV Encontro nacional de pesquisa em Ciência da Informação (XXIV ENANCIB)**. Vitória, ES, p. 15. 04-08 nov. 2024.

MANARIS, B. Natural language processing: A human-computer interaction perspective. **Advances in Computers**, [s. l.], v. 47, p.1-66, 1998. DOI:10.1016/S0065-2458(08)60665-8. Acesso em: 19 jun. 2022.

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: http://manu47.magtech.com.cn/Jwk3_jdis/Y2016/V1/I2/1. Acesso em: 15 dez. 2022.

MARTINS, A. L. Potenciais aplicações da Inteligência Artificial na Ciência da Informação. **Informação & Informação**. Londrina, v. 15, n. 1, p. 1-16, jul./jun. 2010. DOI: 10.5433/1981-8920.2010v15n1p1. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/3882>. Acesso em: 20 jun. 2023.

MIKHEEV, M.A.; YAKIMOV, P.Y. Development of the documents comparison module for an electronic document management system. In: **Data Science Session at the 5th International Conference on Information Technology and Nanotechnology - DS-ITNT 2019**, v. 2416, p. 527-533, Samara, Russia, 21-24 maio 2019. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85070973402&doi=10.18287%2f1613-0073-2019-2416-527-533&origin=inward&txGid=dcb8969d1cf9eb76a28fccd362c6816c>. Acesso em: 06 ago.2024.

MINAYO, M. C. S. **O desafio do conhecimento: pesquisa qualitativa em saúde**. 12 ed. São Paulo: Hucitec, 2010.

MORAIS, M. W., RAMALHO, R. A. S. Representação em SKOS de um microtesauro de conhecimentos estratégicos nas organizações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 24, n. 2, p.187-198, out./dez. 2019.

MORESI, E. A. D. Delineando o valor do sistema de informação de uma organização. **Ciência da Informação**, Brasília, v. 29, n. 1, abr. 2000.

MOUTINHO, S. O. M.; ALENCAR, D. F.; MARTINS, P. G. M., CONEGLIAN, C. S. Ciência Da informação e Ciência de Dados: convergências interdisciplinares. **Encontros Bibli**, Florianópolis, v. 29, p.26, 2024.

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION (NARA). Context for electronic records management (ERM). **National Archives**, 2019. Disponível e: <https://www.archives.gov/records-mgmt/initiatives/context-for-erm.html>). Acesso em: 20 jun. 2023.

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.

NAKAMITI, E. K. **Agentes inteligentes artificiais**. 2009. 91 f. Dissertação (Mestrado em Comunicação) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2009. Disponível em: <https://tede2.pucsp.br/handle/handle/5240>. Acesso em: 30 maio 2022.

NEVES, B. C. Strategist librarian on social media: conceptual approach and possibilities of digital marketing in public libraries. **Informação & Sociedade: Estudos**, [s. l.], v. 28, n. 3, 2018. DOI: [10.22478/ufpb.1809-4783.2018v28n3.39203](https://doi.org/10.22478/ufpb.1809-4783.2018v28n3.39203). Acesso em: 27 maio 2022.

NEVES, E. C.; LONGO, R. M. J. Atuação do profissional da informação na gestão do conhecimento. **Revista de Biblioteconomia de Brasília**, Brasília, v. 23-24, n. 2, 2000. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/73314>. Acesso em: 27 maio 2022.

NHACUONGUE, J. A.; DUTRA, M. L. A terminologia em sistemas de recuperação da informação baseada na wordnet.pt. **Informação & Sociedade: Estudos**, [s. l.], v. 30, n. 2, 2020. DOI: [10.22478/ufpb.1809-4783.2020v30n2.50756](https://doi.org/10.22478/ufpb.1809-4783.2020v30n2.50756). Disponível em: <https://brapci.inf.br/index.php/res/v/147979>. Acesso em: 29 maio 2022.

OBUKHOV A.; KRASNYANSKIY M.; NIKOLYUKIN M. Implementation of Decision Support Subsystem in Electronic Document Systems Using Machine Learning Techniques. In: **2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)**, Vladivostok, Russia, 04 out. 2019. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85078569216&doi=10.1109%2fFarEastCon.2019.8934879&origin=inward&txGid=d0bf67a8cdfbcf1b827b97450d62fbca>. Acesso em: Acesso em 06 ago. 2024.

ONAN, A.; ATIK, E.; YALÇIN, A. Machine Learning Approach for Automatic Categorization of Service Support Requests on University Information Management System. In: **International Conference on Intelligent and Fuzzy Systems - INFUS 2020**, Istanbul, Turkey, v. 1197. 21-23 Jul. 2020. Disponível em: https://www.scopus.com/record/display.uri?eid=2-s2.0-85088741109&doi=10.1007%2f978-3-030-51156-2_132&origin=inward&txGid=77fc0a67c502c7a13f249133bc044503. Acesso em: 06 ago.2024.

PEARCE-MOSES, R. **A glossary of archival and records terminology**. Chicago: The Society of American Archivists, 2005.

RAUTENBERG, S.; CARMO, P. R. V. *Big Data* e Ciência de Dados: complementariedade conceitual no processo de tomada de decisão. **Brazilian Journal of Information Studies: Research Trends**, [s. l.], v.13, n.1, p.56-67. 2019. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=6983493>. Acesso em: 06 ago.2023.

RICHARDSON, R. J.; PERES, J. A. S.; WANDERLEY, J. C. V.; CORREIA, L. M.; M. H. M. **Pesquisa social: métodos e técnicas**. 3. ed. São Paulo: Atlas, 2007. 334 p.

RUIZ, M.; ROMÁN, C.; GARRIDO, A. L.; MENA, E. uAIS: An experience of increasing performance of NLP information extraction tasks from legal documents in an electronic document management system. In: **22nd International Conference on Enterprise Information Systems - ICEIS 2020**, v. 1, p. 189-196, Virtual, Online, 5-7 maio 2020. Disponível em: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85090796518&origin=inward&txGid=fc8cbd37a03d262d61ba5a7e4d1d8b2d>. Acesso em: 06 ago.2024.

RUSSELL, S. J.; NORVIG, P. **Inteligência Artificial**. 2. ed. Rio de Janeiro, RJ: Elsevier, 2004.

SAARIKKO, T.; WESTERGRENN, U. H.; BLOMQUIST, T. Digital transformation: Five recommendations for the digitally conscious firm. **Business Horizons**, [s. l.], v. 63, n. 6, p.825-839, nov./dez. 2020.

SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S. Tecnologias da Web Semântica aplicadas à organização do conhecimento: padrão SKOS para construção e uso de vocabulários controlados descentralizados. In: José Augusto Chaves Guimarães; Vera Dodebei. (Org.). **Organização do Conhecimento e Diversidade Cultural**. 1.ed. Marília: Fundepe, v. 3, p.224-233, 2015. Disponível em: <https://isko.org.br/wp-content/uploads/2021/05/Proceedings-ISKO-Brasil-2015.pdf>. Acessado em 29 maio 2022.

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, P. L. A. C.; VIDOTTI, S. A. B. G. Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação? **Datagramazero**, [s. l.], v. 10, n. 3, p. 1-10, jun. 2009. Disponível em: <http://dgz.org.br/jun09/Art_02.htm>. Acesso em: 08 jul. 2024.

SANTOS, F. G. S.; SILVA, M. N.; BARI, V. A. O impacto da gestão documental nos arquivos de engenharia e arquitetura: uma análise do arquivo da DIPOP do Instituto Federal de Educação, Ciência e Tecnologia de Sergipe. **AtoZ: novas práticas em informação e conhecimento**, [s. l.], v. 10, n. 1, p. 85-93, 2021. DOI: <http://dx.doi.org/10.5380/atoz.v10i1.75160>.

SAMBETBAYEVA, M., KUSPANOVA, I., YERIMBETOVA, A., SERIKBAYEVA, S., BAUYRZHANOVA, S. Development Of Intelligent Electronic Document Management System Model Based On Machine Learning Methods. **Eastern-European Journal**

Of Enterprise Technologies, v.1, n.2, p. 68–76. DOI:
<https://doi.org/10.15587/1729-4061.2022.251689>

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n.1, p.41-62, jan./jun. 1996.

SETZER, V. W. Dado, informação, conhecimento e competência. **DataGramZero**, [s. l.], v. 0, p.A01, 1999. Disponível em:
<http://www.brapci.inf.br/index.php/article/download/14562>. Acesso em: 27 jun. 2022.

SICHMAN, J. S. Inteligência Artificial e sociedade: avanços e riscos. **Estudos Avançados**, [s. l.], v. 35, n. 101, p.37-50, abr. 2021. FapUNIFESP (SciELO).
<http://dx.doi.org/10.1590/s0103-4014.2021.35101.004>. Disponível em:

SILVA, N. B. X.; NATHANSOHN, B. M. análise da produção científica em Inteligência Artificial na área da Ciência da Informação no Brasil. In: **XIX Encontro nacional de pesquisa em Ciência da Informação (ENANCIB 2018)**. Londrina, p. 111-126. 2018.

SIQUEIRA, I. S. P.; PEREIRA, A. E. C. Perspectivas de aplicação da inteligência artificial à biblioteconomia e à ciência da informação. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo. v. 22, n. 1/2, p.39-80, jan./jun. 1989.

STANFORD UNIVERSITY. Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence: **Report of the 2015-2016 Study Panel**. Stanford, CA: Stanford University, 2016. Disponível em:
https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/ai100report10032016fnl_singles.pdf. Acesso em: 10 jun. 2024.

TAKEUCHI, H., NONAKA, I. **Gestão do Conhecimento**. São Paulo: Bookman, 2008.

TKACHENKO, A. L.; DENISOVA, L. A. Designing an information system for the electronic document management of a university: Automatic classification of documents. In: **15th International Scientific and Technical Conference: Applied Mechanics and Systems Dynamics, AMSD 2021, Virtual, Online, 9-11 nov. 2021**. DOI:10.1088/1742-6596/2182/1/012035. Disponível em:
<https://www.scopus.com/record/display.uri?eid=2-s2.0-85127676818&doi=10.1088%2f1742-6596%2f2182%2f1%2f012035&origin=inward&txGid=d9c11035561cc71386f164f3f1cadb2a>. Acesso em: 06 ago.2024.

VALENTIM, M. L. P. (Org) **Métodos Qualitativos de pesquisa em Ciência da Informação**. São Paulo: Polis, 2005.

VAJJALA, S.; MAJUMDER, B.; GUPTA, A.; SURANA, H. Practical Natural Language Processing: Comprehensive Guide to Building Real-World NLP Systems. United States of America: **O'Reilly Media**, 2020.

VÉRAS NETO, C. N.; SILVA, M. B.; DIAS, G. A. Avaliação da implantação do GED nos arquivos da UNIMED/NNE como forma de recuperação da informação.

Biblionline, v. 6, n. 1, 2010. Disponível em:

<https://periodicos.ufpb.br/ojs2/index.php/biblio/article/view/4908>. Acesso em: 20 jun. 2023.

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.

WEISS, M. C. Sociedade sensoriada: a sociedade da transformação digital. **Estudos Avançados** [online], [s. l.], v. 33, n. 95, p.203-214, 2019. Disponível em:

<https://doi.org/10.1590/s0103-4014.2019.3395.0013>. Acesso em: 9 set. 2022

WEITZEN, H. S. **O poder da informação: como transformar a informação que você domina em um negócio lucrativo**. São Paulo: Markon, 1991.

WILKINS, J. Document Management vs. Records Management: What's the Difference?. **AIIM - Association for Intelligent Information Management**. 19 nov. 2019. Disponível em:

<https://info.aiim.org/aiim-blog/document-management-vs.-records-management>. Acesso em 16 jul. 2024.

WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO). **WIPO Technology Trends 2019: Artificial Intelligence**. Intellectual Property Office, Geneva, p.158.

2019, <https://www.wipo.int/publications/en/details.jsp?id=4386>. Acesso em 11 jun. 2024.

ZEROUAL, I.; LAKHOAJA, A. Data Science in light of natural language processing: an overview. **Procedia Computer Science**, [s. l.], v. 127, p.82–91, 2018.

APÊNDICE 1 – Prototipagem da *String* – IA, CD e GED

PROTOTIPAGEM STRING		PESQUISA POR TERMOS DE INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS PARA GESTÃO ELETÔNICA DE DOCUMENTOS							
BUSCA PELOS TERMOS	BUSCA	TERMOS + STRING GED		Pesquisa sequencial <i>String</i> GED					
TERMOS	TERMO PARA PESQUISA	SCO	WOS	Sequencia	Termos inclusos	STRING	SCO	WOS	DESCONSIDRAR (contemplado no pesquisa anterior)
artificial intelligence	"artificial intelligence"	16	9	T01	T1	"artificial intelligence"	16	9	
data science	"data science"	2	-	T02	T1-T2	"artificial intelligence" OR "data science"	18	9	
classification	classification	44	18	T03	T1-T3	"artificial intelligence" OR "data science" OR classification	60	25	
data mining	"data mining"	19	4	T04	T1-T4	"artificial intelligence" OR "data science" OR classification OR "data mining"	76	29	
connections	connections	17	6	T05	T1-T5	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections	92	35	
character recognition	"character recognition"	16	6	T06	T1-T6	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition"	103	41	
machine learn*	"machine learn**"	16	4	T07	T1-T7	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**"	111	44	
ontology	ontology	11	3	T08	T1-T8	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology	117	46	
neural network*	"neural network**"	8	3	T09	T1-T9	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**"	122	46	
decision support system	"decision support system"	8	1	T10	T1-T10	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system"	125	47	
allocation	allocation	7	4	T11	T1-T11	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation	131	50	
pattern recogni*	"pattern recogni**"	7	3	T12	T1-T12	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**"	133	50	
natural language processing	"natural language proces"	6	4	T13	T1-T13	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing"	137	52	
feature extraction	"feature extraction"	4	3	T14	T1-T14	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction"	137	52	DESCONSIDRAR
text analy* (analysis e analytic)	"text analy**"	4	2	T15	T1-T15	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**"	137	52	DESCONSIDRAR
fuzzy (environment*, logic, number, set*, syst* fuzzy*		4	1	T16	T1-T16	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy*"	139	52	
learning algorithm	"learning algorithm"	4	-	T17	T1-T17	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm"	139	52	DESCONSIDRAR
robotic*	robotic*	4	-	T18	T1-T18	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic*	142	52	
dialogue	dialogue	3	3	T19	T1-T19	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue	145	54	
predictive analy* (analysis e analytic)	"predictive analy**"	3	1	T20	T1-T20	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy**"	146	54	
recommend* system*	"recommend* system**"	3	1	T21	T1-T21	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy**" OR "recommend* system**"	147	55	
biometrics	biometrics	3	-	T22	T1-T22	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy**" OR "recommend* system**" OR biometrics	149	55	

PROTOTIPAGEM STRING		PESQUISA POR TERMOS DE INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS PARA GESTÃO ELETÔNICA DE DOCUMENTOS							
BUSCA PELOS TERMOS	BUSCA	TERMOS + STRING GED		Pesquisa sequencial <i>String</i> GED					
TERMOS	TERMO PARA PESQUISA	SCO	WOS	Sequencia	Termos inclusos	STRING	SCO	WOS	DESCONSIDERAR (contemplado no pesquisa anterior)
computer vision	"computer vision"	3	-	T23	T1-T23	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy" OR "learning algorithm" OR robotic" OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision"	151	55	
embedding*	embedding*	3	-	T24	T1-T24	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy" OR "learning algorithm" OR robotic" OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding*	154	55	
intelligent system	"intelligent system"	3	-	T25	T1-T25	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy" OR "learning algorithm" OR robotic" OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system"	155	55	
knowledge representation	"knowledge representati	3	-	T26	T1-T26	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy" OR "learning algorithm" OR robotic" OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation"	157	55	
clustering	clustering	2	3	T27	T1-T27	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy" OR "learning algorithm" OR robotic" OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering	159	57	
expert system*	"expert system**"	2	1	T28	T1-T28	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy" OR "learning algorithm" OR robotic" OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system**"	161	58	
information extraction	"information extraction"	2	1	T29	T1-T29	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy" OR "learning algorithm" OR robotic" OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system**" OR "information extraction"	161	58	DESCONSIDERAR
association rule	"association rule"	2	-	T30	T1-T30	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy" OR "learning algorithm" OR robotic" OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system**" OR "information extraction" OR "association rule"	163	58	
augmented reality	"augmented reality"	2	-	T31	T1-T31	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy" OR "learning algorithm" OR robotic" OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system**" OR "information extraction" OR "association rule" OR "augmented reality"	164	58	

PROTOTIPAGEM STRING		PESQUISA POR TERMOS DE INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS PARA GESTÃO ELETÔNICA DE DOCUMENTOS							
BUSCA PELOS TERMOS	BUSCA	TERMOS + STRING GED		Pesquisa sequencial <i>String</i> GED					
TERMOS	TERMO PARA PESQUISA	SCO	WOS	Sequencia	Termos inclusos	STRING	SCO	WOS	DESCONSIDERAR (contemplado no pesquisa anterior)
image segmentation	"image segmentation"	2	-	T32	T1-T32	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy**" OR "recommend* system**" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system**" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation"	165	58	
latent semantic analysis	"latent semantic analysis"	2	-	T33	T1-T33	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy**" OR "recommend* system**" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system**" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis"	165	58	DESCONSIDERAR
speech processing	"speech processing"	2	-	T34	T1-T34	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy**" OR "recommend* system**" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system**" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing"	165	58	DESCONSIDERAR
support vector machine*	"support vector machine"	2	-	T35	T1-T35	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy**" OR "recommend* system**" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system**" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine**"	165	58	DESCONSIDERAR
vector machine*	"vector machine**"	2	-	T36	T1-T36	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy**" OR "recommend* system**" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system**" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine**" OR "vector machine**"	165	58	DESCONSIDERAR
genetic algorithm	"genetic algorithm"	2	-	T37	T1-T37	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy**" OR "recommend* system**" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system**" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine**" OR "vector machine**" OR "genetic algorithm"	166	58	
decision tree	"decision tree"	1	-	T38	T1-T38	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn**" OR ontology OR "neural network**" OR "decision support system" OR allocation OR "pattern recogni**" OR "natural language processing" OR "feature extraction" OR "text analy**" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy**" OR "recommend* system**" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system**" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine**" OR "vector machine**" OR "genetic algorithm" OR "decision tree"	166	58	DESCONSIDERAR

PROTOTIPAGEM STRING		PESQUISA POR TERMOS DE INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS PARA GESTÃO ELETÔNICA DE DOCUMENTOS							
BUSCA PELOS TERMOS	BUSCA	TERMOS + STRING GED		Pesquisa sequencial <i>String</i> GED					
TERMOS	TERMO PARA PESQUISA	SCO	WOS	Sequencia	Termos inclusos	STRING	SCO	WOS	DESCONSIDERAR (contemplado no pesquisa anterior)
deep learning	"deep learning"	1	-	T39	T1-T39	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning"	166	58	DESCONSIDERAR
descriptive model*	"descriptive model**"	1	-	T40	T1-T40	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model**"	167	58	
inductive reasoning	"inductive reasoning"	1	-	T41	T1-T41	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model**" OR "inductive reasoning"	167	58	DESCONSIDERAR
latent dirichlet	"latent dirichlet"	1	-	T42	T1-T42	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model**" OR "inductive reasoning" OR "latent dirichlet"	167	58	DESCONSIDERAR
learning model	"learning model"	1	-	T43	T1-T43	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model**" OR "inductive reasoning" OR "latent dirichlet" OR "learning model"	167	58	DESCONSIDERAR
machine intelligen*	"machine intelligen**"	1	-	T44	T1-T44	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy* OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model**" OR "inductive reasoning" OR "latent dirichlet" OR "learning model" OR "machine intelligen**"	167	58	DESCONSIDERAR

PROTOTIPAGEM STRING		PESQUISA POR TERMOS DE INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS PARA GESTÃO ELETÔNICA DE DOCUMENTOS							
BUSCA PELOS TERMOS	BUSCA	TERMOS + STRING GED		Pesquisa sequencial <i>String</i> GED					
TERMOS	TERMO PARA PESQUISA	SCO	WOS	Sequencia	Termos inclusos	STRING	SCO	WOS	DESCONSIDERAR (contemplado no pesquisa anterior)
morphology	morphology	1	-	T45	T1-T45	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy*" OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend*" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model*" OR "inductive reasoning" OR "latent dirichlet" OR "learning model" OR "machine intelligen*" OR morphology	167	58	DESCONSIDERAR
multi agent system	"multi agent system"	1	-	T46	T1-T46	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy*" OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend*" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model*" OR "inductive reasoning" OR "latent dirichlet" OR "learning model" OR "machine intelligen*" OR morphology OR "multi agent system"	167	58	DESCONSIDERAR
multilayer perceptron	"multilayer perceptron"	1	-	T47	T1-T47	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy*" OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend*" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model*" OR "inductive reasoning" OR "latent dirichlet" OR "learning model" OR "machine intelligen*" OR morphology OR "multi agent system" OR "multilayer perceptron"	167	58	DESCONSIDERAR
random-forest*	"random-forest"	1	-	T48	T1-T48	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy*" OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend*" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model*" OR "inductive reasoning" OR "latent dirichlet" OR "learning model" OR "machine intelligen*" OR morphology OR "multi agent system" OR "multilayer perceptron" OR "random-forest"	167	58	DESCONSIDERAR
speech recognition	"speech recognition"	1	-	T49	T1-T49	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy*" OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend*" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model*" OR "inductive reasoning" OR "latent dirichlet" OR "learning model" OR "machine intelligen*" OR morphology OR "multi agent system" OR "multilayer perceptron" OR "random-forest*" OR "speech recognition"	167	58	DESCONSIDERAR

PROTOTIPAGEM STRING		PESQUISA POR TERMOS DE INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS PARA GESTÃO ELETÔNICA DE DOCUMENTOS							
BUSCA PELOS TERMOS	BUSCA	TERMOS + STRING GED		Pesquisa sequencial <i>String</i> GED					
TERMOS	TERMO PARA PESQUISA	SCO	WOS	Sequencia	Termos inclusos	STRING	SCO	WOS	DESCONSIDERAR (contemplado no pesquisa anterior)
target function	"target function"	1	-	T50	T1-T50	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy*" OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend* system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model*" OR "inductive reasoning" OR "latent dirichlet" OR "learning model" OR "machine intelligen*" OR morphology OR "multi agent system" OR "multilayer perceptron" OR "random-forest*" OR "speech recognition" OR "target function"	167	58	DESCONSIDERAR
text recognition	"text recognition"	1	-	T51	T1-T51	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy*" OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend* system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model*" OR "inductive reasoning" OR "latent dirichlet" OR "learning model" OR "machine intelligen*" OR morphology OR "multi agent system" OR "multilayer perceptron" OR "random-forest*" OR "speech recognition" OR "target function" OR "text recognition"	167	58	DESCONSIDERAR
probabilist* method*	"probabilist* method*"	1	-	T52	T1-T52	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy*" OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend* system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model*" OR "inductive reasoning" OR "latent dirichlet" OR "learning model" OR "machine intelligen*" OR morphology OR "multi agent system" OR "multilayer perceptron" OR "random-forest*" OR "speech recognition" OR "target function" OR "text recognition" OR "probabilist* method*"	168	58	
rule learning	"rule learning"	-	1	T53	T1-T53	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy*" OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend* system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model*" OR "inductive reasoning" OR "latent dirichlet" OR "learning model" OR "machine intelligen*" OR morphology OR "multi agent system" OR "multilayer perceptron" OR "random-forest*" OR "speech recognition" OR "target function" OR "text recognition" OR "probabilist* method*" OR "rule learning"	168	58	DESCONSIDERAR

PROTOTIPAGEM STRING		PESQUISA POR TERMOS DE INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS PARA GESTÃO ELETÔNICA DE DOCUMENTOS							
BUSCA PELOS TERMOS	BUSCA	TERMOS + STRING GED		Pesquisa sequencial <i>String</i> GED					
TERMOS	TERMO PARA PESQUISA	SCO	WOS	Sequencia	Termos inclusos	STRING	SCO	WOS	DESCONSIDERAR (contemplado no pesquisa anterior)
self learning	"self learning"	-	1	T54	T1-T54	"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "feature extraction" OR "text analy*" OR "fuzzy*" OR "learning algorithm" OR robotic* OR dialogue OR "predictive analy*" OR "recommend*" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "information extraction" OR "association rule" OR "augmented reality" OR "image segmentation" OR "latent semantic analysis" OR "speech processing" OR "support vector machine*" OR "vector machine*" OR "genetic algorithm" OR "decision tree" OR "deep learning" OR "descriptive model*" OR "inductive reasoning" OR "latent dirichlet" OR "learning model" OR "machine intelligen*" OR morphology OR "multi agent system" OR "multilayer perceptron" OR "random-forest*" OR "speech recognition" OR "target function" OR "text recognition" OR "probabilist* method*" OR "rule learning" OR "self learning"	168	59	

RESULTADO FINAL - STRING GED

"artificial intelligence" OR "data science" OR classification OR "data mining" OR connections OR "artificial intelligence" OR "character recognition" OR "machine learn*" OR ontology OR "neural network*" OR "decision support system" OR allocation OR "pattern recogni*" OR "natural language processing" OR "fuzzy*" OR robotic* OR dialogue OR "predictive analy*" OR "recommend*" system*" OR biometrics OR "computer vision" OR embedding* OR "intelligent system" OR "knowledge representation" OR clustering OR "expert system*" OR "association rule" OR "augmented reality" OR "image segmentation" OR "genetic algorithm" OR "descriptive model*" OR "probabilist* method*" OR "self learning"

APÊNDICE 2 – Ficha de Anotações

PALAVRA-CHAVE			
AUTOR		ANO	20
TÍTULO		FOLHA	1/1

RESUMO

--

Avaliação de Relevância da String de Busca

Os termos buscados na string apareceram:

Título		Palavra-chave do Autor		Palavra-chave da base		Resumo	
GED	IA/CD	GED	IA/CD	GED	IA/CD	GED	IA/CD

Impressões Iniciais

TÓPICOS	ANOTAÇÕES/PERGUNTAS
Dados para referência	
Observações:	
Desafios enfrentados:	
Aplicação prática:	
Benefícios:	

Método, Técnicas, Ferramentas, Modelos etc. que podem ser utilizados e implantado na área de GED

TÓPICOS	ANOTAÇÕES/PERGUNTAS
Etapas:	
Ferramentas:	
Métodos:	
Técnicas:	
Modelos:	
Algoritmos:	
Arquiteturas:	
Frameworks e bibliotecas:	
Não especificado:	
Conclusão	

**APÊNDICE 3 – Artigo: MAGALHÃES, E. V. S.; ZANIRO, D. L.; SOARES, V. R.
M. AMARAL, R. M.; NEVES, B. C. Inteligência Artificial e Ciência de Dados
na Gestão de Documentos: uma Revisão Sistemática de Literatura**

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

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 qualiquantitativa. 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).

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

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

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

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.

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

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?).

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

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 conseqüentemente 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 (ABNT). NBR ISO 9001:2015: Sistemas de gestão da qualidade - Requisitos. Rio de Janeiro, 2015.

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

BOELL, S. K.; CEZEC-KECMANOVIC, D. Are systematic reviews better, less biased and of higher quality? **19th European Conference on Information Systems (ECIS 2011)**, [n. p.], *Paper 223*, Helsinki, jun. 2011. Disponível em: https://www.researchgate.net/publication/221407655_Are_systematic_reviews_better_less_biased_and_of_higher_quality. Acesso em: 08 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, abr. 2007. Disponível em: <https://doi.org/10.1016/j.jss.2006.07.009>. Acesso em: 08 jun. 2024.

CAO, L. Data Science: A Comprehensive Overview. *Acm Computing Surveys*, [s. l.], v. 50, n. 3, p. 1-42, 29 jun. 2017. **Association for Computing Machinery (ACM)**. Disponível em: <http://dx.doi.org/10.1145/3076253>. Acesso em: 07 jan. 2024.

DYBA, T.; DINGSOYR, T.; HANSSEN, G. K. Applying Systematic Reviews to Diverse Study Types: an experience report. **First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)**, [s. l.], p. 225-234, set. 2007. IEEE. 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. IET. Disponível em: <https://doi.org/10.1049/iet-sen.2013.0029>. Acesso em: 08 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. (Eds.). **Indicadores de Ciência, Tecnologia e Inovação em São Paulo 2010**. São Paulo: FAPESP, p. 4.1-4.71, 2010.

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. Wiley. 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**, [s. l.], v. 6, n. 11, p. 72-85, 2010.

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

KITCHENHAM, B. **Procedures for Performing Systematic Reviews**. Joint Technical Report Software Engineering Group, Keele University (TR/SE-0401), United Kingdom: Empirical Software Engineering and Australia: National ICT Australia Ltd, (0400011T.1), 2004.

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

KITCHENHAM, B.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. **Technical Report EBSE 2007-001**. Keele University and Durham University Joint Report, 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: Eduep, 2008.

NAKAMITI, E. K. **Agentes inteligentes artificiais**. 2009. 91 f. Dissertação (Mestrado em Comunicação) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2009. Disponível em: <https://tede2.pucsp.br/handle/handle/5240>. Acesso em: 05 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.

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: ABRQ, 2005. 223 p.

SICHMAN, J. S. Inteligência Artificial e sociedade: avanços e riscos. **Estudos Avançados**, [online], [s. l.], v. 35, n. 101, p. 37-50, abr. 2021. FapUNIFESP (SciELO). Disponível em: <http://dx.doi.org/10.1590/s0103-4014.2021.35101.004>. Acesso em: 05 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: 05 mar. 2024.

VINAGRE, J.; MONIZ, N. Inteligência Artificial: riscos e promessas. **Revista Ciência Elementar**, Porto, v. 8, n. 4, p. 1-9, 2020. Disponível em:

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

<https://rce.casadasciencias.org/rceapp/art/2020/052/>. Acesso em: 10 mar. 2024. DOI: 10.24927/rce2020.052

WEISS, M. C. Sociedade sensoriada: a sociedade da transformação digital. **Estudos Avançados** [online], [s. l.], v. 33, n. 95, p. 203-214, 2019. Disponível em: <https://doi.org/10.1590/s0103-4014.2019.3395.0013>. Acesso em: 09 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>.