

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

SAMARA SIVIRINO MARQUES

**O USO DE *CHATBOTS* PARA AUXILIAR A RECUPERAÇÃO DA INFORMAÇÃO  
EM REPOSITÓRIOS INSTITUCIONAIS**

São Carlos, SP  
2024

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

SAMARA SIVIRINO MARQUES

**O USO DE *CHATBOTS* PARA AUXILIAR A RECUPERAÇÃO DA INFORMAÇÃO  
EM REPOSITÓRIOS INSTITUCIONAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos, como parte das exigências para a obtenção do título de Mestre em Ciência da Informação.

**Área:** Conhecimento, Tecnologia e Inovação.

**Linha:** Tecnologia, Informação e Representação.

**Orientador:** Prof. Dr. Januário Albino Nhacuongue

São Carlos, SP  
2024

Marques, Samara Sivirino

O uso de chatbots para auxiliar a recuperação da  
informação em repositórios institucionais / Samara  
Sivirino Marques -- 2024.  
329f.

Dissertação (Mestrado) - Universidade Federal de São  
Carlos, campus São Carlos, São Carlos

Orientador (a): Januário Albino Nhacuongue

Banca Examinadora: Paula Regina Dal'Evedove,  
Raimunda Fernanda dos Santos

Bibliografia

1. Chatbots. 2. Processamento de Linguagem Natural. 3.  
Repositórios Institucionais. I. Marques, Samara Sivirino.  
II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática  
(SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Ronildo Santos Prado - CRB/8 7325

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 Samara Svirino Marques, realizada em 30/04/2024:

---

Prof. Dr. Januário Albino Nhacuongue  
Universidade Federal de São Carlos

---

Profa. Dra. Paula Regina Dal'Evedove  
Universidade Federal de São Carlos

---

Profa. Dra. Raimunda Fernanda dos Santos  
Universidade Federal do Rio Grande do Norte (UFRN)

## AGRADECIMENTOS

Minha sincera gratidão a Deus por Sua benevolência em me conceder saúde e força para superar todos os desafios enfrentados ao longo da vida e durante o curso. Em momentos de dificuldade, sua presença foi um farol de esperança, guiando-me através das adversidades e fortalecendo-me para seguir em frente.

Dedico um especial agradecimento ao meu marido, Aparecido, que me incentivou desde o momento da inscrição até o momento da defesa, compreendeu a minha ausência como esposa e me apoiou incondicionalmente durante os dias mais desafiadores, seu amor e apoio foram a luz nos momentos mais cinzentos.

Expresso meus sinceros agradecimentos ao IFSP Campus Hortolândia e à equipe da biblioteca, em especial ao meu chefe, Élcio José da Costa, por todo o apoio oferecido ao longo desta jornada acadêmica.

Desejo expressar minha mais profunda gratidão ao meu orientador, Prof. Dr. Januário Albino Nhacuongue, pela sua orientação excepcional, apoio incansável e confiança depositada ao longo deste trabalho. Sua vasta experiência e dedicação foram pilares fundamentais para o desenvolvimento e a conclusão bem-sucedida desta pesquisa. Agradeço sinceramente por sua orientação perspicaz, que foi fundamental para moldar cada aspecto deste estudo e para meu crescimento acadêmico e profissional.

Aos ilustres membros da banca, expresso meu profundo apreço pela generosa dedicação de seu tempo, pela cuidadosa análise da dissertação e pelas enriquecedoras trocas de experiências. Sua participação atenta e valiosas contribuições foram essenciais para o aprimoramento deste trabalho, enriquecendo sua qualidade e contribuindo significativamente para o meu crescimento acadêmico.

Quero expressar minha mais profunda gratidão a todos aqueles que, direta ou indiretamente, contribuíram para minha formação acadêmica, cada gesto de apoio, conselho e incentivo foi fundamental para minha jornada, moldando minha trajetória e me impulsionando em direção ao sucesso. Por isso, é com sincera gratidão que reconheço a importância de cada indivíduo que cruzou meu caminho, pois sem eles minha jornada não seria a mesma.

## RESUMO

O estudo busca analisar, a partir da literatura nacional e internacional, como o PLN e os chatbots podem contribuir para a recuperação da informação em repositórios institucionais. Três objetivos específicos foram estabelecidos: identificar e mapear os principais problemas nas pesquisas sobre PLN e chatbots na área de recuperação de informação; analisar as soluções propostas em cada estudo, destacando abordagens para aprimorar a recuperação de informação; caracterizar as aplicações específicas de cada estudo no sentido de aprimorar a recuperação de informação em repositórios institucionais, destacando as contribuições práticas e potenciais benefícios para o processo de busca e recuperação de informação. A pesquisa adota a metodologia de análise de conteúdo. Por fim, a análise categorial temática foi empregada para tratar os resultados obtidos, proporcionando uma compreensão mais profunda das tendências e padrões identificados. As categorias de análise são estabelecidas a partir do espaço quadrimensional em que o conceito relevância pode ser formalmente definido, para efeitos de avaliação de sistemas de recuperação de informação. Sendo assim, a pesquisa baseou-se em duas categorias, recursos de informação e representação do problema do usuário. A pesquisa analisou 263 artigos publicados entre 2019 e 2023, selecionados por meio da base de dados *Web of Science*. Os resultados destacam a importância do uso de PLN e *chatbots* na otimização da recuperação de informação em repositórios institucionais. Essas tecnologias têm o potencial de transformar a forma como os usuários interagem e acessam informações, proporcionando uma busca mais eficiente, personalizada e acessível. A partir dos resultados, pode-se inferir que um repositório direcionado por tarefas e a construção de chatbots a partir de ferramentas já existentes, intuitivas e de código aberto, é uma forma de aplicação das pesquisas analisadas. Conclui-se que há um cenário promissor no uso de PLN e *chatbots* para aprimorar a recuperação da informação em repositórios institucionais, simplificando o processo de busca e tornando-o mais eficiente para os usuários. Para futuras pesquisas, há oportunidades de explorar técnicas avançadas de PLN para melhorar a compreensão de consultas complexas pelos *chatbots* e personalizar a experiência do usuário, adaptando o estilo de interação e fornecendo recomendações de conteúdo relevante.

**Palavras-chave:** *Chatbots*; Processamento de Linguagem Natural; Repositórios Institucionais; Recuperação da Informação; Ciência da Informação.

## ABSTRACT

The study aims to analyze, based on national and international literature, how NLP and chatbots can contribute to the retrieval of information in institutional repositories. Three specific objectives were established: identify and map the main problems in research on NLP and chatbots in the area of information retrieval; analyze the solutions proposed in each study, highlighting approaches to improve information retrieval; characterize the specific applications of each study in order to improve information retrieval in institutional repositories, highlighting the practical contributions and potential benefits to the information search and retrieval process. The research adopts the content analysis methodology. Thematic categorical analysis was used to analyze the results obtained, providing a deeper understanding of the trends and patterns identified. The analysis categories were established based on the four-dimensional space in which the concept of relevance can be formally defined, for the purposes of evaluating information retrieval systems. Therefore, the research was based on two categories: information resources and representation of the user's problem. The research analyzes 263 articles published between 2019 and 2023, selected through the Web of Science database. The results highlight the importance of using NLP and chatbots in optimizing information retrieval in institutional repositories. These technologies have the potential to transform the way users interact and access information, providing a more efficient, personalized and accessible search. From the results, it can be inferred that a task-driven repository and the construction of chatbots based on existing, intuitive and open source tools is a way of applying the research analyzed. It is concluded that there is a promising scenario in the use of NLP and chatbots to improve information retrieval in institutional repositories, simplifying the research process and making it more efficient for users. For future research, there are opportunities to explore advanced NLP techniques to improve chatbots' understanding of complex queries and personalize the user experience by adapting the interaction style and providing relevant content recommendations.

**Keywords:** Chatbots; Natural Language Processing; Institutional Repositories; Information Retrieval; Information Science.

## LISTA DE FIGURA

Figura 1: Processo total de Indexação .....	65
Figura 2: Busca no Del.icio.us.....	75

## LISTA DE QUADRO

Quadro 1: Categorias de análise temática dos estudos .....	86
--	----

## LISTA DE TABELAS

Tabela 1: Linguagem natural: vantagens e desvantagens .....	63
Tabela 3: Alguns estudos analisados nos resultados .....	91
Tabela 3: Síntese de possíveis aplicações de <i>chatbots</i> em repositórios institucionais .....	119

## LISTA DE ABREVIATURAS

BERT	<i>Bidirectional Encoder Representations from Transformers</i>
BiLSTM	<i>Bidirectional Long Short-Term Memory</i>
CAPES	Catálogo de Teses e Dissertações da CAPES
CC	Ciência da Computação
CI	Ciência da Informação
DBN	<i>Deep Belief Network</i>
DSR	<i>Design Science Research</i>
FAQ	Perguntas Frequentes
HMN	Modelo de Rede de Correspondência Hierárquica
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
OAI	<i>Open Access Initiative</i>
PLN	Processamento de Linguagem Natural
MIT	<i>Massachusetts Institute of Technology</i>
MS MARCO	<i>Microsoft MAchine Reading COmprehension</i>
NER	Reconhecimento de Entidades Nomeadas
RNAs	Redes Neurais Artificiais
SBERT	<i>Sentence-BERT</i>
SGC	Sistemas de Gerenciamento de Conhecimento
SMART	<i>System for the Manipulation and Retrieval of Text</i>
TICs	Tecnologias de Informação e Comunicação
UFSCar	Universidade Federal de São Carlos
UTS	Universidade de Tecnologia de Sydney

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	13
1.1 Objetivo geral .....	18
1.2 Objetivos específicos .....	18
1.3 Justificativas.....	19
1.4 Estrutura do trabalho.....	23
<b>2 ESTUDOS RELACIONADOS</b> .....	25
<b>3 FUNDAMENTAÇÃO TEÓRICA</b> .....	32
3.1 Ciência da Informação e Inteligência Artificial .....	32
3.2 <i>Chatbots</i> e o Processamento de Linguagem Natural .....	34
3.3 Recuperação da informação .....	42
3.4 Repositório institucional .....	56
3.5 Indexação .....	61
3.6 Folksonomia.....	70
<b>4 MÉTODO E METODOLOGIA</b> .....	80
<b>5 ANÁLISE E DISCUSSÃO DOS RESULTADOS</b> .....	89
<b>CONSIDERAÇÕES FINAIS</b> .....	122
<b>REFERÊNCIAS</b> .....	125
<b>APÊNDICE A – Levantamento e análise dos resultados</b> .....	135

## 1 INTRODUÇÃO

A ubiquidade do fenômeno da informação, segundo Freire (2006), tem norteado diferentes esferas de atuação na sociedade, transformando a apropriação da informação e do conhecimento em forças produtivas. Para Rubin e Rubin (2020), esse fenômeno é característico da economia da informação, que começou com a transição das nações industriais para sociedades baseadas na informação, nas quais a informação passou a ser tratada como ativo de negócio. Essas transformações, por um lado, vêm afetando as práticas dos profissionais de informação, principalmente, no que tange à atribuição de valor aos serviços de informação oferecidos a pessoas e organizações. Por outro lado, vêm transformando domínios de atuação da Ciência da Informação (CI), com destaque para as seguintes áreas temáticas: bibliometria, catalogação, classificação, produção e comunicação científica, gestão da informação, recuperação de informação, mecanismos de busca e indústria da informação.

Desse modo, para Levy (2010), o hipertexto, enquanto técnica de transmissão e tratamento de mensagens, tornou-se um dos principais agentes de transformação da sociedade, pelos seus seis princípios abstratos: a) metamorfose – a extensão, a composição e o desenho do hipertexto estão em constante construção e renegociação em prol dos atores envolvidos; b) heterogeneidade – os nós e conexões de uma rede hipertextual são heterogêneas, isto é, envolvem uma hibridização de linguagens; c) multiplicidade e encaixe de escalas – qualquer nó ou conexão compõe e interfere na estrutura da rede; d) exterioridade – a propagação da organização da rede hipertextual depende de um exterior indeterminado; e) topologia – o curso dos acontecimentos no hipertexto é determinado pela própria estrutura da rede; f) mobilidade dos centros – a rede possui diversos centros móveis, com uma infinidade de ramificações.

A "necessidade de informação" representa a expressão de um conjunto de pessoas ou participantes da rede, os quais estabelecem padrões relacionados a situações ou atividades que compõem a resposta para essa necessidade. A partir da área do conhecimento, o coletivo determina o objetivo, os métodos próprios para recuperar a informação e, desta forma, cabe ao pesquisador, a partir da análise de seu domínio, definir percepções coletivas, "socialmente compartilhadas e vivenciadas pelos membros". No âmbito da gestão de informações, é incumbência dos

profissionais da informação aprimorar serviços e sistemas alinhados aos padrões coletivos (Araújo Júnior, 2007, p. 12).

Nesse contexto, podemos definir um sistema de informação como aquele que tem a informação como elemento central, visando principalmente à sua gestão. Exemplos tangíveis desses sistemas incluem bibliotecas e centros de informação (Marcial et al., 2015, p. 5). Diante disso, torna-se evidente que uma ferramenta crucial para atender a essas demandas informacionais são os repositórios institucionais.

Os repositórios institucionais desempenham um papel crucial na preservação e no acesso à memória acadêmica de uma instituição. Ao servirem como arquivos digitais centralizados, garantem a salvaguarda e acessibilidade dos documentos representativos do conhecimento acumulado ao longo do tempo, incluindo artigos científicos, teses e relatórios de pesquisa. Além disso, promovem o acesso aberto a esses materiais, democratizando o conhecimento e ampliando seu impacto. Assim, contribuem para a continuidade da pesquisa e desenvolvimento acadêmico, enquanto preservam a memória da instituição para as gerações presentes e futuras. Esses repositórios não apenas suprimem as necessidades informacionais, mas também se tornam elementos significativos na rede hipertextual e nos equipamentos coletivos da inteligência, conforme destacado por Levy (2010).

Assim, ao conectar a necessidade de aprimoramento de serviços de informação, a definição de sistema de informação e o papel vital dos repositórios institucionais revela-se uma abordagem integrada e abrangente para a gestão eficiente da informação.

Saracevic (1996) evidencia que a CI é uma ciência que analisa questões científicas, com foco em pesquisa e resolução de problemas em diferentes áreas, e pode ajudar a resolver questionamentos que exigem conhecimento mais específico da informação tanto em problemas sociais quanto em tecnologia da informação. Isso mostra que a informação gerada a cada segundo é vista por diferentes prismas e em diferentes mídias, comprovando a crescente versatilidade dentro do âmbito científico. Dessa forma, a informação adquire o propósito de fornecer soluções e melhorar a compreensão do mundo.

A CI tem participado de fato dessas mudanças, influenciada pelas novas tecnologias da informação, em especial a *Web*. Por muitos anos, a informação foi mantida cativa no universo simbólico divino. Entre manuscritos e prisões, chega-se à

era da Internet, onde a maioria dos textos é publicada na sua língua natural (Barreto, 2007).

A CI e a Ciência da Computação (CC) parecem ser as ciências mais engajadas na melhoria da qualidade da informação recuperada. A CI tem uma perspectiva mais metodológica e tenta estruturar dados e criar métodos e modelos que proporcionem um melhor armazenamento de informações, bem como métodos de estudo para adicionar semântica às informações para que possam ser aplicadas em processos de recuperação. Já a CC busca atuar na aplicação diretamente no desenvolvimento de técnicas computacionais, como algoritmos, que possam viabilizar os métodos propostos e pesquisados (Santarém Segundo, 2010, p. 26).

A Inteligência Artificial (IA), surgida como uma área interdisciplinar nos anos 1950, incorpora disciplinas como computação, cognição e aprendizado, aplicando-as à capacidade das máquinas de raciocinar e aprender (Martins, 2010). Sua versatilidade permite que ela faça contribuições significativas para a CI, abrangendo aprendizado de máquina, Processamento de Linguagem Natural (PLN), recuperação da informação, *chatbots*, mineração de dados, sistemas de recomendação, predição e ciência de dados. A relação da IA com o processamento e análise de grandes volumes de dados facilita a automação de tarefas complexas.

Martins (2010) explica que o PLN é um campo de pesquisa dentro da IA com o objetivo de habilitar computadores a compreenderem instruções em linguagem humana, simplificando a interação com interfaces de sistemas de recuperação de informação automatizados. No entanto, a interpretação nesse contexto é desafiadora devido ao contexto da informação, resultando em ambiguidades e divergências culturais, além das regras gramaticais.

Vajjala et al. (2020, p. 14) descrevem a IA como um ramo da CC voltado para a construção de sistemas capazes de executar tarefas que exigem inteligência humana, frequentemente chamada de "inteligência de máquina". O aprendizado de máquina é um subcampo da IA que se concentra no desenvolvimento de algoritmos capazes de aprender a realizar tarefas automaticamente com base em um grande número de exemplos, sem a necessidade de regras programadas manualmente.

Vajjala et al. (2020) também observam que, similarmente a outros sistemas iniciais de IA, as primeiras tentativas de desenvolver sistemas de PLN baseavam-se na construção de regras específicas para cada tarefa. Isso exigia que os desenvolvedores tivessem conhecimento no domínio para formular regras que

pudessem ser integradas a um programa ou sistema, como os *chatbots*. Esses sistemas também necessitavam de recursos como dicionários e tesouros, que eram normalmente acumulados e digitalizados ao longo do tempo.

As tecnologias integradas aos sistemas de IA abrangem diversas áreas, incluindo PLN, visão computacional, aprendizado de máquina, robótica, geração de voz, algoritmos de otimização, sistemas especialistas, *chatbots*, reconhecimento de objetos e biometria, bem como geração de linguagem natural.

A ideia da recuperação da informação que, segundo Saracevic (1996), caracterizou a essência da CI na sua origem, desenvolveu-se sob três vertentes: descrição intelectual da informação, especificação intelectual da busca e especificação dos sistemas e das técnicas a empregar. Ingwersen (1992) também defende que a recuperação da informação é o campo de pesquisa central da CI, cujo objetivo é investigar processos que possam facilitar a comunicação efetiva da informação

Baeza-Yates e Ribeiro-Neto (2013, p. 4) identificam o problema principal da recuperação da informação a partir do seguinte objetivo:

O objetivo principal de um sistema de recuperação da informação é recuperar todos os documentos que são relevantes à necessidade de informação do usuário e, ao mesmo tempo, recuperar o menor número possível de documentos irrelevantes.

A recuperação da informação está preocupada com processos envolvidos na representação, no armazenamento, na busca e na descoberta de informações que são relevantes para o usuário. Esses processos decorrem de interações entre os principais componentes de um Sistema de Recuperação da Informação (SRI). Os repositórios digitais que, segundo Sayão (2017), se tornaram parte essencial da infraestrutura mundial de pesquisa para apoiar processos de gestão e de curadoria de dados, estão inseridos nesta problemática.

Santos (2015) analisa parte da problemática envolvida nos processos de organização e tratamento das informações em ambientes digitais, mostrando que a participação direta dos usuários nesse processo, por meio da representação colaborativa da informação ou da folksonomia, ainda precisa ser aprofundada pelos estudos da área da CI. Ao analisar o uso de *chatbots* como sistemas auxiliares de recuperação da informação em repositórios institucionais, esta pesquisa se enquadra no bojo desses estudos.

Lynch (2003, p. 332) define repositórios institucionais como uma coleção de serviços que a instituição oferece à comunidade acadêmica para a gestão e disseminação de materiais digitais criado pela própria instituição e por seus membros. Ainda para o autor, o comprometimento da universidade com a gestão destes materiais é talvez mais importante do que o serviço propriamente em si.

Um repositório institucional é um reconhecimento que a vida intelectual e acadêmica das universidades será cada vez mais representada, documentada e compartilhada em formato digital, e que a responsabilidade primária das universidades é exercitar a administração destas riquezas: tanto para torna-los disponíveis quanto para preservá-los. Um repositório institucional é o meio que cada uma das universidades irá afirmar esta responsabilidade tanto para os membros de sua comunidade como para o público em geral. Este é um novo canal para estruturação da contribuição da universidade para o mundo mais amplo, como tal, convida a avaliação política e cultural dessa relação (Lynch, 2003, p. 332, tradução nossa).

O papel fundamental dos repositórios institucionais na recuperação de informação reside na facilitação do acesso e na disseminação eficiente do conhecimento produzido pela comunidade acadêmica. Ao concentrar e organizar as produções científicas em um único local, essas plataformas proporcionam uma maneira sistemática de buscar e recuperar informações relevantes. A indexação cuidadosa e a atribuição de metadados adequados permitem uma recuperação precisa e eficaz, contribuindo para a difusão ampla e acessível do conhecimento científico.

Os estudos que se utilizam da IA no âmbito da Ciência da Informação estão em ascensão, envolvendo aspecto como uso de chatbots e realidade virtual em serviços de biblioteca (Gujral; Shivarama; Choukimath, 2019), o uso de chatbots para perguntas frequentes e conversas guiadas a respeito dos serviços da biblioteca (Mckie; Narayan, 2019); apresentação de chatbots como o Rasa Stack, uma ferramenta em código aberto para uso em bibliotecas (Bagchi, 2020), dentre outros estudos que serão melhor abordados no capítulo 2.

Cumprir destacar a atualidade desses estudos, que remontam a 5 anos ou menos, mais uma vez reforçando sua temática eminente na área. Apesar dos estudos que serão destacados, no decorrer do levantamento bibliográfico não houve nenhuma pesquisa que trabalhasse com o uso de chatbots relacionados à representação de metadados em repositórios institucionais.

Sendo assim, a pesquisa é norteadada pela seguinte questão: considerando a PLN enquanto um mecanismo que permite que computadores compreendam a

linguagem natural utilizada por humanas a partir de programas como os *chatbots*, questiona-se de que maneira esses softwares podem atuar como sistemas auxiliares de recuperação da informação e contribuir para otimização desse processo em repositórios institucionais?

Com o propósito de atingir esse fim, elaboraram-se as seguintes perguntas específicas, que direcionam a análise aprofundada da literatura pertinente na área em questão: quais são os principais desafios que têm orientado as investigações no campo da Recuperação de Informação? Quais são as principais soluções apontadas por cada estudo para aprimorar a eficiência da Recuperação de Informação? Que aplicações cada estudo pode ter para aprimorar a Recuperação de Informação em repositórios institucionais?

Juntas, essas perguntas específicas constituem uma abordagem integral que facilita a compreensão e a avaliação abrangente da influência do PLN e *chatbots* na eficácia dos processos de recuperação de informação. A busca por informações de interesse ou que atendam às necessidades dos usuários representa a tarefa de pesquisa mais crucial na *Web*. Contudo, alcançar esse propósito nos repositórios institucionais tem sido desafiador, devido às dificuldades encontradas na indexação social e no engajamento direto dos usuários na descrição de objetos informacionais.

### **1.1 Objetivo geral**

Analisar, a partir da literatura nacional e internacional, como o PLN e os *chatbots* podem contribuir para a recuperação da informação em repositórios institucionais.

### **1.2 Objetivos específicos**

1 - Identificar e mapear os principais problemas que têm guiado as pesquisas sobre PLN e *chatbots* na área de recuperação de informação;

2 - Analisar as soluções propostas em cada estudo específico, destacando as principais abordagens sugeridas para aprimorar a recuperação de informação;

3 - Caracterizar as aplicações específicas de cada estudo no sentido de aprimorar a recuperação de informação em repositórios institucionais, destacando as

contribuições práticas e potenciais benefícios para o processo de busca e recuperação de informação.

### 1.3 Justificativas

Esta pesquisa se justifica pelo fato de que aborda questões altamente relevantes dentro do contexto da linha de pesquisa "Tecnologia, Informação e Representação", do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos (UFSCar). Além disso, a investigação contribuirá significativamente para a literatura da área, com estudos aprofundados sobre recuperação da informação.

A análise apresentada, que visou examinar, com base na literatura nacional e internacional, como o PLN e dos *chatbots* podem contribuir para a recuperação da informação em repositórios institucionais, encontra seu contexto no cenário dinâmico da transformação digital e inovação na gestão da informação. Com a crescente digitalização do conhecimento e a expansão da produção acadêmica, repositórios institucionais desempenham um papel crucial na disseminação e acesso a esse vasto acervo de dados e informações. Por exemplo, a Estratégia Brasileira para a Transformação Digital (E-Digital) para o ciclo 2022 – 2026 é dividida em dois grupos de eixos temáticos: eixos temáticos habilitadores e eixos de transformação digital. No grupo de transformação digital, o eixo economia baseada em dados, que tem como um dos objetivos específicos promover o desenvolvimento de soluções tecnológicas nas áreas prioritárias de desenvolvimento do país, está atrelado à necessidade de criação de infraestruturas para armazenamento, compartilhamento e reutilização de dados, promovendo o aumento recorrente de seu valor (MCTI, 2022).

O PLN e a introdução dos *chatbots*, nesse contexto, representam uma resposta às demandas por soluções mais eficientes e acessíveis para reaproveitamento e reutilização de dados e informações armazenados em repositórios institucionais. O PLN, ao capacitar sistemas computacionais a compreenderem e interpretarem a linguagem humana, promete aprimorar a busca de informações nos repositórios, tornando-a mais intuitiva e precisa. Por sua vez, os *chatbots*, com sua capacidade de interação natural, têm o potencial de oferecer uma experiência de usuário mais amigável e eficaz.

A análise abrangente, incorporando perspectivas da literatura nacional e internacional, amplia a compreensão sobre potenciais contribuições dessas tecnologias. Ela não apenas busca identificar os benefícios, mas também avaliar as implicações práticas, desafios e oportunidades que o PLN e os *chatbots* apresentam na otimização da recuperação da informação em repositórios institucionais. Nesse contexto, a pesquisa visa fornecer *insights* valiosos para aprimorar a eficiência e a acessibilidade dessas plataformas, alinhando-se com as demandas evolutivas do cenário acadêmico e informacional contemporâneo.

No Brasil, o último levantamento do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), publicado em março de 2022, identificou 151 repositórios de acesso aberto, que contém publicações científicas e dados abertos. Esta situação ilustra o cenário do desenvolvimento tecnológico descrito por Jelusic e Stricevic (2011), caracterizado pela complexidade decorrente da transição da publicação impressa para a publicação eletrônica. Os ambientes digitais revolucionam os paradigmas de atuação bibliotecária, suscitando a compreensão de novos processos de representação, de satisfação de necessidades de usuários e de apropriação da informação.

O avanço tecnológico inseriu as bibliotecas numa dimensão totalmente nova. Trouxe novas possibilidades, mas com elas muitos dilemas e problemas. Os catálogos informáticos, os sistemas de informação das bibliotecas, a publicação eletrônica, a digitalização de coleções e muitas outras possibilidades abriram-se às bibliotecas com o aparecimento das TICs (Jelusic; Stricevic, p. 26, tradução nossa).

Alguns dilemas e problemas citados pelos autores envolvem a recuperação da informação. segundo os autores Baeza-Yates e Ribeiro-Neto (2013), entender os desafios enfrentados na recuperação de informações permite desenvolver soluções eficazes que melhoram a acessibilidade, a usabilidade e as utilidades desses repositórios. Para Ingwersen (1992), a área da recuperação da informação concentra-se nos procedimentos relacionados à representação, armazenamento, pesquisa e obtenção de informações pertinentes à questão que um usuário humano procura resolver ou deseja obter. Conforme o autor, o desafio reside em localizar informações, seja em forma de texto seja em outras mídias, que atendam de maneira ideal ao estado de incerteza e ao âmbito do problema do usuário. Portanto, certos documentos possuem uma relevância superior para requisitos de informação específicos quando comparados a outros documentos, e um texto particular pode adquirir significados distintos conforme as várias necessidades de informação.

A indexação automática viabiliza o acesso e a recuperação da informação no cenário informacional atual, caracterizado pelo volume excessivo da informação. Geralmente, o processo se baseia na construção de um índice invertido cuja estrutura é composta pelo vocabulário e pelas ocorrências. O vocabulário é o conjunto das palavras do documento e, para cada palavra do vocabulário, o índice armazena os documentos correspondentes (Baeza-Yates; Ribeiro-Neto, 2013). Desse modo, a tarefa de representação pode ser feita com maior rapidez e as consultas do usuário processadas em linguagem natural. Contudo, na prática, muitos problemas podem ocorrer.

Ao analisar a relevância, conceito chave dos sistemas de recuperação de informação, Mizzaro (1998) mostra-nos que existem muitos tipos de relevância, que podem ser classificados em um espaço quadrimensional. Na dimensão representação da necessidade do usuário, por exemplo, é preciso considerar as diferenças entre necessidade de informação real, que condiciona a construção da necessidade de informação percebida. A necessidade percebida é uma representação da necessidade real. O problema é que, às vezes, o usuário pode não perceber corretamente a sua necessidade real. Além disso, para o sucesso da tarefa de busca, o usuário deve ser capaz de traduzir a sua necessidade percebida no formulário, utilizando a expressão correta. Esta situação torna problemática a recuperação da informação, principalmente, nos repositórios institucionais que utilizam a indexação social.

A outra abordagem para lidar com o conceito de "assunto/conteúdo", conforme Ingwersen (1992), é focar no processo de representação do indexador humano. Nesse cenário, a questão de o que constitui um documento pode ser respondida através do trabalho de um indexador que, por meio da classificação ou indexação de documentos, busca resumir ou substituir o conteúdo da mensagem em cada documento ou trecho de texto. Este processo é baseado na abordagem cognitiva, formada pelo contexto sócio-histórico-cultural em que o indexador está inserido. Esse contexto pode ser problemático no contexto da interoperabilidade de informação entre sistemas na *Web*.

Segundo Lima, Fujita e Redigolo (2021), a recuperação da informação vai além da simples indexação e busca por palavras-chave. Ela envolve a compreensão do conteúdo, do contexto e das necessidades dos usuários para fornecer resultados precisos e relevantes. A representação adequada dos itens de informação, como a criação de metadados descritivos e a categorização por meio de etiquetas, ajuda a

classificar e organizar o conteúdo de maneira que seja mais fácil de ser encontrado e compreendido, sendo o PLN e a folksonomia opções para forma facilitar o uso de termos de busca pelo pesquisador.

Os autores Santarém Segundo (2010), Santos e Corrêa (2015), Vajjala et.al (2020), Santos, Carvalho e Rodrigues (2021), discorrem que a folksonomia pode ser compreendida enquanto uma prática de representação que permite que os próprios usuários classifiquem e “etiquetem” os recursos de acordo com sua compreensão e com o seu contexto pessoal. Isso proporciona uma flexibilidade e adaptabilidade que os sistemas de classificação tradicionais muitas vezes não conseguem oferecer. Os termos usados nas *tags* podem ser coloquiais, específicos para um nicho ou culturalmente relevantes, o que torna a classificação mais próxima do uso real da linguagem.

Contudo, para que o repositório institucional tenha um padrão é necessário que se tenha uma política de indexação clara e bem explicada, para que o autor, ao fazer o autoarquivamento, possa saber qual o modelo de metadado assunto vai ser inserido no campo designado.

A adoção de novas interfaces, como os *chatbots*, oferece aos usuários a capacidade de acessar informações na internet por meio de aplicativos de mensagens já familiares, dispensando a necessidade de configurações especiais e possibilitando a utilização de linguagem natural.

Os *chatbots* podem revelar um papel significativo na recuperação de informações, particularmente no contexto da representação e indexação. Por exemplo, *chatbots* podem interagir com os usuários para compreender suas necessidades informativas e, em seguida, utilizar algoritmos de PLN e técnicas de recuperação de informação para identificar documentos relevantes e fornecer respostas pertinentes. Além disso, *chatbots* podem ser treinados para entender diferentes domínios e estilos de linguagem.

Os avanços recentes nas técnicas de IA e PLN tornaram a implementação de *chatbots* mais acessível e flexível em relação às suas aplicações e manutenção, além de incrementar sua habilidade em simular conversações humanas.

Para fundamentar ainda mais a justificativa de uma pesquisa nessa área, é possível examinar a presença de outros estudos já realizados, conforme pode ser visto no próximo capítulo.

## 1.4 Estrutura do trabalho

A pesquisa está estruturada em cinco capítulos, com o primeiro dedicado à introdução. Neste capítulo inicial, é fornecida uma explicação inicial dos conceitos abordados no texto, juntamente com a apresentação da problemática a ser investigada, dos objetivos a serem alcançados com a pesquisa e da justificativa que embasa a relevância do estudo. Essa seção introdutória estabelece o contexto para o restante do trabalho, delineando as questões a serem exploradas e os propósitos da investigação.

No segundo capítulo, foram realizadas revisões de estudos anteriores que abordam temas cruciais para a pesquisa, examinando trabalhos acadêmicos, pesquisas científicas e projetos que se concentram na intersecção entre CI e Inteligência Artificial (IA), além de estudos relacionados a *chatbots*, PLN, recuperação de informação e outras áreas relevantes. Essa revisão abrangente da literatura proporcionou uma base sólida para o desenvolvimento e a contextualização do presente estudo, fornecendo *insights* valiosos sobre o estado atual da pesquisa nesse campo multidisciplinar.

No terceiro capítulo foi feita a fundamentação teórica, que visa proporcionar uma compreensão aprofundada dos conceitos-chave explorados na pesquisa. Nesta seção, foram discutidos diversos aspectos pertinentes, tais como a interação entre CI e IA, o funcionamento dos *chatbots* e do PLN, os princípios subjacentes à recuperação de informação, o papel desempenhado pelos repositórios institucionais na gestão da informação, a importância da indexação e a concepção da folksonomia. A análise detalhada desses conceitos fornece um arcabouço teórico sólido para embasar a investigação em curso, permitindo uma compreensão mais profunda e contextualizada dos temas abordados.

No quarto capítulo, foi abordado o método e a metodologia empregados na pesquisa, compreendendo uma detalhada descrição das etapas, procedimentos e técnicas adotadas com o propósito de atingir os objetivos estabelecidos. Também foi apresentada uma análise minuciosa do processo de pesquisa, incluindo a seleção e aplicação de ferramentas metodológicas específicas, bem como a justificativa por trás de suas escolhas. A explanação cuidadosa desses aspectos proporcionará uma compreensão clara e abrangente do rigor metodológico empregado no estudo.

No quinto capítulo, foi feita a análise e discussão dos resultados, descrevendo detalhadamente os resultados obtidos por meio da pesquisa, seguida de uma análise aprofundada desses resultados. Além disso, foi promovida uma discussão sobre as descobertas, conclusões e possíveis implicações para o campo da CI e áreas correlatas. A partir dessa análise crítica e reflexiva, busca-se fornecer *insights* valiosos, bem como contribuições significativas para o avanço do conhecimento e prática no campo da pesquisa.

Ao término, o trabalho apresenta as considerações finais, seguidas pela lista de referências bibliográficas e apêndice. Nesta seção, foram destacados os principais pontos discutidos ao longo da pesquisa, bem como possíveis sugestões para pesquisas futuras. As referências bibliográficas garantirão a credibilidade e a integridade do trabalho, fornecendo uma base sólida para os argumentos apresentados.

## 2 ESTUDOS RELACIONADOS

O interesse pelos estudos sobre *chatbots* e recuperação de informação na Ciência da Informação (CI) está em ascensão, refletindo o desejo de explorar o papel dessas tecnologias na facilitação do acesso às informações. Na interseção entre esses dois temas, observa-se um campo dinâmico, no qual os pesquisadores se dedicam a aplicá-los de maneira eficaz para atender às necessidades informacionais dos usuários. Esse interesse evidencia o reconhecimento do potencial transformador dos *chatbots* na recuperação de informação, resultando em melhorias significativas na acessibilidade ao conhecimento. Os estudos apresentados foram organizados de forma cronológica e abrangem pesquisas globais.

No artigo de Gujral, Shivarama e Choukimath (2019), é realizado um mapeamento da aplicação e do potencial impacto da IA em bibliotecas acadêmicas. A implementação da IA em quatro domínios - educacional, informativo, assistivo e redes sociais - é minuciosamente detalhada, juntamente com suas implicações. Dentre os temas explorados, destacam-se agentes de conversação, como *chatbots*, serviços de biblioteca participativa baseados em IA e o impacto na pesquisa e descoberta de informações, além da utilização da realidade virtual. A incorporação da IA em serviços de referência virtual representa um novo modelo de atendimento online para bibliotecas, mostrando como os bibliotecários estão sempre na vanguarda das tecnologias, buscando envolver e aprimorar os serviços para seus usuários. No entanto, é importante ressaltar que adições significativas compreendem experiências de realidade virtual que estabelecem uma conexão entre os usuários e as bibliotecas, além de fortalecerem as habilidades de alfabetização informacional.

No estudo de Mckie e Narayan (2019), é examinado um protótipo desenvolvido pela Universidade de Tecnologia de Sydney (UTS), que se concentra na exploração de *chatbots*, incluindo perguntas frequentes (FAQ) e conversa guiada. Os autores enfatizam a importância do planejamento e contextualização na construção de *chatbots* para bibliotecas, resultando na personalização das respostas de acordo com o papel na jornada do usuário e na simulação de uma conversa humana. Além disso, o estudo aborda as barreiras enfrentadas pelos estudantes na pesquisa acadêmica e destaca como os *chatbots* podem superá-las, facilitando o acesso a informações relevantes. Essa pesquisa evidencia o potencial dos *chatbots* como ferramentas eficazes para melhorar a experiência dos usuários nas bibliotecas, oferecendo

soluções personalizadas e acessíveis para as necessidades informacionais dos estudantes.

Gundakanal e Kaddipujar (2019) abordam a evolução e aplicação da IA, com ênfase em *machine learning*, nas bibliotecas. O estudo destaca os desafios enfrentados na implementação da IA nesse ambiente, oferecendo *insights* valiosos sobre as competências necessárias para os bibliotecários na sociedade artificialmente inteligente. O crescimento tecnológico tem impactado diversos setores, proporcionando tanto habilidades e oportunidades quanto desafios. A influência da IA é observada em atividades cotidianas, desde a navegação na *Web* até a interação com assistentes virtuais como Siri e *Google Assistant*. Essas transformações têm um impacto significativo nas bibliotecas, remodelando a coleta, o acesso e a distribuição de informações. Tecnologias emergentes, como a internet das coisas, redefinem a capacidade dos profissionais de processar e analisar dados. O estudo explora como tecnologias avançadas, especialmente a IA, podem impulsionar as bibliotecas modernas na era digital, proporcionando novas oportunidades para melhorar os serviços e atender às demandas dos usuários de maneira mais eficaz.

No artigo de Anelli et al. (2019), foi apresentado um sistema desenvolvido com o objetivo de auxiliar os usuários na exploração do conhecimento disponibilizado pela Biblioteca Digital da Puglia, que engloba informações sobre bens digitais na região da Puglia, na Itália. O estudo demonstra a interação com a biblioteca digital por meio de um assistente virtual, ressaltando como, ao ser publicada como *Linked Open Data*, ela pode ser integrada de forma ágil e dinâmica com fontes externas de conhecimento, como aquelas de natureza geográfica. Essa abordagem promove uma experiência mais enriquecedora para os usuários, permitindo que explorem e ampliem seu conhecimento de maneira eficaz e integrada.

Em Bagchi (2020), foi proposto a conceituação de um novo *chatbot* para bibliotecas, utilizando uma plataforma de código aberto recentemente desenvolvida com IA, chamada Rasa, e sugere sua possível adoção por instituições bibliotecárias. Além de apresentar a essência da tecnologia *chatbot* e sua aplicação atual em bibliotecas, o artigo explora os fundamentos técnicos do Rasa Stack que podem ser aproveitados no desenvolvimento de *chatbots* para bibliotecas. Além disso, reflete sobre o potencial de pesquisas futuras nessa direção, destacando a importância de investigações adicionais para aprimorar e ampliar a aplicação dessa tecnologia no contexto bibliotecário. Essa abordagem representa um avanço significativo no campo

da automação e interação digital nas bibliotecas, promovendo uma experiência mais eficiente e personalizada para os usuários.

No estudo de Erickson e Kim (2020), são discutidos os Sistemas de Gerenciamento de Conhecimento (SGC) e os *chatbots* como elementos-chave na entrega eficiente de dados para diversos usuários. Os *front-ends* dos *chatbots*, utilizados por funcionários remotos, representantes de *call center* e clientes de bibliotecas, são conectados ao *back-end* SGC para fornecer informações em tempo real. Uma evolução notável dos *chatbots* inclui a capacidade de fornecer respostas aprendidas a conjuntos de pesquisas e sequências de ações na tela do computador. A revisão de literatura abrange estudos em bibliotecas, ambientes educacionais e locais de trabalho, destacando a importância da integração entre *chatbots* e SGC para uma entrega abrangente e eficaz de informações. Recomenda-se fortemente considerar essa integração em futuras pesquisas, pois pode aprimorar significativamente a experiência do usuário e a eficiência na obtenção de informações.

O estudo conduzido por Ali, Naeem e Bhatti (2020) visa avaliar o potencial de aplicação de ferramentas de IA em bibliotecas acadêmicas no Paquistão, com foco nos setores de serviços técnicos e usuários de bibliotecas. Utilizando uma abordagem auto exploratória, a pesquisa incluiu entrevistas qualitativas com 10 bibliotecários chefes de bibliotecas de universidades, representando tanto o setor público quanto o privado. Os resultados, apresentados de forma descritiva, revelam que os bibliotecários estão conscientes das tecnologias de IA incluindo serviços baseados em PLN, como *Google Assistant*, *Voice Search* e *Google Translate*. Métodos de reconhecimento de padrões, como mineração de dados de texto, são empregados para recuperar materiais e realizar pesquisas online. O estudo oferece sugestões aos bibliotecários para a incorporação de ferramentas de IA, especialmente para aqueles que ainda não adotaram essas tecnologias ou buscam soluções mais avançadas. Isso destaca a importância de se adaptar às inovações tecnológicas para melhor atender às necessidades em constante evolução dos usuários das bibliotecas acadêmicas.

A pesquisa conduzida por Silva (2020) enfatiza o papel significativo dos *chatbots* na recuperação de informação dentro da área da CI. O estudo explora a influência da CI no desenvolvimento desses *chatbots*, os quais são programas de *software* projetados para interagir com usuários de maneira semelhante a uma conversa humana. A análise dos *chatbots* em relação à recuperação de informação na CI ressalta não apenas sua contribuição para esse campo, mas também como a

CI molda e aprimora continuamente sua funcionalidade. Ao examinar a interação entre os *chatbots* e a recuperação de informação, o estudo oferece uma compreensão mais profunda das implicações práticas e teóricas dessa tecnologia emergente no contexto da CI. A metodologia adotada foi exploratória, buscando familiaridade com o problema e abrindo caminhos para futuras discussões. Os resultados evidenciam a importância dos *chatbots* na facilitação do acesso ao conhecimento por meio da interação em linguagem natural, ressaltando a necessidade de uma organização adequada das informações para o correto funcionamento desses agentes conversacionais. Conclui-se que os *chatbots* representam uma ferramenta promissora para a recuperação da informação, com a CI desempenhando um papel crucial em seu desenvolvimento e aprimoramento.

O estudo realizado por Batista (2021) investigou o uso de *chatbot* para suporte técnico de primeiro nível, visando facilitar o acesso à informação e agilizar o atendimento ao usuário. O trabalho explora cinco temas na fundamentação teórica: *chatbots*, linguagem natural, recuperação de informação, modelos de recuperação da informação e IA. A metodologia detalha a construção do *chatbot*, incluindo etapas do processo, levantamento histórico, tratamento de dados, criação da base de conhecimento e o serviço em si. A pesquisa utilizou a metodologia de revisão sistemática de literatura, com seis questões de pesquisa definidas para orientar a investigação. Os resultados mostraram que o *chatbot* desenvolvido obteve boa avaliação dos usuários em relação ao tempo de resposta e à coerência das informações. A satisfação dos usuários com o *chatbot* foi positiva, porém houve uma preferência equilibrada em relação ao atendimento humano, devido à questão emocional. Apesar dos resultados satisfatórios, ainda são necessários ajustes para lidar com contextos mais complexos. A pesquisa conclui que, embora os objetivos tenham sido alcançados, ainda há espaço para melhorias na tecnologia de *chatbot* para substituir plenamente os atendentes humanos no suporte técnico.

O estudo de Carvalho (2022) destaca a crescente importância das pesquisas em métodos de recuperação de informação, especialmente diante da expansão da *internet* e do avanço tecnológico. No contexto atual, a utilização de novas interfaces, como *chatbots*, emerge como uma abordagem promissora para o acesso intuitivo a informações na *Web*. Este estudo propõe um modelo conceitual para avaliar a eficácia dos *chatbots* como SRI, visando resultados de qualidade e minimização de problemas de comunicação. O modelo inclui o enriquecimento dos termos do usuário com

metadados do repositório, com o intuito de mitigar ruídos terminológicos. A pesquisa adota uma abordagem exploratória, compreendendo uma revisão bibliográfica para fundamentação teórica e a elaboração do modelo proposto. Os resultados indicam que a inclusão de metadados melhora a qualidade dos resultados e enfrenta a subjetividade nos diálogos. Conclui-se que os sistemas de recuperação de informação são mais eficientes ao incorporar metadados em interfaces de *chatbots*, promovendo inclusão, obtenção de melhores resultados, redução do número de interações e diminuição da dependência de tecnologias complexas.

A pesquisa de Lima (2022) tem como objetivo analisar o impacto dos agentes autônomos computacionais, ou *bots*, na execução de processos informacionais e na interação com seres humanos no contexto da *internet*, sob a perspectiva da CI, utilizando uma abordagem bibliográfica e descritiva, o estudo envolveu um mapeamento sistemático da literatura, análise documental e observação direta sistemática de *bots* em ação. Foram realizados estudos de casos múltiplos com *bots* reais para examinar processos informacionais, benefícios, limitações e sua influência em temáticas da CI. Os resultados destacam que os *bots* desempenham diversos processos informacionais, com impactos tanto benéficos quanto maléficos. Observou-se que, apesar de suas capacidades, eles enfrentam limitações em atividades que exigem cognição, empatia e criatividade, habilidades consideradas exclusivas dos seres humanos. Em alguns casos, problemas éticos, morais e legais foram identificados. Conclui-se que as temáticas da CI são relevantes para pesquisas envolvendo *bots*, e os profissionais da informação devem considerar a IA e seus produtos como entidades que interagem com informações na rede.

A autora Guimarães (2022) destaca em sua pesquisa a criação de um *chatbot* para recuperar informações na base de dados do Catálogo de Teses e Dissertações da CAPES. Embasada em metodologias de estudos de usuários e referenciais teóricos de práticas informacionais, explora elementos sociais que moldam a subjetividade e influenciam a realidade. Adotando a abordagem *Design Science Research* (DSR), o estudo desenvolveu iterativamente o *chatbot*, chamado Kika, em três encontros virtuais com participantes de graduação, pós-graduação e pesquisadores. O artefato recuperou dados do Catálogo da CAPES (2013 a 2020) e a avaliação focou na experiência do usuário, considerando confiabilidade e facilidade de interação no pseudodiálogo para acesso e recuperação de informações. A conclusão destaca que a construção do *chatbot*, alinhada às práticas informacionais,

promove a interação do sujeito informacional com o artefato, considerando sua evolução durante o *design*.

O estudo da autora Monterei (2022) destaca as aplicações da IA, com ênfase em *machine learning*, nas bibliotecas. O objetivo principal é mapear os benefícios e impactos que o aprendizado de máquina pode oferecer para o desenvolvimento de produtos e serviços nesse contexto. A pesquisa adota uma abordagem qualitativa e quantitativa, com foco exploratório, de natureza pura, utilizando pesquisa bibliográfica. Conclui-se que este estudo proporciona ao pesquisador a oportunidade de refletir e identificar novos fenômenos nas relações interdisciplinares entre a CI e a IA.

No artigo de Tella e Ajani (2022), a aplicação da IA em sistemas bibliotecários é abordada, com referência a estudos de taxonomia e análise do impacto da IA, aprendizado de máquina, automação e robótica nas profissões da informação. São explorados exemplos de serviços robóticos relevantes para bibliotecas, como inventário automatizado, entrega de livros aos usuários e interação com crianças e adultos. É ressaltada a importância da comunicação aberta entre os conselhos de gerenciamento da biblioteca e os bibliotecários para lidar eficazmente com questões relacionadas aos robôs. Há ênfase na necessidade de desenvolver um plano estratégico para enfrentar tecnologias emergentes, como os robôs humanóides, e destaca-se a importância de estabelecer parcerias com instituições de origem e fornecedores de robôs. Essa abordagem sublinha a importância da adaptação das bibliotecas às inovações tecnológicas emergentes para melhor atender às necessidades dos usuários e otimizar os serviços oferecidos.

No estudo de Idemudia e Makinde (2022), foi direcionado à implementação da IA em bibliotecas, analisando as perspectivas e desafios específicos enfrentados pelas bibliotecas acadêmicas na Nigéria. A pesquisa concentrou-se na identificação das áreas da biblioteca que poderiam se beneficiar da aplicação da IA destacando diferentes componentes da IA, como sistemas especialistas, reconhecimento de padrões, PLN e robótica, todos com potencial para serem incorporados nas bibliotecas acadêmicas do país. Essas aplicações incluem atividades como aquisição de materiais bibliográficos, catalogação, classificação, indexação, entre outras. A revisão da literatura enfatizou a robótica como um exemplo de superinteligência artificial, capaz de revolucionar os serviços oferecidos pelas bibliotecas acadêmicas. Conclui-se que, apesar dos desafios inerentes, as bibliotecas acadêmicas nigerianas têm a

capacidade de adotar a IA em suas operações, garantindo eficiência, eficácia e qualidade nos serviços bibliotecários.

No estudo de Tsimara, Oikonomou e Katsirikou (2023), na era contemporânea, a comunicação direta e o amplo acesso à Internet delineiam um cenário onde os dados estão abundantemente disponíveis. No entanto, a disseminação fragmentada de informações, especialmente no âmbito científico, torna desafiador correlacionar conceitos, prejudicando a busca pelo conhecimento. Apesar da eficácia dos mecanismos de busca na recuperação de dados, suas limitações em classificar resultados pela relevância para a pergunta são evidentes, principalmente em consultas que abrangem mais de um conceito. Para enfrentar essa lacuna, surgiu o desenvolvimento da *Web Semântica*, buscando emular o funcionamento do cérebro humano. Os defensores da *Web Semântica* destacam suas aplicações bem-sucedidas em bibliotecas, pesquisa em CI, indústria, biologia, medicina e humanidades, proporcionando uma experiência de usuário mais interativa. O objetivo do artigo é descrever o uso das tecnologias de *Web Semântica* e IA em bibliotecas, ressaltando sua importância para aprimorar o acesso à crescente necessidade de recuperação de texto completo. O exemplo da Biblioteca da Universidade de Piraeus ilustra de forma eficaz como essas tecnologias são empregadas para recuperar informações de fontes disponíveis para seus usuários.

Após examinar a literatura disponível, identificou-se uma lacuna de pesquisa que ressalta a necessidade de investigar o uso de *chatbots* para facilitar a recuperação de informações em repositórios institucionais. Embora existam estudos que tratam do emprego de *chatbots* na recuperação de informações, até o momento nenhum deles se dedicou especificamente a explorar essa aplicação dentro do contexto específico dos repositórios institucionais.

De modo a melhor abordar essa lacuna, para que então se cumpra com os objetivos desta pesquisa, no próximo capítulo abordaremos as fundamentações teóricas adotadas para a investigação, em especial o que diz respeito à inteligência artificial no contexto da Ciência da Informação, *chatbots* e Processamento de Linguagem Natural.

## 3 FUNDAMENTAÇÃO TEÓRICA

### 3.1 Ciência da Informação e Inteligência Artificial

A CI é um campo multidisciplinar que aborda o estudo da informação sob diferentes perspectivas. O avanço tecnológico tem provocado transformações rápidas e significativas na forma como acessamos, armazenamos e consumimos dados. A IA, como campo de pesquisa, oferece soluções abrangentes em diversas áreas, conectando os fundamentos teóricos da Organização do Conhecimento na CI aos mecanismos que facilitam a vida do usuário e atuam como suporte estratégico.

Segundo Coneglian (2020), na CI a IA desempenha um papel crucial ao facilitar diversos processos, incluindo a representação, organização e recuperação da informação. Isso não apenas permite a criação de novos modelos que incorporam essas ferramentas e conceitos inovadores, mas também tem o potencial de impactar de maneira significativa as pesquisas e soluções desenvolvidas nesse campo. A crescente popularização dos serviços de IA não só influencia as práticas tradicionais, mas também introduz paradigmas que podem transformar a concepção de processos convencionais.

Pereira (2019) conceitua a IA como uma ampla área de pesquisa subdividida em diversas subáreas, cada uma adotando abordagens distintas e tratando de problemas complexos para os quais ainda não existem soluções completamente satisfatórias.

Já Cozman e Neri (2021, p. 21) caracterizam a IA como a área dedicada à criação de artefatos artificiais que manifestam comportamento inteligente. Contudo, a definição de "comportamento inteligente" é desafiadora, uma vez que o conceito de "inteligência" é fluido e flexível para os seres humanos. A aceitação da inteligência limitada em certos animais e a rápida adaptação aos artefatos digitais com evidentes restrições cognitivas destacam a complexidade associada à compreensão do que constitui inteligência no contexto da IA.

Nos últimos anos, a IA tem sido objeto de estudo em diversas áreas do conhecimento. Originado na área de CC, esse campo de pesquisa passou por significativa evolução, impulsionado pelo avanço tecnológico nas últimas décadas. Esse progresso resultou na criação de diversas tecnologias que se incorporaram ao cotidiano, sendo fundamentadas nos princípios da IA.

Uma evidência clara do avanço tecnológico atual pode ser exemplificada por soluções inovadoras, como os *chatbots*. Essa categoria de IA, conforme discutida por Coneglian (2020), busca simular a interação humana por meio de conversas. A abordagem dos *chatbots* reflete não apenas a sofisticação tecnológica, mas também a busca por uma experiência de usuário mais intuitiva e próxima à comunicação humana. A análise de Coneglian (2020) destaca a relevância dessas soluções no cenário tecnológico contemporâneo, evidenciando como a IA está sendo aplicada para replicar aspectos complexos da interação humana.

Com isso, a IA proporciona avanços notáveis na automatização de processos informacionais, contribuindo para a eficiência em bibliotecas, arquivos e sistemas de informação. Sistemas de recomendação, baseados em IA, personalizam experiências de usuários, facilitando o acesso ao conteúdo relevante. Além disso, técnicas de PLN e aprendizado de máquina são aplicadas na análise de grandes conjuntos de dados, agilizando a categorização e extração de informações valiosas.

A interligação entre CI e IA ressalta a importância crucial da IA na transformação digital das práticas informacionais. A colaboração entre profissionais de CI e especialistas em IA é vital para a criação de soluções inovadoras, permitindo que as instituições enfrentem a crescente complexidade da informação na era digital. Nesse cenário, a compreensão das interfaces entre CI e IA é fundamental para explorar totalmente o potencial dessas disciplinas e impulsionar avanços significativos na gestão e acesso à informação.

No estudo apresentado por Saracevic (1996) na *International Conference on Conceptions of Library and Information Science*, a IA é destacada como uma das áreas-chave de interesse para ambas, CC e CI, sendo considerada uma linha de pesquisa pertencente às Ciências Cognitivas. De acordo com o autor, a relevância crucial dessa ciência cognitiva reside na interação de abordagens extremamente diversificadas na abordagem de questões relacionadas ao cérebro e à mente, abrangendo desde as humanidades até as ciências da vida, das ciências sociais às matemáticas, da lógica às engenharias. Embora todas essas áreas possuam potencial interesse para a CI, a proximidade mais iminente é estabelecida pelas questões investigadas e soluções experimentadas na IA.

De acordo com Coneglian (2020), a IA desempenha um papel significativo em sistemas de recuperação da informação, que são componentes fundamentais na subárea de sistemas de informações. Esses sistemas têm como principal objetivo

recuperar diversos tipos de dados, incluindo textos, imagens e sons, em resposta a consultas dos usuários. A ideia central é que, a partir de uma pergunta formulada pelo usuário, o sistema seja capaz de apresentar resultados relevantes à questão proposta.

Conforme Russell e Norvig (2016), a aplicação da IA em sistemas que permitem a classificação automática de conteúdo é fundamental para a recuperação eficiente de informações, demandando uma organização lógica e semanticamente coerente. Nesse contexto, as ontologias desempenham um papel crucial, podendo ser desenvolvidas tanto de forma manual quanto automática no ambiente digital. Essas estruturas facilitam a compreensão e a categorização de informações, contribuindo para a eficácia dos sistemas de recuperação da informação.

Outro campo de extensa pesquisa na IA é o Processamento de Linguagem Natural (PLN), que utiliza técnicas computacionais para que os computadores possam compreender instruções escritas em linguagem natural. Desta forma, melhora as interfaces de comunicação entre humanos e máquinas mesmo com a complexidade dos idiomas humanos, que envolvem ambiguidades e nuances dependentes de diversos fatores além da gramática. O PLN busca interpretar frases considerando contexto, ambiguidades, regras gramaticais, questões culturais e conceitos abstratos.

No PLN é possível permitir que as interfaces de sistemas automatizados de recuperação da informação compreendam as intenções do usuário e comuniquem-se de forma clara, facilitando a operação desses sistemas.

Dentre as diversas subdivisões da IA, os *chatbots* emergem como uma área específica de interesse nesta pesquisa. A próxima seção oferecerá uma análise mais aprofundada e detalhada sobre esse tema, explorando sua relevância, funcionamento e aplicações específicas no contexto da IA.

### **3.2 Chatbots e o Processamento de Linguagem Natural**

Mukherjee e Patra (2023) definem *chatbot* com um aplicativo de *software* desenvolvido com o propósito de emular interações humanas em diálogos com usuários finais. Embora a presença de IA não seja uma característica universal em todos os *chatbots*, a contemporaneidade dessas aplicações frequentemente se vale de métodos avançados de IA, notadamente o PLN, a fim de discernir e interpretar as indagações do usuário, bem como automatizar a formulação de respostas correspondentes.

Para os autores, os *chatbots* têm a capacidade de agilizar o processo de acesso às informações por parte dos usuários, uma vez que são aptos a oferecer respostas imediatas a indagações e demandas, seja por meio de entradas em formato de texto, entradas de áudio, ou por meio de uma combinação de ambas, dispensando a imprescindibilidade de intervenção humana ou a condução de pesquisas manuais (Mukherjee; Patra, 2023).

Para Crowston, Allen e Hechman (2012), o PLN é uma abordagem computacional para a análise de texto que envolve a compreensão e geração de linguagem humana de forma automatizada. Essa tecnologia permite que os computadores interpretem, analisem e gerem texto de maneira semelhante ao que os seres humanos fazem. O PLN envolve várias tarefas, como reconhecimento de entidades, análise sintática, análise semântica, tradução automática, sumarização de texto, entre outras.

Essas técnicas são fundamentadas em teorias linguísticas e algoritmos computacionais que permitem aos sistemas processar e compreender a linguagem escrita ou falada. O PLN é utilizado em uma ampla gama de aplicações, como assistentes virtuais, sistemas de recomendação, análise de sentimentos em redes sociais, extração de informações de grandes volumes de texto, entre outros. Para a presente pesquisa, vamos nos aprofundar nos conceitos e aplicações do PLN por meio de *chatbots*.

Segundo Martinez (2010) os *chatbots* e assistentes virtuais são aplicações de PLN que desempenham um papel crucial na interação entre humanos e computadores. Os *chatbots* são programas de computador que simulam conversas humanas por meio de mensagens de texto ou fala, sendo utilizados em diversos setores como atendimento ao cliente e comércio eletrônico. Eles funcionam por meio de algoritmos de PLN que analisam a entrada do usuário, compreendem a intenção por trás da mensagem e geram respostas apropriadas. Existem dois tipos principais de *chatbots*: os baseados em regras, que seguem um conjunto predefinido de regras, e os baseados em aprendizado de máquina, mais avançados e capazes de melhorar sua interação com os usuários ao longo do tempo.

Por outro lado, o autor supracitado explica que os assistentes virtuais são programas de IA que oferecem suporte e assistência aos usuários em diversas tarefas. Eles podem compreender comandos de voz e texto, realizar ações específicas e fornecer informações úteis. As funcionalidades dos assistentes virtuais incluem

agendamento de compromissos, navegação na web, controle de dispositivos inteligentes e assistência em tarefas cotidianas como previsão do tempo e recomendações de restaurantes. Essas aplicações estão em constante evolução, enfrentando desafios como compreensão de contexto e personalização da interação.

Em relação aos *chatbots*, Zeroual e Lakhouaja (2018) explicam que o propósito PLN reside em habilitar esses sistemas de conversação a compreender e gerar respostas em linguagem natural de maneira eficaz, capacitando os *chatbots* a analisar as mensagens dos usuários, discernir a intenção subjacente às perguntas e fornecer respostas pertinentes e coesas. Adicionalmente, o PLN desempenha um papel fundamental no aprimoramento da interação entre os *chatbots* e os usuários, conferindo às conversas um caráter mais natural e significativo.

Com avanços em tecnologias de PLN, aprendizado de máquina e processamento de linguagem profunda, os *chatbots* e assistentes virtuais prometem uma interação mais natural e personalizada entre humanos e máquinas, transformando a maneira como interagimos com a tecnologia e facilitando a realização de tarefas de forma eficiente e conveniente.

As versões mais recentes dos *chatbots* podem ser denominadas "assistentes virtuais inteligentes" ou "agentes virtuais" e demonstram uma capacidade não apenas de compreender diálogos fluentes por meio da utilização de modelos de linguagem sofisticados, mas também de automatizar tarefas de relevância substancial. Além de sua proeminência como assistentes virtuais de consumo, exemplificados por aplicativos como o "Apple Siri", o "Google Assistant", o "Amazon Alexa", esses agentes virtuais estão ganhando cada vez mais destaque em ambientes empresariais, onde desempenham um papel fundamental no auxílio a clientes e colaboradores (Mukherje; Patra 2023).

O início da comunicação computacional remonta a década de 50/60, a partir dos trabalhos de Alan Turing e Joseph Weizenbaum. Mais tarde, em 1966, Weizenbaum desenvolveu o Eliza, o pioneiro no campo dos *chatbots*. que fazia uso de algoritmos de correspondência de padrões. A ferramenta imitava a linguagem de um psicoterapeuta, usando ferramentas e regras rígidas (Vajjala et al., 2020).

Segundo os autores Adamopoulou e Moussiades (2020), com o avanço recente da tecnologia de IA, os *chatbots* estão experimentando um aumento significativo em sua popularidade. Atualmente, alguns *chatbots* empregam algoritmos altamente sofisticados para oferecer respostas mais precisas e específicas. É fundamental notar

que as capacidades de aprendizado profundo incorporadas aos *chatbots* de IA possibilitam a melhoria gradual da precisão das conversas ao longo do tempo.

Essas tecnologias de IA estão construindo redes de respostas apropriadas por meio de suas interações com os indivíduos. Consequentemente, as respostas fornecidas por um *chatbot* de IA se tornam mais robustas à medida que ele é utilizado por mais tempo. Em comparação com um *chatbot* que simplesmente combina informações com base em algoritmos, um *chatbot* de IA que utiliza aprendizado profundo tem a capacidade de oferecer respostas mais abrangentes e precisas, especialmente no que se refere às intenções subjacentes à consulta.

Para Adamopouloy e Moussiades (2020), anteriormente, os *chatbots* eram primariamente sistemas de entrada e saída de texto, limitados a responder a um conjunto restrito de perguntas diretas com respostas pré-programadas. Frequentemente, quando confrontados com tópicos complexos ou situações não previstas pelos desenvolvedores, os *chatbots* encontravam dificuldades, atuando de forma comparável a uma espécie de FAQ interativas, onde se destacavam ao fornecer respostas para as quais haviam sido treinados.

Ao longo dos anos, houve a introdução de mais regras e técnicas de PNL nos *chatbots*, permitindo que os usuários finais interagissem com eles em um estilo mais conversacional. De fato, os *chatbots* modernos têm a capacidade de aprender à medida que são expostos a uma variedade crescente de expressões humanas e, consequentemente, podem se adaptar ao contexto. A compreensão da Linguagem Natural é uma técnica amplamente adotada por *chatbots* de IA contemporâneos para identificar as necessidades dos usuários (Adamopouloy; Moussiades 2020)

Essas tecnologias se apoiam em técnicas de aprendizado de máquina e aprendizado profundo, dois aspectos da IA com distinções sutis entre si. Tais métodos são utilizados para construir uma base de conhecimento progressivamente mais abrangente, que é aprimorada por meio das interações do usuário. Esse aprimoramento amplia a capacidade dos *chatbots* de prever com precisão as necessidades dos usuários e fornecer respostas cada vez mais eficazes ao longo do tempo (Adamopouloy; Moussiades 2020)

Vajjala et al. (2020) também dissertam sobre as aplicações e conceitos dos *chatbots*. Para os autores, *chatbots* podem ser compreendidos como sistemas interativos que permitem que usuários interajam com eles em linguagem natural, a partir de textos ou ferramentas de discurso. O desenvolvimento desses sistemas

trouxe, segundo os autores, o início de uma nova era: a da interface conversacional, que em breve não vai requerer o uso de mouse ou telas para interação.

Como aplicações dos *chatbots*, os autores destacam as áreas de vendas e e-commerce; notícias e ferramentas de descobrimento de conteúdo; serviços de atendimento ao cliente, medicina e área jurídica. Mais especificamente, os autores dissertam sobre o uso muito comum em áreas de “Perguntas e Respostas”, comuns em *websites*. Nesses sistemas, a partir de uma pergunta do usuário, a ferramenta busca pela correta resposta como retorno ao usuário (Vajjala et al., 2020).

O PLN engloba uma variedade de operações que permitem aos sistemas computacionais lidar com texto de forma inteligente e eficaz. Entre essas operações resultantes do PLN, destacam-se segundo os autores Martinez (2010) e Crowston e Allen (1995):

**1. Tradução Automática:** é uma das aplicações mais conhecidas do PLN, permitindo a conversão de texto de um idioma para outro de forma automatizada. Sistemas avançados de tradução automática utilizam técnicas de PLN, como modelos de linguagem e alinhamento de texto, para produzir traduções precisas e fluentes.

**2. Desambiguação de Sentido da Palavra:** é essencial para lidar com a polissemia, ou seja, palavras que possuem múltiplos significados. O PLN emprega técnicas de análise contextual para determinar o significado correto de uma palavra com base no contexto em que é utilizada.

**3. Resumo:** envolve a geração de um resumo conciso e informativo a partir de um texto mais extenso. Algoritmos de PLN podem identificar as informações mais relevantes e condensá-las em um resumo que capture os pontos-chave do texto original.

**4. Anotação Sintática:** consiste em identificar e marcar os elementos sintáticos de um texto, como sujeito, objeto, verbos, entre outros. Essa operação é fundamental para análises mais avançadas, como análise semântica e geração de estruturas gramaticais.

**5. Compreensão e Classificação Supervisionada de Documentos de Texto:** envolve a extração de informações relevantes e a classificação do conteúdo com base em categorias predefinidas. Algoritmos de classificação supervisionada, como redes neurais e algoritmos de aprendizado de máquina, são frequentemente utilizados para essa tarefa.

**6. Extração de Conteúdo:** consiste em identificar e extrair informações específicas de um texto, como datas, nomes, locais, eventos, entre outros. Essa operação é útil para indexação de documentos, geração de metadados e análises de conteúdo.

**7. Design de Ferramentas de Recomendação e Sistemas de Suporte a Decisões:** é empregado no design de ferramentas de recomendação personalizadas, que utilizam análises de texto para sugerir produtos, conteúdos ou ações com base nas preferências do usuário. Além disso, sistemas de suporte a decisões utilizam análises de texto para fornecer *insights* e informações relevantes para tomadas de decisão.

**8. Expansão de Consultas:** é uma técnica que visa melhorar a precisão e abrangência dos resultados de busca, adicionando sinônimos, termos relacionados ou conceitos semânticos às consultas de pesquisa. O PLN é utilizado para identificar termos relevantes e expandir as consultas de forma inteligente.

**9. Estruturas de Perguntas e Respostas:** é empregado na criação de sistemas de perguntas e respostas, que permitem aos usuários fazer perguntas em linguagem natural e receber respostas precisas e contextualizadas. Esses sistemas utilizam técnicas de processamento de linguagem para compreender as perguntas e encontrar as respostas adequadas nos documentos ou bases de conhecimento.

**10. Análise de Sentimentos:** é uma operação que visa identificar e classificar as emoções expressas em um texto, como positivas, negativas ou neutras. Algoritmos de PLN são utilizados para analisar o tom e a polaridade do texto, sendo aplicados em áreas como monitoramento de redes sociais, análise de feedback de clientes e detecção de sentimentos em avaliações online.

Essas operações exemplificam a diversidade de aplicações e funcionalidades que o PLN oferece, contribuindo para aprimorar a interação entre humanos e sistemas computacionais no processamento e compreensão de texto.

Desta forma os assistentes de voz estão se tornando cada vez mais sofisticados, graças aos avanços na tecnologia de PLN e algoritmos de aprendizado de máquina. Eles são capazes de aprender com as interações dos usuários, melhorar sua capacidade de comandos e fornecer respostas precisas e relevantes. Tornaram-se parte essencial da vida cotidiana de muitas pessoas, facilitando as tarefas diárias e fornecendo um nível de conveniência e interação que antes era inimaginável. Com

o contínuo avanço da IA, espera-se que esses assistentes de voz se tornem ainda mais poderosos e integrados em nossas vidas.

Tais avanços no desenvolvimento de *chatbots* mostram o potencial da IA para melhorar a qualidade das interações humanas e oferecer uma experiência de conversação mais enriquecedora e satisfatória. Na nossa análise, essa experiência pode ser aplicada aos processos de recuperação de informação, de modo a melhorar a precisão e relevância na satisfação das necessidades de informação do usuário.

Os *chatbots* e assistentes virtuais, como discutido por Martinez (2010), estão imersos em desafios e avanços que moldam dinamicamente a interação entre humanos e computadores. No que diz respeito aos desafios, a compreensão contextual emerge como uma barreira central, destacando a dificuldade de interpretar não apenas as palavras individuais, mas o contexto subjacente das mensagens dos usuários. A personalização da interação é outro ponto crítico, exigindo a adaptação de respostas com base nas preferências individuais e histórico de interações, um desafio em constante evolução.

A privacidade dos dados surge como uma preocupação contemporânea e desafio constante, lidar com informações sensíveis demanda medidas robustas para garantir segurança. Desta forma, a ambiguidade linguística, incluindo ironia e sarcasmo, complica a tarefa dos *chatbots* em interpretar o tom e o contexto, destacando a complexidade da linguagem natural. A integração multicanal é um desafio técnico e de usabilidade, como aponta o autor, busca por uma experiência consistente em diferentes canais, como texto, voz e redes sociais, destaca a demanda dos usuários por coesão na interação (Martinez, 2010).

Ainda de acordo com Martinez (2010), no que tange aos avanços, ressalta-se o papel crucial do aprendizado de máquina, incluindo algoritmos como redes neurais, que aprimoram significativamente a compreensão e precisão dos *chatbots*. O PLN avançado permite uma análise mais sofisticada da linguagem, incluindo a capacidade de analisar contexto e gerar respostas mais contextuais e relevantes. A integração de dados externos e APIs enriquece a capacidade dos *chatbots* de acessar informações em tempo real, fornecendo respostas mais precisas e atualizadas. As funcionalidades multifuncionais dos assistentes virtuais, vão além do agendamento e incluem uma variedade de tarefas, contribuindo para sua utilidade e versatilidade.

A experiência do usuário é aprimorada com interfaces mais amigáveis, respostas mais rápidas e precisas, e uma interação mais natural. Martinez (2010)

destaca que esses avanços refletem o cenário dinâmico e promissor dos *chatbots* e assistentes virtuais, continuando a se desenvolver para atender às crescentes demandas dos usuários por interações mais inteligentes e personalizadas.

A interação essencial entre PLN e *chatbots* é salientada por Crowston, Allen e Hechman (2012), que ressaltam a importância fundamental do PLN na eficácia operacional dos *chatbots*. No âmbito da Compreensão de Linguagem Natural, delineada pelos autores, os *chatbots* são capacitados a entender e interpretar a linguagem natural dos usuários, incluindo o reconhecimento de palavras-chave, a compreensão do contexto das mensagens e a identificação da intenção subjacente às perguntas dos usuários.

A geração de respostas, conforme abordado por esses autores, baseia-se na análise realizada pelo PLN, possibilitando que os *chatbots* elaborem respostas relevantes e coerentes para as perguntas dos usuários, indicando que o PLN auxilia os *chatbots* na seleção de informações apropriadas e na estruturação de respostas compreensíveis.

A Personalização da Interatividade, conforme discutido por Crowston, Allen e Hechman (2012), é facilitada pelo PLN, permitindo que os *chatbots* personalizem a interação com os usuários com base em informações contextuais e histórico de conversas. Isso resulta em uma experiência mais adaptada às necessidades individuais de cada usuário, conforme apontado por esses pesquisadores.

Os autores enfatizam que a Melhoria Contínua dos *chatbots*, é impulsionada pelo PLN, por meio da análise de dados e feedback dos usuários. Essa abordagem contribui para aprimorar a capacidade dos *chatbots* em compreender nuances da linguagem, lidar com perguntas mais complexas e oferecer respostas mais precisas (Crowston; Allen; Hechman, 2012).

A eficiência e escalabilidade dos *chatbots* são maximizadas pela automação da compreensão e geração de linguagem natural proporcionada pelo PLN. Essa automatização torna os *chatbots* mais eficientes e capazes de lidar com um grande volume de interações de maneira rápida e consistente, resultando em um atendimento ao cliente mais ágil e eficaz (Crowston; Allen; Hechman, 2012)

O PLN é a tecnologia-chave que capacita os *chatbots* a interagir de forma inteligente e natural com os usuários, oferecendo uma experiência de conversação mais próxima à comunicação humana. A combinação dessas duas tecnologias têm revolucionado a forma como os usuários interagem com sistemas automatizados.

Os autores Idemudia e Makinde (2022) abordam a implementação da IA em bibliotecas, com foco nas perspectivas e desafios específicos enfrentados pelas bibliotecas acadêmicas na Nigéria. Ao analisar as possibilidades da IA, o estudo identifica áreas da biblioteca que podem se beneficiar dessas tecnologias, destacando componentes como sistemas especialistas, reconhecimento de padrões, PLN e robótica, todos com potencial para melhorar as operações bibliotecárias.

Como exemplo de superinteligência artificial, capaz de transformar radicalmente os serviços oferecidos pelas bibliotecas acadêmicas. Além disso, o estudo ressalta o potencial dos *chatbots* como agentes conversacionais para aprimorar a interação e o atendimento aos usuários. Essas ferramentas de IA podem simular conversas humanas e oferecer respostas automáticas e personalizadas, melhorando significativamente a experiência do usuário (Idemudia; Makinde, 2022).

Além de oferecer suporte direto aos usuários, os autores destacam o potencial transformador da IA, especialmente dos *chatbots*, na modernização e melhoria dos serviços bibliotecários, garantindo eficiência, eficácia e qualidade no atendimento aos usuários, mesmo diante dos desafios específicos enfrentados pelas bibliotecas acadêmicas na Nigéria.

Após a explanação sobre os conceitos relacionados à IA, chatbots e PLN, pretende-se no próximo capítulo abordar o universo da recuperação da informação, para posteriormente, relacionar essas ideias.

### **3.3 Recuperação da informação**

Esta seção aborda diversos conceitos essenciais relacionados à recuperação da informação, incluindo a definição do termo, a relevância, os principais problemas de recuperação da informação, a influência das novas tecnologias nesse campo e a interconexão entre a representação da informação e sua recuperação.

Desde a publicação do “Manual de Documentação”, de Paul Otlet em 1937 (Lopez Yopez, 1989) e do MEMEX de Vannevar Bush em 1945 (Barreto, 2007), que numerosos estudos vêm denotando métodos e técnicas para melhorar o processo de recuperação da informação (Santarém Segundo, 2010).

A recuperação da informação desempenha um papel de destaque e possui uma estreita ligação com a CI. Em um texto seminal dentro deste campo, Saracevic (1996) discute as interdisciplinaridades inerentes à CI, destacando que a recuperação da

informação não apenas faz parte dessa ciência, mas é sua faceta mais preponderante e a que mais promove interações interdisciplinares.

O termo "recuperação da informação" (*information retrieval*) foi formalmente definido por Mooers em 1951, que sinalizava que esse domínio de estudo aborda questões que englobam desde a descrição da informação até a especificação de buscas, abrangendo ainda os sistemas e técnicas empregados na localização de informações. A partir dessa definição, surgiram pesquisas em diversas áreas do conhecimento que abordam a recuperação da informação, uma vez que os processos relacionados a essa área são fundamentais para uma ampla gama de atividades.

Naturalmente, a CI desempenhou um papel central nesses estudos, colaborando com a CC para a criação de sistemas de recuperação da informação mais eficazes. Esses sistemas visam processar informações com maior celeridade e melhorar os processos para atender às necessidades informacionais dos usuários de forma mais abrangente.

Os autores Baeza-Yates e Ribeiro-Neto, (2013, p. 1) conceituam esse domínio

recuperação da informação trata da representação, armazenamento, organização e acesso a itens de informação, como documentos, páginas Web, catálogos online, registros estruturados e semiestruturados, objetivos multimídias etc. A representação e a organização dos itens de informação devem fornecer aos usuários facilidade de acesso às informações de interesse.

Afirmam que há pouco tempo a recuperação da informação era considerada uma área de interesse restrita apenas a bibliotecários e especialistas em informação. Um ponto de vista que prevaleceu por muitos anos, apesar do fato de que as ferramentas de recuperação de informações para aplicativos multimídia e de hipertexto se tornaram difundidas entre os usuários modernos de computadores pessoais.

Apesar da notável influência dos autores na área da CC, é evidente que os conceitos por eles apresentados estão estreitamente relacionados com a CI. Isso ilustra que o processo de recuperação da informação requer a consideração de diversos elementos, que estão intrinsecamente relacionados aos usuários e à forma como estes acessam as informações.

No início dos anos 1990, um fato mudou essa percepção para sempre: o surgimento da *World Wide Web*. Inventada em 1989 por Tim Berners-Lee, a *web* tornou-se um repositório universal da cultura e do conhecimento humano. O seu sucesso baseia-se na concepção de uma interface de utilizador padrão que seja

sempre a mesma, independentemente do ambiente informático utilizado para o executar, e que permita a cada utilizador criar o seu próprio documento.

Segundo Baeza-Yates e Ribeiro-Neto (2013), como resultado, milhões de usuários construíram bilhões de documentos que compõem o maior repositório de conhecimento humano da história. A consequência imediata disso é que encontrar informações úteis na *web* nem sempre é uma tarefa fácil e geralmente requer uma consulta a um buscador, que tem tudo a ver com a recuperação da informação e suas tecnologias. Dessa forma, quase da noite para o dia, a recuperação da informação ganhou destaque ao lado de outras tecnologias.

Os autores Baeza-Yates e Ribeiro-Neto (2013) apontam que os usuários de sistemas modernos de informação e recuperação, assim como usuários de mecanismos de busca, necessitam de informações de diferentes níveis de complexidade.

Uma descrição completa das necessidades do usuário não prevê necessariamente a melhor fórmula de consulta para o Sistema de Recuperação de Dados. Por sua vez, Baeza-Yates e Ribeiro-Neto (2013) explicam que o usuário pode primeiro solicitar que essas informações sejam traduzidas em uma consulta ou série de consultas a serem enviadas ao sistema. Em sua forma mais comum, essa tradução gera um conjunto de palavras-chave ou termos de indexação que resumem as necessidades de informação do usuário. No caso de uma consulta do usuário, o principal objetivo de um SRI é recuperar informações úteis ou relevantes para o usuário. O foco é a recuperação de informações, não a recuperação de dados.

Ainda de acordo com os autores, uma forma de ser eficaz em seus esforços para satisfazer as necessidades de informação dos usuários, é que seja possível ao SRI “interpretar” o conteúdo dos itens de informação, ou seja, uma coleta e distribuição de informações, classificadas por relevância, de acordo com a solicitação do usuário. Nessa “interpretação” do conteúdo do documento, informações sintáticas e semânticas são extraídas do texto do documento e utilizadas para atender às necessidades de informação do usuário.

Fusco (2010) e Coneglian (2020) enfatizam que o objetivo primordial da recuperação da informação reside na entrega de resultados que possuam um valor real e relevância para os usuários, buscando estabelecer uma conexão mais eficaz entre os itens bibliográficos e as demandas informacionais dos usuários.

Ampliando a conexão entre os campos da CI e CC no contexto da Recuperação da Informação, Santarém Segundo destaca que:

A Ciência da Informação e a Ciência da Computação aparecem como as ciências mais envolvidas com a busca pela melhoria da qualidade da informação recuperada. A Ciência da Informação apresenta uma visão mais metodológica e tem procurado estruturar os dados e criar métodos e 27 modelos que proporcionem um melhor armazenamento da informação, assim como vem estudando métodos que agreguem semântica à informação, e conseqüentemente possam ser aplicadas no processo de recuperação. A Ciência da Computação tem procurado atuar na aplicação dos modelos citados, diretamente no desenvolvimento de técnicas computacionais, como algoritmos, que possam viabilizar as metodologias sugeridas e pesquisadas (Santarém Segundo, 2010, p. 26-27).

A perspectiva apresentada pelo autor ressalta as distintas contribuições teóricas e metodológicas que as diversas áreas oferecem ao campo da recuperação da informação. Nesse contexto, a CC tem desempenhado um papel mais proeminente na criação de soluções práticas, fazendo uso de técnicas computacionais e algoritmos. Enquanto isso, a CI concentra-se em considerações epistemológicas e metodológicas, desenvolvendo modelos teóricos destinados a aprimorar o aspecto semântico das soluções. É importante ressaltar que esses papéis são complementares, e ambas as áreas colaboram profundamente nesse domínio, evidenciando a afirmação de Saracevic (2007, p. 1915) “a relevância ainda continua sendo uma noção básica na CI, e particularmente na recuperação de informação”.

Outro aspecto de relevância está relacionado à perspectiva que a recuperação da informação pode adotar. Nesse sentido, Baeza-Yates e Ribeiro-Neto afirmam que:

Na visão centrada no computador, a recuperação de informação consiste principalmente na construção de índices eficientes, no processamento de consultas com alto desempenho e no desenvolvimento de algoritmos de ranqueamento, a fim de melhorar os resultados. Na visão centrada no usuário, a recuperação de informação consiste principalmente em estudar o comportamento do usuário entender suas principais necessidades e determinar como esse entendimento afeta a organização e a operação do sistema de recuperação (Baeza-Yates, Ribeiro-Neto, 2013, p. 1).

A perspectiva apresentada pelos autores evidencia que as duas abordagens são complementares e devem ser consideradas em estudos relacionados à recuperação da informação. Não é suficiente possuir um sistema de alta qualidade, com índices excepcionais, se ele não atender adequadamente às necessidades dos usuários ou se a sua usabilidade for deficiente. Este trabalho busca abordar ambas as dimensões, buscando otimizar tanto o processo de processamento de informações, por meio da integração de novas tecnologias, quanto o aprimoramento da interação

entre o sistema e o usuário, ao oferecer uma abordagem alternativa que permite consultas em linguagem natural.

Nhacuongue e Dutra (2020) destacam quatro fases do PLN: a) final de 1940 a década de 1960 – centrada na tradução automática, através de análise da estrutura de frases para solucionar os problemas de ambiguidade sintática e semântica; b) final dos anos 1960 até final dos anos 1970 - influenciada pela Inteligência Artificial na construção de bases de conhecimento para manipulação e representações do significado; c) final dos anos 1970 até final dos anos 1980 – caracterizada pelo enfoque gramatical-lógico ou pelo desenvolvimento da teoria gramatical e pelo uso da lógica para desenvolvimento de algoritmos de análise; d) dos anos 1990 em diante - marcada pela crescente influência do léxico na aprendizagem de máquinas e reconhecimento da fala.

Um dos objetivos do processamento de linguagem natural é a criação do índice invertido ou dicionário de termos, que mapeia os termos de acordo com a sua ocorrência em partes de um documento. O índice invertido garante a velocidade de indexação, principalmente, em sistemas automatizados. Para isso, o primeiro passo consiste na coleta de documentos que serão indexados. Em seguida, ocorre a tokenização do texto. Um *token* é uma instância de uma sequência de caracteres que formam uma unidade semântica útil para processamento. A seguir, ocorre o pré-processamento linguístico, produzindo uma lista de tokens normalizados, que serão utilizados na indexação. O processamento linguístico inclui algumas das seguintes operações: eliminação de *stop-words* ou termos comuns como artigos e preposições; normalização ou canonização de termos, para permitir correspondências apesar das diferenças superficiais nas suas sequências de caracteres (por exemplo, correspondência entre os termos *ônibus* e *onibus*) ou *stemming* e lematização, cujo objetivo é reduzir as formas flexionais ou derivadas de uma palavra a uma forma básica comum (Manning; Raghavan; Schutze, 2008).

É fundamental destacar também que, ao explorar o campo da recuperação da informação, é imprescindível considerar os sistemas de recuperação da informação. Esses sistemas, conforme definidos por Santarém Segundo (2010) são modelos que englobam todo o sistema de representação, armazenamento, administração e recuperação de informações.

Baeza-Yates e Ribeiro-Neto (2013) afirmam que o desafio da recuperação de informações reside na sua complexidade fundamental: um SRI visa não apenas

recuperar todos os documentos que sejam pertinentes à necessidade de informação do utilizador, mas também obter o maior número possível de documentos irrelevantes. A dificuldade não está apenas na extração da informação contida nos documentos, mas também na habilidade de avaliar a sua relevância.

Nesse sentido, é importante reconhecer que nenhum SRI pode oferecer respostas perfeitas que atendam a todas as necessidades de todos os utilizadores o tempo todo.

Baeza-Yates e Ribeiro-Neto (2013) explicam que a recuperação de dados, dentro do contexto de um SRI, engloba a tarefa de identificar quais documentos na coleção contêm as palavras-chave fornecidas na consulta do utilizador. No entanto, isso frequentemente não é suficiente para atender às necessidades informativas do utilizador. De fato, um utilizador de um SRI está mais interessado em obter informações abrangentes sobre um determinado tema do que apenas encontrar dados que correspondam exatamente à consulta em questão. Por exemplo, um utilizador de um SRI pode estar disposto a aceitar documentos nos resultados que contenham sinônimos dos termos utilizados na consulta. Portanto, em um SRI, os resultados podem ser aproximados, e pequenas imprecisões podem passar despercebidas.

Por outro lado, Baeza-Yates e Ribeiro-Neto (2013) afirmam que em um Sistema de Recuperação de Dados, a presença de um único item incorreto entre milhões de objetos recuperados representa uma falha completa. Um Sistema de Recuperação de Dados, como um banco de dados relacional, lida com dados que possuem estrutura e semântica bem definidas. Em contrapartida, em um SRI, lidamos com texto em linguagem natural que não possui uma estrutura rigorosa. A recuperação de dados é uma solução adequada para os utilizadores de sistemas de banco de dados, mas não resolve o desafio de obter informações abrangentes sobre um determinado assunto ou tópico, que é o foco principal de um SRI.

Os modelos computacionais de recuperação de informação para Silva, Santos e Ferneda (2013) são fundamentais para organizar e recuperar informações de maneira eficiente. Além disso, os subsistemas de recuperação da informação desempenham papéis específicos no processo de busca e recuperação de dados. Vamos descrever criteriosamente sobre os modelos computacionais e os subsistemas de recuperação da informação.

Conforme Silva, Santos e Ferneda (2013) o Modelo Booleano, um dos paradigmas clássicos de recuperação de informação, tem uma ampla aplicação em

sistemas de busca e recuperação de dados. Sua fundamentação na teoria dos conjuntos e na Álgebra de Boole, desenvolvida por George Boole no século XIX, permite a abordagem binária da recuperação de informações, classificando documentos como relevantes ou não para uma consulta específica.

Os operadores lógicos, tais como *AND*, *OR* e *NOT*, são utilizados para combinar os termos de busca. O operador *AND* recupera documentos que contêm todos os termos da consulta, enquanto o *OR* recupera aqueles que contêm pelo menos um termo, e o *NOT* exclui documentos com um termo específico. A simplicidade e o formalismo claro tornam o modelo facilmente implementável em sistemas de recuperação de informação (Silva, Santos e Ferneda, 2013)

A relevância binária classifica os documentos como relevantes ou não, baseando-se na presença ou ausência dos termos de busca, sem fornecer uma ordenação dos resultados. O modelo encontra aplicação em ambientes onde a classificação binária da relevância é suficiente, como em bancos de dados relacionais e catálogos de bibliotecas.

Silva, Santos e Ferneda (2013) conclui explicando que apesar de suas limitações, como a falta de granularidade na representação da relevância e a ausência de ordenação dos resultados, o modelo booleano permanece relevante em sistemas de recuperação de informação. Especialmente em cenários onde a simplicidade e a classificação binária atendem às necessidades, sua influência duradoura é evidente.

Baeza-Yates e Ribeiro-Neto (2013) também buscam aprofundar-se nos conceitos a respeito do modelo booleano, indicando que neste contexto, os elementos da matriz podem ser de duas formas: 1 quando há o termo de busca no documento e 0, quando o termo é ausente. Para o modelo booleano, não há satisfação parcial das condições de consulta, ou o documento é relevante ou não o é.

Para os autores, esse critério binário impede uma recuperação de qualidade. Outro ponto que pode atrapalhar a busca por meio de operadores booleanos é o fato de que eles exigem um nível de precisão que nem todos os usuários conseguem expressar em suas buscas. Como vantagens desse sistema, Baeza-Yates e Ribeiro-Neto (2013) destacam o formalismo e a simplicidade do modelo.

A respeito do modelo vetorial, Silva, Santos e Ferneda (2013) o destaca como um dos modelos clássicos de recuperação de informação, sendo amplamente empregado devido à sua eficácia e simplicidade. Neste modelo, a representação tanto

dos documentos quanto das consultas do usuário ocorre em um espaço vetorial, permitindo a comparação parcial com base na similaridade de seus vetores.

Para os autores, no âmbito do modelo vetorial, cada documento é caracterizado por um conjunto de termos de indexação, onde cada termo possui um peso associado que indica a relevância no conteúdo do documento. De maneira análoga, uma expressão de busca é representada por um conjunto de termos, cada um com pesos correspondentes, refletindo a importância de cada termo na consulta realizada pelo usuário.

Uma das principais vantagens desse modelo, em comparação com o Modelo Booleano, é a consideração da relevância como um conceito contínuo, expresso numericamente por um valor entre zero e um. Essa abordagem viabiliza o desenvolvimento de técnicas avançadas de recuperação de informação, como *clustering*, *feedback* de relevância, classificação e reformulação de consultas, contribuindo para resultados mais precisos e personalizados (Silva, Santos e Ferneda, 2013).

É relevante mencionar que o Modelo Vetorial foi concretizado no sistema SMART (*System for the Manipulation and Retrieval of Text*), desenvolvido por Salton na década de 1960. Esse sistema desempenhou um papel pioneiro na aplicação prática do modelo vetorial, servindo como base para o desenvolvimento de diversas técnicas e abordagens na área de recuperação de informação (Silva, Santos e Ferneda, 2013).

Baeza-Yates e Ribeiro-Neto (2013) também refletem sobre o modelo vetorial, afirmando que este modelo reconhece as limitações do modelo booleano e, para superá-lo, propõe um quadro em que casamentos parciais podem ocorrer. Para tanto, é necessário que pesos não binários sejam atribuídos aos termos de indexação, essas definições de peso serão posteriormente utilizadas para identificar o grau de similaridade entre o documento armazenado e a busca do usuário. Na recuperação, os documentos serão ordenados a partir do referido grau de similaridade, fornecendo uma resposta mais precisa.

Como principais vantagens deste modelo, os autores destacam o fato da ponderação dos termos refletir positivamente na qualidade da recuperação; a possibilidade do casamento parcial permite recuperação de documentos mais próximos do que desejado na consulta; a forma de apresentação dos resultados de maneira ordenada de acordo com o grau de proximidade com a busca. Como única

desvantagem, os autores afirmam que esse modelo encara os termos de indexação como mutuamente independentes. De acordo com Silva, Santos e Ferneda (2013), o modelo probabilístico representa um modelo clássico de recuperação de informação que utiliza a teoria matemática das probabilidades para aprimorar a precisão dos resultados de busca. Este modelo parte do pressuposto da existência de um conjunto ideal de documentos que atende a cada possível busca realizada no sistema.

No contexto do modelo probabilístico, o processo de recuperação de informação é caracterizado por um grau de incerteza no julgamento da relevância dos documentos em relação à expressão de busca. Ao contrário do modelo booleano, que trata a relevância como um conceito binário (relevante/não relevante), o Modelo Probabilístico adota uma abordagem mais realista, considerando a probabilidade de relevância dos documentos em relação à consulta do usuário (Silva, Santos e Ferneda, 2013).

Para Silva, Santos e Ferneda (2013), uma das características distintivas do modelo probabilístico é a utilização do *relevance feedback*, um processo iterativo no qual o usuário fornece *feedback* sobre a relevância dos documentos recuperados. Esse processo iterativo refinado contribui para melhorar continuamente a precisão dos resultados de busca, aproximando-se do conjunto ideal de documentos relevantes.

Além disso, o modelo probabilístico reconhece que a atribuição de relevância é uma tarefa subjetiva que cabe ao usuário, diferenciando-se de outros modelos que pressupõem uma relevância exata. Essa abordagem flexível e adaptável torna o modelo probabilístico uma escolha eficaz para sistemas de recuperação de informação que enfrentam desafios relacionados à incerteza e subjetividade na avaliação da relevância dos documentos.

Segundo Baeza-Yates e Ribeiro-Neto (2013), o modelo probabilístico pode ser compreendido a partir da seguinte ideia: em uma consulta, há um conjunto de documentos ideal, que contém apenas os documentos relevantes, que a partir de uma descrição adequada serão recuperados. Sendo assim, o processo de consulta deve ser compreendido enquanto um “processo de especificação das propriedades do conjunto ideal”, que não são conhecidas no momento da busca. Deve-se então, esforçar-se para identificar uma estimativa inicial dessas propriedades, que irá gerar uma descrição probabilística preliminar.

A fim de melhorar essa descrição probabilística do conjunto ideal, é necessária uma interação com o usuário. Como exemplo, os autores sugerem que os usuários possam ver os documentos recuperados e decidir quais são relevantes e quais podem ser excluídos, e, a partir dessas informações, o sistema pode refinar a descrição do conjunto ideal. Com a repetição, espera-se que essa descrição fique cada vez mais precisa e, no entanto, essa hipótese possui como desvantagem o fato que a relevância para o usuário pode ser afetada por variáveis externas.

Conforme Silva, Santos e Ferneda (2013), o modelo fuzzy representa uma abordagem inovadora na área de recuperação de informação, fundamentada na lógica fuzzy. Seu principal objetivo é lidar de maneira mais flexível e adaptável com a imprecisão, incerteza e a natureza parcial das informações. Neste contexto, a representação dos documentos e das consultas é concebida de forma mais aberta, reconhecendo e incorporando a diversidade e incerteza inerentes ao processo de recuperação de informações.

No âmbito do modelo fuzzy, busca-se ampliar o conceito de representação dos documentos, abandonando a abordagem convencional baseada em palavras-chave. A proposta assume que cada consulta determina um conjunto difuso, enquanto cada documento possui um grau de pertencimento a esse conjunto. Essa perspectiva permite uma avaliação mais flexível da relevância dos documentos em relação às consultas dos usuários, considerando não apenas a presença ou ausência de termos, mas também a relação semântica entre eles.

Uma das notáveis vantagens apresentadas pelo modelo fuzzy, conforme Silva, Santos e Ferneda (2013), é sua capacidade natural de lidar com a imprecisão e incerteza. Essa característica permite a recuperação de documentos com graus variados de relevância, atendendo às nuances das consultas dos usuários. Adicionalmente, o modelo fuzzy pode empregar instrumentos como tesouros para identificar termos semanticamente relacionados, ampliando a abrangência da busca e considerando diferentes nuances de significado.

Embora o modelo fuzzy não tenha uma implementação tão difundida quanto outros modelos clássicos de recuperação de informação, sua proposta oferece uma perspectiva interessante para lidar com a complexidade e subjetividade das consultas de usuários. Essa abordagem possibilita uma recuperação mais flexível e adaptável de informações, especialmente em ambientes nos quais alcançar uma precisão exata pode se mostrar desafiador.

De acordo com Baeza-Yates e Ribeiro-Neto (2013), a teoria dos conjuntos fuzzy associa uma função de pertinência aos elementos da classe, a partir de valores em um intervalo de 0 a 1, em que 0 representa a não pertinência e 1, à pertinência completa e os graus entre esse intervalo representam os elementos marginais. Sendo assim, a pertinência em um conjunto fuzzy é gradativa e não abrupta, como no modelo booleano.

Existem três operações básicas em conjuntos fuzzy: o complemento de um conjunto, a união de dois ou mais conjuntos e a intersecção de dois ou mais conjuntos. Esse grau de pertinência em um conjunto disjuntivo é computado a partir de uma soma algébrica, e, o grau de pertinência em um conjunto conjuntivo é atribuído a partir de um produto algébrico. Essa forma de cálculo é mais adequada para sistemas como Repositórios Institucionais, diferente das funções min. e máx., mais comumente utilizadas (Baeza-Yates; Ribeiro-Neto, 2013). As Redes Neurais Artificiais (RNAs), conforme Silva, Santos e Ferneda (2013), são modelos computacionais inspirados no funcionamento do sistema nervoso humano, visando simular a capacidade de aprendizado e processamento de informações dos neurônios cerebrais. Na esfera da recuperação de informação, essas redes são aplicadas para aprimorar a precisão e eficiência dos sistemas de busca e recuperação de informações.

O princípio de operação das redes neurais artificiais baseia-se em conexões entre unidades de processamento, denominadas neurônios artificiais, organizadas em camadas e interconectadas por pesos que indicam a força das conexões. Durante o treinamento, essas redes ajustam esses pesos para aprender padrões nos dados, otimizando sua capacidade de realizar tarefas específicas, como classificação e recuperação de informações.

No contexto da recuperação de informação, as redes neurais desempenham um papel crucial ao relacionar as expressões de busca dos usuários com os documentos de um acervo. Isso possibilita uma busca mais eficiente e precisa, uma vez que cada expressão de busca ativa neurônios na rede, propagando sinais para os documentos relacionados e auxiliando na identificação da relevância desses documentos em relação à consulta do usuário.

Uma das notáveis vantagens das redes neurais na recuperação de informação é sua capacidade de lidar com padrões complexos e não-lineares nos dados, conforme destacado por Silva, Santos e Ferneda (2013). Essa característica possibilita uma análise mais sofisticada e adaptável das informações, permitindo que

as redes neurais sejam treinadas para reconhecer padrões sutis e contextuais. Isso, por sua vez, conduz a resultados mais personalizados e relevantes para os usuários, contribuindo para uma experiência de recuperação de informação mais eficaz.

Os algoritmos genéticos representam técnicas de otimização e busca inspiradas no processo de evolução natural, visando encontrar soluções para problemas complexos por meio de uma abordagem fundamentada em evolução e seleção natural. Na área de recuperação de informação, esses algoritmos desempenham um papel crucial ao aprimorar a eficiência e relevância dos resultados de busca (Silva, Santos e Ferneda, 2013).

Baeza-Yates e Ribeiro-Neto (2013) também explicam sobre a concepção das redes neurais enquanto uma representação gráfica do cérebro humano com seus neurônios, como um modelo de recuperação da informação. Para os autores, os nodos correspondem às unidades de processamento e as arestas, as conexões sinápticas. Para demonstrar a força entre esses elementos, um peso é atribuído a cada uma das redes neurais e o estado do nodo é definido pelo nível de ativação.

Os autores continuam dizendo que a rede neural é composta por três camadas: termos da consulta, termos dos documentos e uma para os documentos. Os termos de consulta iniciam o processo de inferência e enviam sinais aos termos dos documentos, uma vez que as arestas são bidirecionais nessa relação, e novos sinais serão direcionados aos nodos dos termos, que poderão direcionar novos sinais aos nodos dos documentos, repetindo todo o processo. A cada nova interação, os sinais vão se tornando mais fracos, até que cessem e há a recuperação (Baeza-Yates; Ribeiro-Neto, 2013).

O funcionamento dos algoritmos genéticos é embasado em princípios biológicos, incluindo seleção, recombinação e mutação. Inicialmente, uma população de soluções candidatas é gerada aleatoriamente e avaliada por uma função de aptidão, mensurando o quão bem cada solução aborda o problema em questão. As soluções mais aptas são então selecionadas para reprodução, combinando-se por meio de operadores genéticos como cruzamento e mutação, gerando novas soluções que herdaram características das soluções parentais conforme Silva, Santos e Ferneda (2013).

Na recuperação de informação, a aplicação dos algoritmos genéticos visa aprimorar a relevância dos resultados de busca, adaptando as representações dos documentos conforme as preferências e interesses dos usuários, conforme destacam

Silva, Santos e Ferneda (2013). Essa abordagem evolutiva permite ajustes dinâmicos nas representações dos documentos ao longo do tempo, de acordo com a interação dos usuários com o sistema.

Os autores ainda citam que uma das principais vantagens dos algoritmos genéticos na recuperação de informação é a sua capacidade de lidar com problemas complexos e multidimensionais, oferecendo soluções eficientes e adaptáveis em ambientes dinâmicos e com grande volume de informações. Adicionalmente, esses algoritmos possibilitam a personalização dos resultados de busca, considerando as preferências individuais dos usuários, contribuindo para uma experiência de recuperação de informação mais ajustada e relevante (Silva, Santos e Ferneda, 2013).

Em conjunto, esses modelos representam uma gama diversificada de abordagens que podem ser aplicadas de forma complementar na recuperação de informação, permitindo uma análise mais abrangente e adaptável dos dados. A combinação desses modelos pode contribuir significativamente para a evolução e aprimoramento dos sistemas de busca e recuperação de informações, atendendo às demandas cada vez mais complexas e diversificadas dos usuários (Silva, Santos e Ferneda, 2013)

De acordo com Silva, Santos e Ferneda (2013, p. 29), os sistemas de recuperação de informação “têm por função representar o conteúdo dos documentos do corpus e apresentá-los ao usuário de uma maneira que lhe permita uma rápida seleção dos itens que satisfazem total ou parcialmente a sua necessidade de informação [...]”

Os subsistemas de recuperação da informação desempenham um papel fundamental na eficiência e eficácia dos sistemas que visam recuperar informações relevantes para os usuários. Segundo Lancaster (1986), os SRI consistem em dois componentes essenciais: a entrada de informação (*input*) e a saída de informação (*output*). O ciclo intrínseco desses subsistemas envolve estágios cruciais, como representação, armazenamento e recuperação, estabelecendo um processo contínuo e retroalimentado.

A formação de uma coleção, seja física ou digital, inicia-se com a seleção de documentos, considerando as necessidades da comunidade a ser atendida. Os processos de catalogação, indexação, classificação e resumo são cruciais na organização desse acervo, abrangendo desde a estruturação até a posterior recuperação da informação pelo usuário.

O subsistema de entrada em sistemas de recuperação de informação, conforme destacado por Silva, Santos e Ferneda (2013), assume a crucial responsabilidade de receber e interpretar as consultas dos usuários, que essencialmente representam suas necessidades informacionais. A interação entre o usuário e o sistema é profundamente impactada pela qualidade e precisão da consulta, tornando esse componente central no processo.

Silva, Santos e Ferneda (2013) ainda observam que diferentes formas de interação podem ser adotadas pelo subsistema de entrada, como consultas de texto livre, consultas estruturadas, consultas por voz e consultas por imagem. Cada modalidade de interação apresenta desafios e oportunidades específicas, sendo a escolha da abordagem mais adequada dependente do contexto de uso e das preferências dos usuários.

Adicionalmente, Silva, Santos e Ferneda (2013) ressaltam que o subsistema de entrada pode ser aprimorado por meio da inclusão de recursos para auxiliar os usuários na formulação de consultas. Esses recursos englobam sugestões de termos, correção ortográfica, expansão de consultas e sugestões de consultas relacionadas, todos voltados para a melhoria da precisão e relevância das consultas, simplificando assim a recuperação das informações desejadas.

A usabilidade e eficácia do subsistema de entrada são consideradas aspectos fundamentais para assegurar uma interação eficiente e satisfatória entre o usuário e o SRI, conforme apontado por Silva, Santos e Ferneda (2013). Nesse contexto, uma interface intuitiva, responsiva e adaptável às preferências dos usuários desempenha um papel crucial ao contribuir para uma formulação precisa e eficaz das consultas, resultando, conseqüentemente, em resultados de busca mais relevantes e satisfatórios.

Diversas propostas na literatura delineiam o fluxo desses subsistemas no SRI. Lancaster (1978) identifica elementos-chave, designando-os como subsistemas, tais como seleção e aquisição, indexação, vocabulário, busca e interação usuário-sistema. Ele enfatiza a interrelação desses componentes, sublinhando que afetam mutuamente usuários, acervos, vocabulários, indexação e recuperação. A visão integral de Lancaster oferece uma compreensão crucial para o desempenho consistente do SRI na entrega de resultados aos usuários.

Para o autor, a interação parte da identificação do usuário, considerando sua área de atuação, atividade desenvolvida e outros elementos relevantes. Essas

respostas orientam princípios para formar o acervo, tratar documentos, determinar classificações e até mesmo estabelecer limites para empréstimos e prazos (Lancaster, 1978).

### **3.4 Repositório institucional**

Os cursos de pós-graduação vêm crescendo gradualmente, a partir de uma maior produção científica, assim como a ocorrência de eventos em todo mundo. Com isso, as Instituições de Ensino Superior, começam a criar instrumentos para reunir e divulgar a sua produção intelectual na Web, a fim de ampliar sua visibilidade (Arakaki et al., 2019).

A intensificação crescente das tecnologias de informação e comunicação destaca uma diversidade de fontes de informação, modificando, tanto em amplitude quanto em agilidade, a capacidade de comunicação da informação em todos os níveis e setores da sociedade. No âmbito da ciência, das universidades e instituições de pesquisa, essa realidade cria oportunidades significativas para a divulgação de resultados de pesquisa.

Contudo, mesmo diante desse avanço, uma parcela substancial da informação gerada nas universidades permanece indisponível para a comunidade acadêmica da própria instituição e para a comunidade científica em geral. Esse cenário é influenciado por vários fatores, incluindo a lentidão nos processos de publicação, os custos associados à publicação de documentos e a dispersão de documentos, sejam eles impressos ou eletrônicos, nas coleções.

A falta de visibilidade é particularmente notável no caso de documentos não publicados ou de difícil acesso, formando uma grande massa de informação gerada pelas universidades. Dentro dessa massa, informações valiosas frequentemente se perdem devido à ausência, na maioria dos casos, de formas eficientes de acesso a elas. Para enfrentar esse desafio, a implementação de repositórios digitais, ou repositórios de informação, surge como uma estratégia para minimizar essa falta de visibilidade e garantir o acesso à produção intelectual.

A utilização das Tecnologias de Informação e Comunicação (TICs) demanda uma orientação intelectual no interior da organização, conforme evidenciado por Santos e Vidotti

As estruturas de acesso automático, de armazenamento e de preservação da informação para sua localização e uso, requisitam uma fundamentação garantida pela racionalização, pela sistematização e pela compreensão: das estruturas de sinais e de símbolos contidos nos diferentes tipos de dados; da operacionalização no processo de comunicação e de transferência de informações; de linguagens natural e artificial; e de análises semântica e semiótica. (Santos; Vidotti, 2009, p. 7)

De acordo com Cerrão e Castro (2018), a CI tem como finalidade a criação de padrões e modelos de estrutura que concedam qualidade e confiabilidade no armazenamento e na disseminação da informação e para que isso aconteça, são necessários padrões e estruturas que possam ser empregados na recuperação da informação. A padronização da representação e da descrição dos recursos informacionais é primordial para o funcionamento de repositórios institucionais.

Os repositórios institucionais surgem como ambientes informacionais fornecedores de dados digitais, destinados a gerenciar, reunir, armazenar e a disseminar a produção científica de uma instituição, possibilitando a preservação da memória científica institucional. As universidades e instituições de pesquisa disponibilizam em seus repositórios institucionais materiais produzidos por seus pesquisadores de forma gratuita e acesso livre.

No contexto abordado por Salcedo e Bezerra (2020), os repositórios institucionais são caracterizados como parte de um "paradigma pós-custodial", transcendendo a concepção tradicional de serem meros depositários do conhecimento produzido internamente. Os autores fundamentam sua abordagem na filosofia do *Open Access Initiative* (OAI), proporcionando às instituições acadêmicas a capacidade de retomar o controle sobre as publicações científicas e reduzir a hegemonia das editoras. A implementação desse enfoque de acesso aberto não apenas outorga maior autonomia às instituições no gerenciamento e disseminação de sua produção acadêmica, mas também desempenha um papel crucial na democratização do conhecimento e na ampliação do acesso à informação.

Repositórios digitais são conjuntos de documentos coletados, organizados e disponibilizados eletronicamente. Nesse contexto, os documentos assumem novas configurações e são chamados de objetos digitais ou estruturas de dados digitalmente codificadas. Essas estruturas compreendem o conteúdo da informação, metadados e identificadores segundo Warner et al. (2006) em uma classificação abrangente, os repositórios podem ser temáticos, concentrando-se em um domínio específico do conhecimento, ou institucionais, focados na produção de uma instituição específica.

Segundo Marcondes e Sayão, (2009, p. 09) “um repositório institucional é uma biblioteca digital destinada a guardar, preservar e garantir livre acesso, via internet, à produção científica no âmbito de uma dada instituição.”

Os repositórios institucionais, também conhecidos como repositórios de acesso aberto, representam coleções que capturam e preservam a produção intelectual de uma ou mais universidades ou comunidades (Crow, 2002). Esses repositórios também podem ser entendidos como um conjunto de serviços oferecidos por uma universidade aos membros de sua comunidade para gerenciar e disseminar materiais digitais criados pela instituição e seus membros (Lynch, 2003). A característica distintiva dos repositórios institucionais é sua orientação para a informação produzida no ambiente das instituições, sendo desenvolvidos, implementados e mantidos por elas.

O canal de informação dominante para a disseminação do conhecimento científico são os periódicos científicos. No início dos anos 1990 surgiram impasses que prejudicaram a divulgação por esses meios, como o alto custo das assinaturas dos periódicos, as dificuldades em publicar, o aumento significativo da quantidade de títulos, entre outros. Como alternativa à questão, os repositórios institucionais foram planejados e inseridos no ambiente acadêmico. Por influência de alguns membros da academia que defendiam a disponibilização de acesso aberto da produção científica originária de recursos públicos, esses repositórios institucionais deveriam ter como características o acesso aberto, disseminação e visibilidade dos resultados das pesquisas científicas (Freitas, 2019).

A visibilidade da produção acadêmica é uma das principais vantagens dos repositórios institucionais, permitindo que a informação "invisível" seja tornada visível e acessível. Esses repositórios divulgam e preservam a produção intelectual das instituições, sendo amplamente adotados em universidades de renome em todo o mundo (Rumsey, 2006; Hockx-Yu, 2006). Importantes exemplos incluem o repositório do *Massachusetts Institute of Technology* (MIT), responsável pelo desenvolvimento do *DSpace*, um dos principais softwares de código aberto mantidos em colaboração com várias universidades (Drake, 2004).

Repositórios institucionais estão em desenvolvimento em todo o mundo, amparados pelo movimento de acesso aberto à produção científica e preservação da memória institucional nas Instituições de Ensino Superior. Os repositórios institucionais atuam como ferramentas de gestão ao potencializarem as atividades de

criação, compartilhamento, disseminação e uso de conhecimentos científicos provenientes do ambiente acadêmico (Leite; Costa, 2006, p. 211).

Para os autores Fachin, Blatmann e Caldin (2019, p. 89)

“[...] acesso aberto é um meio de disseminação do conhecimento sem restrições de fronteiras, taxas, com o conhecimento do autor na disponibilização do conteúdo, de forma a socializar rapidamente as pesquisas e os pesquisadores”.

Conforme observado por Salcedo e Bezerra (2020), o crescimento contínuo da produção institucional a cada ano coloca desafios significativos na gestão e recuperação de documentos digitais. Nesse ambiente digital em expansão, a busca por informações torna-se complexa, especialmente quando o pesquisador não possui uma compreensão clara do tipo específico de informação ou documento que procura. A recuperação de dados frequentemente resulta em conjuntos extensos de itens informacionais, muitas vezes pouco relevantes para o usuário, mas que seguem uma lógica de representação estabelecida pela instituição e por especialistas encarregados das tarefas de indexação e classificação. Isso, por sua vez, pode gerar ansiedade e frustração entre os pesquisadores.

O repositório institucional possibilita o aumento da eficácia da preservação da produção intelectual de pesquisadores de instituições acadêmicas para a visibilidade de ambos – tanto da instituição, como do pesquisador, prestando contas do desenvolvimento científico desses ambientes à população.

No processo de autoarquivamento, é comum que o próprio autor do documento realize a prática de representação. No entanto, em comparação com a indexação realizada por bibliotecários especialistas, essa representação tende a ser menos abrangente e mais sujeita a falhas na recuperação. Segundo Freitas (2019, p. 13), “se uma publicação não puder ser encontrada eletronicamente, ela é pouco útil e, em termos práticos, também não existe”. Esta colocação exemplifica bem a situação descrita anteriormente, o próprio autor do documento faz o autoarquivamento sem uma política de indexação clara e definida.

Fujita (2003) destaca o papel crucial do indexador na indexação, sendo um elemento fundamental dentro do sistema de informação. A responsabilidade principal do indexador é compreender o conteúdo de um documento e realizar uma análise conceitual que represente de maneira adequada o assunto, garantindo assim a

correspondência com as consultas dos usuários. No entanto, para assegurar essa correspondência, é imprescindível adotar uma política de indexação.

Essa política, conforme explicado por Fujita (2003), orienta os princípios e critérios que direcionam as decisões para otimizar o serviço e racionalizar os processos. Assim, torna-se uma decisão administrativa indispensável para um sistema de recuperação de informação, uma vez que somente após sua definição o sistema pode estabelecer suas características principais. É fundamental destacar que essa política deve ser formalmente descrita nos documentos oficiais do sistema, como manuais de indexação.

Além disso, a política de indexação funciona como um guia para as ações profissionais, potencializando os resultados satisfatórios dos repositórios institucionais em relação à qualidade dos metadados.

Contudo, os repositórios institucionais proporcionam a visibilidade da propriedade intelectual das Instituições de Ensino Superior (Camargo; Vidotti, 2011). Os ambientes informacionais digitais devem ser criados empregando a abordagem centrada no usuário, segundo Loureiro e Albagli (2008, p. 06)

“a caracterização de um ambiente informacional envolve identificar os principais atores, suas atribuições e relações, bem como os demais elementos que definem as condições de acesso, necessidades e usos da informação”.

De acordo com Leite (2009), o repositório institucional consiste, portanto, em um serviço de informação científica — em ambiente digital e interoperável — dedicado ao gerenciamento da produção intelectual de uma instituição de ensino superior. O repositório tem a finalidade de reunir, armazenar, organizar, preservar, recuperar e, especialmente, a vasta disseminação da informação científica produzida na instituição. Em um sentido geral, o aporte dos repositórios institucionais está em especial na reestruturação e evolução do sistema de comunicação científica por meio de processos de gestão da informação científica, proporcionando, o aumento da visibilidade dos resultados de pesquisa, do pesquisador e da instituição

Lynch (2003, p. 332) define repositórios institucionais como uma coleção de serviços que a instituição oferece à comunidade acadêmica para a gestão e disseminação de materiais digitais criado pela própria instituição e por seus membros. Ainda para o autor, o comprometimento da universidade com a gestão destes materiais é talvez mais importante do que o serviço propriamente em si.

Um repositório institucional é um reconhecimento que a vida intelectual e acadêmica das universidades será cada vez mais representada, documentada e compartilhada em formato digital, e que a responsabilidade primária das universidades é exercitar a administração destas riquezas: tanto para torná-las disponíveis quanto para preservá-las. Um repositório institucional é o meio que cada uma das universidades irá afirmar esta responsabilidade tanto para os membros de sua comunidade como para o público em geral. Este é um novo canal para estruturação da contribuição da universidade para o mundo mais amplo, como tal, convida a avaliação política e cultural dessa relação (Lynch, 2003, p. 332, tradução nossa).

Leite (2009, p. 33) afirma ainda, que os repositórios institucionais procuram atender a essas duas demandas, que estão necessariamente ligadas: a maximização do acesso (auxiliando para a produção de conhecimento) proporciona a maximização do impacto do que é produzido (visibilidade).

### **3.5 Indexação**

De acordo com Pinto (2001), a indexação é um conceito abrangente aplicado em diversas áreas do conhecimento, abrangendo disciplinas como Economia, Demografia, Comércio e CI. Na perspectiva desta última, a indexação se apresenta como um processo intrínseco à Organização da Informação, como delineado por Dias e Naves (2013).

Piedade (1983) define a indexação como a inclusão do registro de um documento em um repositório de informações, sendo corroborada por Baptista, Araújo Júnior e Carlan (2010), que a consideram uma prática integrante do processo operacional da análise da informação, envolvendo a definição de termos que possibilitam a descrição do documento.

Originada da necessidade humana de recuperar informações registradas em documentos, a prática da indexação teve sua evolução vinculada à memória como instrumento primordial de gravação e recuperação de informações, avançando posteriormente junto com os desenvolvimentos tecnológicos que culminaram nos processos contemporâneos de organização informacional (Silva; Fujita, 2004).

Segundo Fujita, Rubi e Boccato (2009), os registros históricos indicam a presença da indexação desde os tempos das tábuas de argila, datando do século II

antes de Cristo (a.C). Embora não denominada como tal na época, as práticas consistiam em representações condensadas que permitiam o acesso a documentos.

A história da indexação se entrelaça com o surgimento da história bibliográfica e ganha maior destaque com o advento das publicações periódicas, gerando uma crescente demanda pela organização desses objetos informacionais. Este processo toma forma mais concreta a partir do século XIX, acompanhando o considerável aumento de documentos a serem indexados (Fujita; Rubi; Boccato, 2009).

As autoras ainda pontuam que o termo indexação refere-se à descrição e identificação de um objeto informacional com base no assunto que aborda. Complementando essa definição, Rubi (2009) destaca que a indexação envolve a identificação do conteúdo do documento por meio do processo de análise de assunto e a representação desse conteúdo por meio de conceitos.

O surgimento do conceito de indexação foi a partir da elaboração de índices e atualmente está ligado ao conceito de análise de assunto. Com a necessidade de recuperação cada vez mais rápida, precisa e especializada por parte de Instituições, “a construção de índices passou a contar com um aparato metodológico e instrumental mais diversificado e muito mais voltado para o contexto de cada documento” (Silva; Fujita, 2004, p. 136).

Assim, indexação é um processo de representação temática da informação que é muito utilizado nas unidades de informações para busca e recuperação de documentos a partir do seu conteúdo. A norma ABNT NBR 12676 (1992, p. 02), explica a indexação como o “ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação”.

Para Silva e Fujita (2004, p. 154), o objetivo da indexação é o de traduzir o conteúdo informacional do documento, tendo em vista sua recuperação, e para tanto, é necessário efetuar um exame do documento a fim de reconhecer conceitos pelos quais a tematicidade do recurso será reconhecida. Essa tematicidade é verificada pelo indexador por meio da leitura do documento, com precisão informacional da comunidade usuária do sistema de informação.

Para Fujita (2003, p. 61), a indexação é uma atividade ligada ao tratamento temático da informação, ou seja, uma "combinação metodológica altamente estratégica entre o tratamento do conteúdo do documento e a sua recuperação por um usuário", esclarecendo uma conexão direta entre o processo e a finalidade da

indexação, objetivando a disponibilização de informações apropriada às necessidades dos usuários.

A obtenção do documento mais relevante para a questão de busca ocorre quando a indexação é realizada de forma a identificar conceitos altamente pertinentes ao conteúdo do documento. Esse processo resulta em uma correspondência precisa com o assunto pesquisado nos índices (Fujita, 2003).

Mais adiante, Fujita (2012) afirma que a indexação em análise documentária, sob o aspecto dos sistemas de informação, é também conhecida como o componente mais importante porque preserva os resultados de uma estratégia de busca.

Para garantir a eficácia desse processo, é fundamental empregar instrumentos específicos de representação e recuperação da informação, conhecidos como linguagens controladas, tais como tesouros e listas de cabeçalhos de assunto. Além disso, é viável utilizar a linguagem natural na representação desses documentos. O autor Raitt (1980) afirmou que a principal estratégia de busca em sistemas on-line de recuperação de informação é a linguagem natural. Contudo, a linguagem controlada é a que aumenta a revocação e a precisão na busca da informação. Na tabela a seguir, estão descritas as vantagens e desvantagens da utilização da linguagem natural na representação de conteúdos de documentos:

**Tabela 1:** Linguagem natural: vantagens e desvantagens

VANTAGENS	DESVANTAGENS
1. Permite o imediato registro da informação em uma base de dados, sem necessidade de consulta a uma linguagem de controle.	Os usuários da informação, no processo de busca, precisam fazer um esforço intelectual maior para identificar os sinônimos, as grafias alternativas, os homônimos etc.
2. Processo de busca é facilitado com a ausência de treinamentos específicos no uso de uma linguagem de controle	Haverá alta incidência de respostas negativas ou de relações incorretas entre os termos usados na busca (por ausência de padronização).
3. Termos de entrada de dados são extraídos diretamente dos documentos que vão constituir a base de dados.	Custos de acesso tendem a aumentar com a entrada de termos de busca aleatório
4. Temas específicos citados nos documentos podem ser encontrados.	Uma estratégia de busca que arrole todos os principais conceitos e seus sinônimos deve ser elaborada para cada base de dados (ex: nomes comerciais de substâncias químicas não ocorrem no Chemical Abstracts).
5. Elimina os conflitos de comunicação entre os indexadores e os usuários, pois ambos terão acesso aos mesmos termos.	Perda de confiança do usuário em uma possível resposta negativa.

**Fonte:** Lopes (2002, p.48)

Fujita (2003), Lancaster (2004) e Gomes (2019) convergem ao destacar que a indexação engloba duas fases fundamentais: a análise conceitual e a tradução. Oliveira (2017, p. 41) corrobora esse entendimento, ressaltando que “análise conceitual e tradução são etapas fundamentais para condução da indexação, sendo a primeira destinada à compreensão do conteúdo do documento e a segunda, à adequação do termo de acordo com a linguagem da unidade de informação”.

Rubi e Fujita (2003), alinhados com Oliveira (2017), observam a falta de consenso entre as correntes teóricas sobre o número de estágios no processo de indexação. No entanto, todas as abordagens concordam na identificação dos conceitos presentes em um documento e sua subsequente tradução para a linguagem do sistema. Robredo (2005), por exemplo, explica as quatro etapas no processo de indexação:

- Análise conceitual do conteúdo significativo do documento, envolvendo a identificação do assunto;
- Expressão dessa análise por meio de um conjunto de palavras, frases ou códigos que representem o assunto;
- Tradução das descrições dos assuntos relevantes para a linguagem de indexação;
- Organização das descrições padronizadas dos assuntos de acordo com a sintaxe da linguagem de indexação.

Fujita (2003) destaca que a identificação de conceitos no processo de indexação compreende duas etapas distintas. A primeira etapa ocorre durante a identificação dos conceitos presentes no conteúdo, visando reconhecer o tema subjacente. A segunda etapa ocorre após essa identificação preliminar, durante o processo de tradução para termos específicos das linguagens documentárias.

E a autora também afirma quanto à compreensão do documento a ser indexado, resalta-se a importância da leitura cuidadosa de determinadas partes, como o título, a introdução e as primeiras frases de capítulos e parágrafos; ilustrações, tabelas, diagramas e suas explicações; conclusão; palavras ou grupos de palavras sublinhadas ou impressas com tipo diferente.

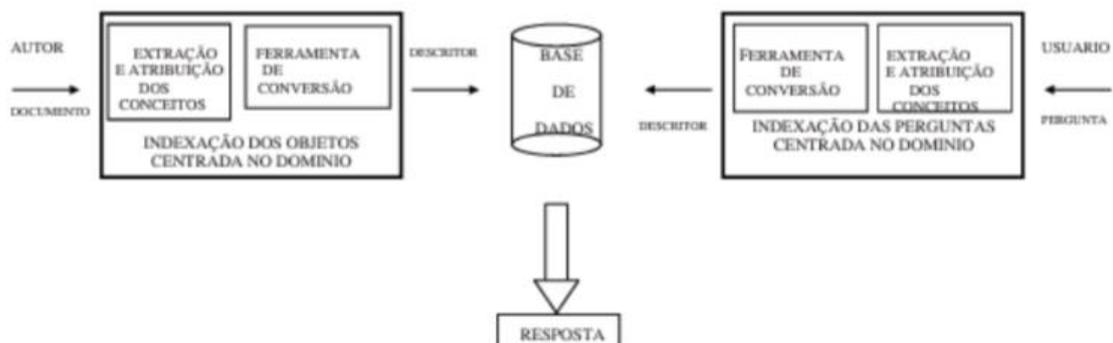
A Norma Brasileira NBR 12676/1992 estabelece condições exigíveis para a prática normalizada do exame de documentos, determinação de seus assuntos e seleção de termos de indexação (Associação Brasileira..., 1992).

Esta norma, que subsidia o processo de indexação, identifica três estágios no processo: exame do documento e estabelecimento do assunto de seu conteúdo, identificação dos conceitos presentes no assunto e tradução desses conceitos nos termos de uma linguagem de indexação (Associação Brasileira... 1992). Segundo a norma, os estágios do processo de indexação se sobrepõem.

A NBR 12676/1992 busca esclarecer as diretrizes que orientam o processo de análise de assunto, uma vez que a literatura não apresenta referências aplicadas de pesquisa ou relatos de experiência que tenham utilizado esses instrumentos no processo de indexação.

A NBR 12676/1992 (Associação Brasileira..., 1992, p. 2), conceitua indexação como “ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação”. A Norma divide o processo de indexação em três estágios, os quais se sobrepõem, consistindo em: a) exame do documento e estabelecimento do assunto de seu conteúdo; b) identificação dos conceitos presentes no assunto; c) tradução desses conceitos nos termos de uma linguagem de indexação.” (Associação Brasileira..., 1992, p. 2).

Gil Leiva (2012) apresenta o processo de indexação da seguinte forma: a partir do documento que é elaborado por um autor, há o processo de indexação por meio da extração e atribuição de conceitos que serão utilizados a partir de uma ferramenta de conversão para que se tornem termos descritores. Esses descritores serão incluídos em uma base de dados, que após a busca pelos termos utilizados pelo usuário em sua pergunta, irão trazer a resposta mais adequada a partir dessas variáveis. Isso pode ser melhor visualizado na figura a seguir:



**Figura 1:** Processo total de Indexação

**Fonte:** Gil Leiva (2012)

Lancaster (2004) estabelece uma diferenciação entre indexação por extração, também conhecida como indexação derivada, e indexação por atribuição. Na indexação por extração, as palavras ou expressões empregadas pelo autor no documento analisado são escolhidas pelo indexador para representar seu conteúdo. Em outras palavras, nesse tipo de indexação, são utilizados os mesmos termos empregados pelo autor na apresentação de suas ideias no texto, ou seja, uma linguagem natural. Isso implica a seleção de palavras encontradas no título, nas palavras-chave ou na própria estrutura textual do documento. Por outro lado, a indexação por atribuição envolve a atribuição de descritores ao documento utilizando uma linguagem originada de uma fonte que não é o próprio documento.

Portanto, são estabelecidos os cabeçalhos a serem utilizados com o objetivo de controlar sinônimos, diferenciar homógrafos e estabelecer ligações entre termos por meio de relações hierárquicas, associativas e de equivalência, ou seja, emprega-se uma linguagem artificial ou controlada. Conforme Lancaster (2004, p.19), "a indexação por atribuição envolve o esforço de representar a substância da análise conceitual por meio do emprego de termos extraídos de alguma forma de vocabulário controlado".

A linguagem de indexação é caracterizada por Silva e Brito (2018) como um conjunto de termos padronizados que representam um conceito da linguagem natural, visando à padronização dos sistemas de indexação.

Moura, Silva e Amorim (2002) apontam para a complexidade intrínseca ao desenvolvimento de linguagens de indexação, uma vez que demanda compreensão de variáveis que podem impactar no processo. Essa complexidade emerge do compromisso com a garantia literária, onde um vocabulário controlado deve efetivamente representar os conceitos presentes na literatura de assunto. Tal princípio, contudo, pode dificultar a interação entre usuários e sistemas de informação, notadamente para iniciantes que não possuem domínio abrangente dos assuntos e categorias, assim como para profissionais quando as categorias de assunto não foram devidamente inseridas no documento.

Considerando as ideias apresentadas, Fujita e Santos (2016) conceituam a indexação como um procedimento encarregado da descrição de documentos, empregando a análise do conteúdo temático para representá-los por meio de

conceitos derivados de uma linguagem documental. Diante dessas considerações, explora-se o processo de política de indexação como um instrumento fundamental para orientar a execução da prática indexadora.

Rubi (2012) afirma que a política de indexação representa uma orientação administrativa adotada por uma unidade de informação, pautando as decisões por meio de diretrizes pré-definidas que visam a condução eficiente do trabalho do bibliotecário de forma lógica e objetiva. Sua elaboração deve ocorrer no contexto da administração da biblioteca, alinhando-se aos objetivos da instituição e refletindo sua filosofia nas resoluções estabelecidas, indo além da simples consideração da indexação como um processo técnico.

Rubi (2009) destaca que a política de indexação precisa abordar tanto as condutas teóricas quanto as práticas a serem seguidas pelos profissionais indexadores durante a execução da indexação. A autora ressalta a importância de estruturar e registrar essa política em um manual de indexação, permitindo não apenas a consulta, mas também a avaliação e eventual modificação, caso seja identificada a necessidade de ajustes.

Destaca-se a relevância da política de indexação como fundamento para a elaboração de um manual específico nesse contexto. Nesse sentido, Fujita e Santos (2016) esclarecem que as políticas diferem dos manuais no sentido de que as primeiras são bases para que se construam os segundos.

Assim, a formalização da política de indexação segundo Nunes (2004) encontra expressão concreta no manual de indexação de uma Unidade de Informação. Ao buscar uma definição concisa, percebe-se que essa política se configura como "uma diretriz que explicita as escolhas técnicas (por isso política) que a biblioteca faz, considerando fundamentalmente duas variáveis: o seu usuário e o seu acervo".

Nunes (2004) ressalta ainda que as diretrizes estabelecidas por meio da política de indexação proporcionam economia de tempo e recursos, tanto materiais quanto financeiros, ao promover a racionalização do processo indexador.

Pinto (2001) destaca que a política de indexação deve ser delineada em uma Unidade de Informação alinhada aos objetivos e à missão da organização em que está inserida. Ele ressalta a importância de que essa política esteja de acordo com as necessidades dos usuários e seja formalizada em um manual, permitindo que os indexadores tenham acesso às regras estabelecidas e possam segui-las. Essa abordagem contribui para mitigar a subjetividade inerente a essa atividade.

Segundo Fujita (2016), a política de indexação tem como objetivo orientar e sistematizar a atividade indexadora, buscando o aprimoramento da indexação e a facilitação da recuperação da informação. A autora ainda esclarece que a função dessa política reside em:

circunstanciar as condições daquele momento para o respaldo das decisões a respeito de elementos da política de indexação, como por exemplo, a exaustividade e a especificidade da indexação, o processo de indexação e a linguagem de indexação no que tange a sua especificidade e exaustividade (Fujita, 2016, p. 16).

Além das considerações sobre sua definição, Fujita (2016) destaca que a política de indexação é uma necessidade indispensável para uma unidade de informação, que, por sua vez, requer a sistematização dos processos e diretrizes da indexação. A autora também enfatiza que a política de indexação pode ser ajustada para aprimorar as condições de infraestrutura ou atender às demandas dos usuários.

Em relação à sua elaboração, Fujita (2016) explica que a política de indexação emana sempre de um consenso entre usuários, indexadores, bibliotecários de referência e dirigentes, atores principais envolvidos no processo cotidiano, onde as ações e mudanças acontecem. Essa elaboração não depende de um padrão de conduta uniforme, mas, essencialmente, das necessidades, comportamentos, decisões estratégicas e, sobretudo, da sociocognição.

Quanto aos participantes do processo de elaboração de uma Política de Indexação, destaca-se a relevância do profissional indexador, pois é ele quem torna evidente o conhecimento relacionado aos aspectos intrínsecos à política de indexação, graças à sociocognição inerente aos profissionais nessa área. A característica sociocognitiva do indexador é fundamental nesse processo, uma vez que ele desempenha um papel crucial na geração do conhecimento da organização informacional segundo Fujita (2016).

A perspectiva sociocognitiva é essencial na indexação, pois permeia o contexto da sociocognição, permitindo a compreensão dos autores e processos envolvidos na indexação, visando ao tratamento adequado da informação. Como destacado por Fujita, Rubi e Boccato (2009) a abordagem sociocognitiva, portanto, tem como foco o sujeito que realiza uma determinada atividade e sua cognição em relação ao seu contexto de produção.

Gomes (2019, p. 94) ressalta que “a política de indexação é considerada um elemento norteador da indexação, que auxilia os profissionais indexadores na tomada das decisões sensíveis ao tratamento temático da informação”.

Segundo Guimarães (2004), diante dessas definições, é possível destacar a crescente imprescindibilidade da política de indexação para as unidades de informação. A política de indexação exerce uma influência direta na otimização de serviços, racionalização de processos e consistência das operações de tratamento e organização de conteúdos informacionais. Isso se torna particularmente relevante ao considerar a dinâmica indexadora, via termo e conceito, intrínseca ao processo de organização da informação, essencial para as unidades de informação.

Guimarães (2004) ainda enfatiza a necessidade de um monitoramento rigoroso e seleção aprofundada de conteúdos, visando possibilitar a construção de novos conhecimentos por meio de fontes confiáveis e autorizadas. Este monitoramento e seleção são mediados pela elaboração de uma política de indexação que, por meio de seus elementos, atenda às necessidades do usuário, considerando os princípios da instituição na qual está inserida.

Dessa forma, percebe-se que a indexação é um tema central na PI, destacando a relevância do indexador e aspectos específicos, como usuário, conceito e tema.

Fujita et al. (2022) o resultado do processamento temático em repositórios e catálogos, abarca metadados temáticos que podem ser examinados sob diversas perspectivas: semântica, sintática, morfológica e pragmática. Processamento temático significa atribuição de assuntos por meio de linguagem livre ou vocabulário controlado. O registro no catálogo da biblioteca do metadado assunto sempre foi uma atribuição dos bibliotecários catalogadores, que são especialistas em informação, mas, nos repositórios, as teses e dissertações são arquivadas pelos próprios autores, que, no processo de autoarquivamento, completam os metadados e atribuem temas, através de palavras-chave em linguagem livre.

Eventualmente os repositórios que orientam os autores sobre o uso de vocabulários controlados, essa atribuição pode incluir termos controlados por vocabulário. Nesta óptica, são realizados estudos sobre a análise dos metadados assuntos para compreender o processo de elaboração de teses e dissertações, em particular, a indexação e o controle de vocabulário.

Em universidades, os repositórios são institucionais e contém toda a produção científica das áreas do conhecimento abrangidas pelas atividades acadêmicas à semelhança de um catálogo bibliográfico online. Essa

circunstância é própria do ambiente universitário onde a produção científica abrange diferentes áreas do conhecimento e são especializadas com inovação científica e tecnológica. (Fujita; Tolare, 2019, p. 96).

No processo de indexação de assuntos em repositórios institucionais, o autor ao fazer o arquivamento do seu trabalho acadêmico no repositório institucional da instituição e essa não possuir uma política de indexação de assunto claramente definida para orientar o autor, é provável que o mesmo insira o metadado assunto a partir da linguagem natural, sem a tradução a partir de uma linguagem controlada. Esse fator pode gerar uma série de impeditivos para a recuperação da informação, uma vez que dificilmente há uma conformidade entre o da forma como o conhece e não a partir do uso adequado em uma linguagem documental, dificultando a recuperação por outros usuários.

Além disso, o preenchimento do campo “assunto” em linguagem natural no repositório pode ser uma implicação negativa nesse tipo de indexação, causando dificuldade na recuperação da informação. Como exemplos, pode ocorrer do autor descrever erros ortográficos, descritores abrangentes e com multiplicidade de sentidos, abreviações de termos, uso de siglas, entre outros.

Nesse sentido, é preciso repensar a forma com que é feita a representação da informação nesses ambientes, visando otimizar a recuperação efetiva da informação e dar apoio ao momento do autoarquivamento dos documentos.

### **3.6 Folksonomia**

A folksonomia desempenha um papel crucial na contemporaneidade, especialmente no ambiente digital, onde a quantidade exponencial de informações torna essencial a implementação de sistemas eficientes de organização. Este fenômeno, baseado na colaboração dos usuários, utiliza *tags* ou palavras-chave para categorizar recursos digitais, proporcionando uma estrutura flexível e dinâmica para a organização da informação.

A importância da folksonomia reside na sua capacidade de empoderar a comunidade de usuários, permitindo-lhes participar ativamente na classificação e recuperação de conteúdo. Ao promover a categorização colaborativa, a folksonomia não apenas reflete a diversidade de perspectivas, mas também contribui para uma recuperação de informações mais precisa e adaptável.

Com isso, pode-se estabelecer o fundamento para examinar os diferentes tipos de folksonomia, investigando de que maneira esses sistemas contribuem de forma positiva para a organização e acessibilidade das informações no cenário digital.

Na folksonomia restrita, os usuários têm um conjunto limitado de *tags* predefinidas para escolher ao categorizar o conteúdo. Essa abordagem busca impor uma estrutura mais controlada e consistente na classificação, visando melhorar a precisão na recuperação da informação.

A folksonomia restrita, também referida como genérica e específica, conforme apontado por Moreiro González (2011), caracteriza-se pelo fato de que apenas o autor da história participa do processo de indexação. Seguindo essa abordagem, o autor é responsável por criar as etiquetas conhecidas como *endo-tags*, conforme descrito por Moulaison (2008). Essas *endo-tags* são etiquetas específicas geradas pelos próprios autores, com o objetivo de facilitar o compartilhamento de informações.

Ao contrário da restrita, a folksonomia ampla oferece total liberdade aos usuários para criar suas próprias *tags*, sem restrições predeterminadas. Isso resulta em uma variedade maior de termos, refletindo a diversidade de perspectivas e interpretações dos usuários sobre o conteúdo.

Conforme Van der Wal (2005), a folksonomia ampla manifesta-se em contextos colaborativos que facilitam a participação de todos os intervenientes, permitindo a construção de um vocabulário emergente por meio da avaliação das preferências de descrição. Em contraste, o segundo tipo restringe a etiquetagem a um grupo específico de intervenientes, resultando em descrições mais específicas, pois se aproximam da autoria do texto. O modelo proposto por Santarém Segundo (2017), conhecido como Representação Interativa, visa implementar uma folksonomia assistida para converter repositórios digitais em ferramentas eficazes de descrição de conteúdo, proporcionando orientação na escolha de termos pelos usuários, por meio de sugestões de termos e de uma interface direcionada, buscando ampliar as relações semânticas através da consistência no uso de etiquetas.

No âmbito geral, segundo o autor, a representação interativa pressupõe um sistema contínuo de alimentação de informações, permitindo o aprimoramento constante da descrição. Esse aprimoramento ocorre a partir das etiquetas atribuídas pelos usuários que contribuem com conteúdo para o repositório. Além disso, envolve a mediação de profissionais na avaliação das etiquetas, culminando na criação de uma linguagem que integra o controle de vocabulário. Todos esses recursos

convergem para a construção de uma ontologia no domínio específico ao qual o repositório pertence.

A folksonomia assistida combina elementos de controle e liberdade. Nesse tipo, os usuários podem sugerir novas *tags*, mas há intervenção de sistemas de recomendação ou algoritmos para organizar e otimizar as *tags* sugeridas. Isso busca equilibrar a flexibilidade da ampla folksonomia com a necessidade de estrutura e consistência.

A abordagem da folksonomia assistida, conforme afirmado por Santarém Segundo e Vidotti (2011), enriquece a descrição de recursos de várias maneiras. A intervenção de sistemas de recomendação ou algoritmos na folksonomia assistida desempenha um papel essencial ao aprimorar a consistência e qualidade das *tags* atribuídas pelos usuários, culminando em etiquetas mais precisas e relevantes. Isso, por sua vez, contribui para uma descrição mais abrangente e precisa do recurso.

Adicionalmente, a folksonomia assistida desempenha um papel crucial na identificação de sinônimos e variantes das *tags*, mitigando a duplicação de informações e facilitando a recuperação do recurso por meio da busca por termos relacionados. Ao enriquecer a descrição do recurso com *tags* mais específicas e pertinentes, a abordagem permite a agregação de informações contextuais, como categorias, temas e conceitos associados ao recurso, proporcionando um entendimento mais abrangente do mesmo (Santarém Segundo; Vidotti, 2011).

Por fim, a folksonomia assistida simplifica a navegação e descoberta de recursos de interesse ao estabelecer conexões significativas entre os diferentes elementos, capacitando os usuários a encontrar conteúdo relevante com maior facilidade.

As características da folksonomia na literatura da CI, conforme destacadas por Mathes (2004), evidenciam que ela se trata de uma classificação social realizada pelos usuários de sites ou ferramentas, com o objetivo de categorizar objetos informacionais atribuindo-lhes palavras-chave. A natureza coletiva desses termos é uma característica essencial inerente à folksonomia, representando o conteúdo de recursos disponíveis na *Web* em diferentes formatos, como textos, livros, artigos, fotografias, vídeos, músicas, links, entre outros.

Na análise da folksonomia, Santos e Corrêa (2015) destacam considerações relevantes sobre a participação dos usuários no processo de descrição de objetos digitais, enfatizando a representação colaborativa da informação. Esses autores

abordam estratégias específicas utilizadas pelos usuários durante o processo de etiquetagem e examinam os efeitos dessa participação na organização e recuperação eficaz dos documentos em ambientes digitais.

A perspectiva de Santos e Corrêa (2015) reforça a importância da participação direta dos usuários no processo de descrição de objetos digitais, especialmente na representação colaborativa da informação. Eles contribuem para debates substanciais na literatura científica, explorando as estratégias adotadas pelos usuários durante a representação dos recursos e analisando os impactos dessa participação na organização e recuperação dos documentos em ambientes digitais.

[...] dizer que a folksonomia é um tipo de vocabulário descontrolado. Isso não quer dizer que o esquema seja uma desordem total [...] na verdade, trata-se de um mecanismo de representação, organização e recuperação de informações que não é feito por especialistas anônimos, o que muitas vezes pode limitar a busca por não trazer determinadas palavras-chave, mas sim um modo onde os próprios indivíduos que buscam informação na rede ficam livres para representá-la, organizá-la e recuperá-la, realizando estas ações com base no senso comum. (Aquino, 2007, p.10)

A representação colaborativa da informação em repositório por meio da folksonomia tem algumas vantagens. Primeiro, permite que os usuários organizem a informação de acordo com suas próprias necessidades e perspectivas, tornando a busca e a recuperação de informações mais personalizadas. Além disso, a folksonomia pode ajudar a identificar padrões e conexões entre diferentes conteúdos, fornecendo *insights* adicionais.

No entanto, é importante ressaltar que a folksonomia também apresenta alguns desafios. Como as *tags* são criadas pelos usuários, pode haver variações nas terminologias e na forma como a informação é categorizada, o que pode dificultar a busca e a recuperação de informações precisas. Além disso, a falta de regras predefinidas pode levar a uma sobrecarga de informações ou a uma categorização inconsistente.

Neste contexto, Santos (2017) explica que é possível deduzir que as deficiências dos sistemas que empregam a folksonomia derivam, em termos gerais, de erros ortográficos, polissemias, *tags* comuns personalizadas, ambiguidade terminológica em etiquetas que representam os materiais informacionais e, mais recentemente, da utilização de *chatbots* programados para considerar *tags* que são importantes para a disseminação rápida de notícias falsas na rede. Esses elementos têm impactos negativos no processo de recuperação da informação, exigindo uma atenção mais cuidadosa dos profissionais da informação na construção, atualização

e aprimoramento de ambientes colaborativos com o objetivo de minimizar tais implicações adversárias e agregar valor às potencialidades da participação dos usuários nesses ambientes, observando as garantias éticas, culturais e de uso.

A folksonomia constitui uma abordagem colaborativa para representação da informação em repositórios e em diversos sistemas que permitem a participação colaborativa dos usuários em diferentes níveis: de um para um; um para muitos; muitos para um e muitos para muitos, entre outros exemplos. Essa prática possibilita que os usuários atribuam *tags* ou etiquetas aos conteúdos digitais, conferindo uma organização mais flexível e personalizada da informação. No entanto, é importante notar que tal abordagem pode enfrentar desafios em termos de consistência e precisão na categorização.

Dessa forma, de maneira geral, nota-se que, na prática de indexação das produções científicas em repositórios institucionais, é permitido empregar os mesmos termos utilizados pelo autor na apresentação de suas ideias no texto, ou seja, a linguagem natural. Isso implica a escolha das palavras-chave do resumo do documento para representá-lo, caracterizando uma indexação por extração, (Lancaster, 2004).

A implementação da folksonomia assistida segue um passo a passo que visa oferecer suporte eficaz ao usuário no momento de definir os termos mais apropriados para as *tags* associadas ao seu trabalho em um repositório digital.

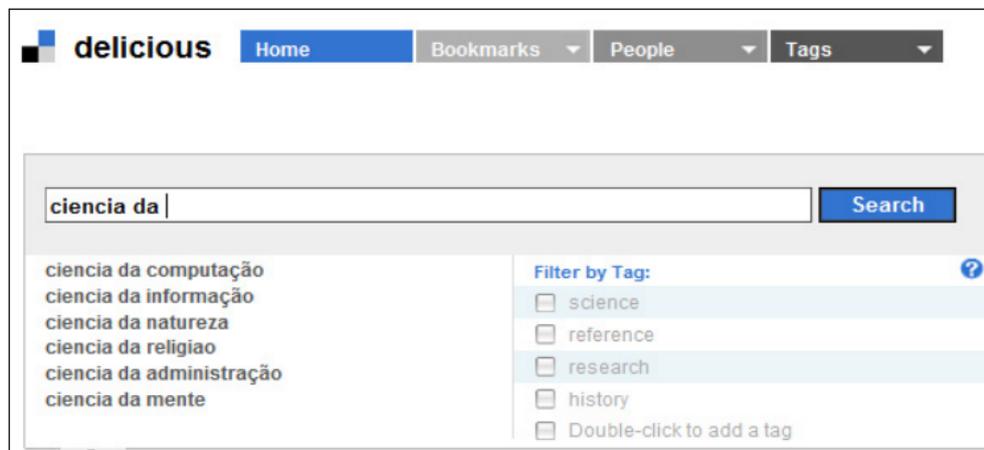
Já no âmbito da folksonomia assistida, é possível observar a atuação do profissional da informação e das ferramentas de Inteligência Artificial mediando o processo de atribuição de *tags* ou rótulos. Por meio dessa prática, podemos estabelecer padrões de entrada e filtrar as frases, evitando erros gramaticais e de escrita, reduzindo possíveis confusões terminológicas e estabelecendo relações semânticas entre etiquetas, entre outras atividades. Assim, o usuário pode escolher quais *tags* utilizar a partir das palavras previamente disponíveis SRI, preservando sua liberdade e criatividade, ao mesmo tempo em que aprimora a eficiência das *tags* na representação e recuperação da informação.

A folksonomia assistida representa um processo de suporte ao usuário durante a definição dos termos mais apropriados para as *tags* que serão associadas ao seu trabalho armazenado em um repositório digital. Este processo é composto por duas partes fundamentais. Na primeira etapa, para a implementação efetiva da folksonomia assistida, é necessário modificar a interface de comunicação do usuário com o

repositório, ou então desenvolver uma nova interface. Isso se destina à inserção de informações no campo de palavras-chave, que serve como referência para a inclusão de conteúdo nas *tags* (Santarém Segundo, 2010).

A pesquisa conduzida pelo autor ressalta que no primeiro passo do processo de folksonomia assistida, é crucial apresentar ao usuário, durante a digitação da *tag*, um conjunto de informações de maneira sistemática. Essas informações são previamente inseridas no sistema, servindo como sugestões através do recurso *autocomplete*.

A obtenção dessas sugestões é fundamentada em uma busca sintática, onde o sistema utiliza uma análise sintática para identificar termos relacionados e fornece recomendações relevantes ao usuário. Esse procedimento visa facilitar e aprimorar a escolha de *tags* pelos usuários, promovendo consistência e qualidade na categorização do conteúdo no repositório digital (Santarém Segundo, 2010),



**Figura 2:** Busca no Del.icio.us  
**Fonte:** Santarém Segundo (2010, p. 179)

Na figura acima, o autor destaca que a abordagem de inserção de dados, exemplificada no site *Delicious*, caracteriza-se pela apresentação de sugestões à medida que o usuário digita o termo a ser registrado como *tag*. Essa funcionalidade, comum em plataformas da *Web 2.0*, destaca-se pela interação rápida com o usuário sem a necessidade de recarregar a página, característica típica das interfaces ricas. A funcionalidade de "sugestão", inicialmente introduzida pelo Google em sua ferramenta de busca, é aplicada de maneira adaptada no *Delicious* e na concepção da folksonomia assistida. Nesse contexto, a ferramenta é empregada para facilitar o processo de descrição do recurso pelo usuário.

Desta forma, após o usuário realizar a descrição das *tags*, decidindo se aceita ou não as sugestões oferecidas, o sistema recebe essas informações e inicia uma segunda etapa no desenvolvimento da folksonomia assistida.

Na segunda etapa, conforme delineado por Santarém Segundo (2010), o repositório recebe os termos fornecidos pelo usuário e conduz uma pesquisa de relacionamento em relação ao conjunto interno de informações disponíveis na ferramenta. Esse processo engloba a busca por conexões dentro de uma estrutura de representação do conhecimento, seja ela um tesouro ou ontologia, viabilizando buscas hierárquicas horizontais e verticais para estabelecer relações entre os termos.

Dentro desse mesmo passo da folksonomia assistida, o sistema não se limita à recuperação de termos relacionados na estrutura de representação do conhecimento, mas também busca informações no conjunto de *tags* já inseridas no sistema, especialmente em seus relacionamentos horizontais. A busca, realizada em níveis predefinidos e com a possibilidade de parametrização, permite uma adaptação contínua conforme o repositório é expandido com novos depósitos (Santarém Segundo, 2010).

O autor propõe um processo no qual o sistema apresenta ao usuário um conjunto de termos relacionados ao que foi descrito, permitindo aceitação total, parcial ou descarte. Nessa abordagem, a decisão final fica a cargo do usuário, preservando sua liberdade na escolha de *tags* que melhor representem seu recurso digital no contexto do repositório científico. O modelo de folksonomia assistida busca consistência nas *tags*, desencorajando o uso de abreviações, plurais/singulares e palavras que possam dificultar a recuperação da informação.

O processo de gravação é efetuado quando o usuário seleciona definitivamente os termos a serem utilizados como *tags*, armazenando as informações no banco de dados. Esse processo visa conferir maior significado às *tags* em relação ao objeto depositado, contribuindo para uma representação mais precisa e contextualizada no âmbito do repositório utilizado (Santarém Segundo, 2010).

Segundo Corrêa e Santos (2018), a inclusão livre de metadados por usuários ou grupos de pessoas e a utilização de linguagem natural para a representação da informação são características fundamentais da folksonomia, uma vez que ela adota uma abordagem centrada no usuário, essa estrutura promove a participação ativa na produção e consumo de informações em ambientes colaborativos. Assim, as *tags* podem refletir “diversas perspectivas, como assunto, forma, propósito, tempo, tarefa

ou status afetivo ou reações críticas e uma miríade de outros motivos” (Viera; Garrido, 2011, p. 8).

Corrêa e Santos (2018) destacam, além disso, que, devido à flexibilidade na atribuição de termos, a folksonomia possui a característica da diversidade de etiquetas em um ou vários idiomas. Além disso, ressalta-se a existência de conceitos que podem ser integrados em notas de escopo em instrumentos de controle terminológico.

Conforme mencionado por Yu e Chen (2020), a associação e o mapeamento de folksonomias têm o potencial de enriquecer consideravelmente os vocabulários e aprimorar a qualidade dos metadados de assunto existentes. Essa prática também oferece a vantagem de mitigar a ambiguidade e a complexidade associadas à folksonomia em sistemas de informação colaborativos. Nos repositórios digitais de bibliotecas universitárias, onde deficiências comuns são observadas na folksonomia tradicional, é possível atenuar essas limitações ao adotar uma abordagem de folksonomia assistida.

Essa abordagem, com o suporte aos usuários por meio de um conjunto predefinido de termos, auxilia na seleção dos descritores mais apropriados para representar seus trabalhos na plataforma digital (Santos, 2013). Essa prática não apenas minimiza as falhas relacionadas à ambiguidade e sinonímia, mas também aprimora a eficiência na recuperação da informação.

Os sistemas híbridos de organização do conhecimento combinam a flexibilidade colaborativa da folksonomia com a estruturação formal dos instrumentos de controle terminológico. Essa abordagem visa superar as limitações individuais, resultando em representações mais precisas e contextualizadas do conhecimento. Ao integrar a contribuição dos usuários com a organização formal, esses sistemas promovem ambientes colaborativos, melhorando a eficiência na recuperação da informação e enriquecendo a experiência do usuário na exploração e compartilhamento de conhecimentos em ambientes digitais. Essa fusão representa uma evolução significativa na organização do conhecimento, oferecendo soluções mais adaptáveis e inteligentes.

Santos e Albuquerque (2021) destacam a importância crucial dos sistemas híbridos de organização do conhecimento, desempenhando um papel fundamental ao integrar as potencialidades da folksonomia com os formalismos e métodos dos instrumentos de controle terminológico. Esta abordagem híbrida busca combinar a

flexibilidade e a riqueza semântica da folksonomia com a estruturação e precisão dos instrumentos de controle terminológico, como ontologias, taxonomias e tesouros.

Os autores explicam que a folksonomia, fundamentada na colaboração dos usuários para atribuir *tags* ou etiquetas a recursos de informação, trazendo consigo uma riqueza de vocabulário e contextos diversos. No entanto, sua natureza descentralizada e não controlada pode resultar em ambiguidades, redundâncias e falta de consistência na representação da informação. Por contraste, os instrumentos de controle terminológico proporcionam uma estrutura formal e controlada para a organização do conhecimento, garantindo maior precisão e consistência nos termos utilizados (Santos e Albuquerque, 2021).

A integração da folksonomia com os instrumentos de controle terminológico em sistemas híbridos oferece benefícios significativos. Tais sistemas combinam a flexibilidade e a diversidade de vocabulário da folksonomia com a estruturação e a semântica dos instrumentos de controle terminológico, resultando em representações mais precisas, consistentes e contextualizadas do conhecimento (Santos; Albuquerque, 2021).

Santos e Albuquerque (2021) apontam ainda que os sistemas híbridos de organização do conhecimento possibilitam a criação de ambientes mais colaborativos e dinâmicos, nos quais os usuários podem contribuir com suas próprias *tags* enquanto se beneficiam da estruturação e organização proporcionadas pelos instrumentos de controle terminológico. Essa abordagem não apenas aprimora a recuperação da informação, tornando-a mais eficiente e relevante, mas também enriquece a experiência do usuário ao explorar e compartilhar conhecimentos em ambientes web.

Portanto, a relevância dos sistemas híbridos de organização do conhecimento reside na capacidade de combinar o melhor da folksonomia e dos instrumentos de controle terminológico, criando ambientes mais eficazes e inteligentes para a representação e a troca de conhecimentos em diversos contextos e domínios de informação. Essa abordagem híbrida representa uma evolução significativa na forma como lidamos com a organização e a recuperação da informação em ambientes digitais, promovendo maior interação e colaboração entre os usuários e os sistemas de informação (Santos; Albuquerque, 2021).

Segundo Santos et al. (2021), os SRI são definidos como sistemas de comunicação com a finalidade de indicar ou proporcionar acesso às informações nele registradas. Como exemplo, os catálogos online de bibliotecas têm o propósito de

indicar exclusivamente a localização dos recursos informacionais no acervo físico das unidades de informação. Nesse contexto, em contraste com esses tipos de catálogos, os repositórios digitais também se caracterizam como SRI, sendo capazes de usufruir das potencialidades oferecidas pela folksonomia.

A partir do respaldo teórico exposto nos capítulos anteriores, segue-se agora para as características metodológicas da presente pesquisa.

## 4 MÉTODO E METODOLOGIA

De acordo com Gil (2002), a metodologia desempenha um papel crucial na realização de qualquer atividade de pesquisa, pois fornece um conjunto de abordagens, técnicas e processos sistematizados para investigar um problema específico. Ela serve como um guia detalhado que orienta a condução da pesquisa e garante sua consistência e padrão.

A sistematização de abordagens, técnicas e processos para investigação ocorre a partir do enquadramento da pesquisa no campo teórico ou conceitual e no campo empírico. Assim, quanto ao objetivo geral, esta pesquisa é classificada como exploratória, na medida em que visa proporcionar maior familiaridade com os problemas de recuperação da informação em repositórios institucionais, investigando propostas de solução que possam contribuir para sua mitigação. Quanto ao delineamento, esta pesquisa é classificada como bibliográfica porque é desenvolvida a partir de fontes bibliográficas (Gil, 2002).

Os estudos analisados foram coletados na base de dados *Web of Science* durante o mês de fevereiro de 2024, utilizando a estratégia de busca avançada. A busca avançada permite a combinação de termos e a limitação da busca por áreas específicas de maior conteúdo informacional, como título, resumo, palavras-chave e descritores de assunto.

A escolha da base de dados *Web of Science* para fundamentar nossa pesquisa é respaldada por diversas razões que ressaltam a qualidade e confiabilidade dessa plataforma. Em primeiro lugar, a *Web of Science* é reconhecida por sua abrangência, englobando uma ampla variedade de fontes acadêmicas, como revistas científicas, conferências, livros e patentes. Essa diversidade proporciona uma visão holística e abrangente do campo de estudo em que estamos inseridos. Além disso, a plataforma é conhecida por seu rigoroso processo de indexação, assegurando que apenas trabalhos de alta qualidade e relevância sejam incorporados à sua base de dados. Isso confere confiabilidade aos resultados obtidos, uma vez que estamos acessando informações provenientes de fontes academicamente respeitáveis.

Como exemplo de sua proeminência, o Centro de Gestão e Estudos Estratégicos (CGEE), uma organização social vinculada ao Ministério da Ciência, Tecnologia e Inovação (MCTI) utiliza a *Web of Science* para monitorar a produção científica e tecnológica brasileira e do mundo, através de indicadores que identificam

tendências e temas emergentes, para nortear a tomada de decisões estratégicas sobre ciência e tecnologia.

Os recursos avançados de busca oferecidos pela *Web of Science* constituem outro ponto crucial. Eles nos permitem criar consultas específicas e refinadas, facilitando a identificação de estudos pertinentes ao tema de nossa pesquisa. Essa funcionalidade é essencial para a precisão e eficácia na seleção de informações relevantes. A capacidade de rastrear citações e visualizar conexões acadêmicas é um diferencial valioso da *Web of Science*. Isso nos possibilita analisar o impacto e a relevância de estudos ao longo do tempo, enriquecendo nossa compreensão do contexto em que essas pesquisas foram conduzidas.

A atualização contínua da base de dados é uma característica essencial, garantindo que tenhamos acesso às publicações mais recentes e relevantes. Isso é crucial para manter nossa pesquisa alinhada com os avanços mais recentes no campo de estudo, conferindo-lhe relevância e atualidade.

Para realizar um levantamento abrangente sobre o tema proposto, foram selecionados criteriosamente os termos-chave "*chatbots*", "*information retrieval*" e "*natural language processing*". A expressão de busca elaborada para nortear a pesquisa foi cuidadosamente delineada como  $((ALL=(Chatbot^*)) AND ALL=(information\ retrieval)) OR ALL=(natural\ language\ processing)) AND ALL=(information\ retrieval)$ . Essa estrutura de busca foi desenhada para identificar trabalhos que abrangessem tanto a interseção quanto as áreas individuais de *chatbots*, recuperação de informação e PLN. Após a execução da busca na base de dados, foram recuperados 4.629 documentos.

Ao aplicar os critérios de inclusão e exclusão, a pesquisa buscou refinar a seleção de documentos para garantir uma análise mais específica e relevante. Inicialmente, o filtro temporal foi estabelecido, abrangendo artigos publicados no período de 2019 a 2023. Essa escolha visa concentrar a revisão nas contribuições mais recentes, considerando o dinamismo e a evolução constante das tecnologias associadas aos *chatbots*, PLN e recuperação de informação. Dessa forma, foram identificados e selecionados 1.953 documentos dentro desse intervalo temporal.

Em seguida, foi aplicado o filtro adicional relacionado ao tipo de documento, restringindo a análise apenas a artigos. Essa decisão foi tomada para assegurar que a pesquisa se baseasse em estudos acadêmicos aprofundados e revisados por pares,

proporcionando uma fundação sólida para a revisão da literatura. O resultado dessa filtragem resultou em 957 documentos.

A delimitação do idioma dos documentos também foi criteriosamente considerada, optando-se por incluir apenas aqueles em inglês e português. Essa escolha é justificada pelo objetivo geral da pesquisa de abranger tanto a produção acadêmica internacional quanto a nacional, visando uma visão mais abrangente do panorama de pesquisa sobre o tema. Com essa restrição, foram mantidos 548 documentos relevantes e alinhados ao escopo da investigação.

A determinação da área de conhecimento foi uma etapa essencial para garantir a pertinência e foco na pesquisa. Ao selecionar documentos associados aos termos-chave específicos, como PLN, *chatbots* e recuperação de informação, a pesquisa buscou abranger campos específicos que estão intrinsecamente relacionados ao escopo da investigação. Essa abordagem foi respaldada pela escolha criteriosa de áreas de busca, incluindo Sistemas de Informação em Ciência da Computação, Inteligência Artificial em Ciência da Computação e Ciência da Informação e Biblioteconomia. Essas áreas foram escolhidas devido à sua estreita relação com os tópicos de interesse, assegurando que os documentos selecionados estejam profundamente imersos em contextos relevantes e alinhados aos objetivos delineados para a pesquisa.

Ao eleger a área de busca específica de Ciência da Informação e Biblioteconomia, foi possível recuperar um total de 71 documentos. Contudo, a delimitação dessa área pode, por vezes, resultar em uma limitação na abrangência dos estudos disponíveis, uma vez que a temática dos *chatbots*, por sua natureza técnica, tem uma presença mais robusta na Ciência da Computação. Diante desse cenário, a estratégia de incorporar áreas adicionais, tais como Sistemas de Informação em Ciência da Computação e Inteligência Artificial em Ciência da Computação, foi adotada com o intuito de ampliar o escopo da pesquisa.

É relevante reconhecer que, na CI, a abordagem direta aos *chatbots* pode ser limitada, dado o enfoque mais técnico inerente a essa temática, que está profundamente associada à CC. No entanto, é crucial ressaltar que essa faceta técnica carrega consigo uma dimensão conceitual de grande interesse para a CI. A incorporação de *chatbots* não apenas representa uma expressão tecnológica, mas também introduz conceitos fundamentais, como interação humano-máquina, eficiência na recuperação de informação e disseminação de conhecimento.

Nesse contexto, a compreensão e a inclusão gradual do universo dos *chatbots* podem ser percebidas como uma extensão natural do escopo da CI. A compreensão dos fundamentos técnicos, mesmo não sendo o foco central da área, capacita os profissionais da CI a dialogar de forma mais eficaz com a vertente tecnológica, possibilitando, principalmente, a exploração das potencialidades conceituais oferecidas por essa tecnologia em ascensão.

Não foram encontrados documentos duplicados durante a recuperação. A partir dessas áreas de busca, foram identificados e mantidos 546 documentos, destacando assim a relevância e o impacto positivo dessa abordagem expandida. Esta inclusão adicional não apenas ampliou a quantidade de material disponível para análise, mas também enriqueceu a diversidade de perspectivas, contribuindo para uma visão mais abrangente e informada sobre a interseção entre *chatbots* e CI. Assim, essa escolha estratégica não só justifica, mas também fortalece a pesquisa, garantindo uma abordagem mais holística e aprofundada sobre essa temática multidisciplinar.

Foram selecionados para análise apenas os artigos que possibilitavam o acesso ao texto completo, totalizando 263 documentos. Essa escolha teve como objetivo assegurar uma avaliação abrangente e detalhada, considerando todo o escopo das contribuições acadêmicas disponíveis. A delimitação dessa seleção contribui para a confiabilidade e relevância dos resultados obtidos, permitindo uma análise mais precisa e substancial das informações contidas nos artigos científicos em questão.

Considerando o limite temporal estabelecido pela pesquisa, o total de artigos divididos por ano de publicação, foram os seguintes: 40 artigos em 2019, 47 artigos em 2020, 39 artigos em 2021, 70 artigos em 2022 e 67 documentos em 2023. Destaca-se o crescimento da publicação na temática nos últimos dois anos, em que o número de artigos publicados quase dobrou.

A partir desse conjunto de documentos recuperados, aplicou-se a análise de conteúdo, conceituada como “[...]um conjunto de técnicas de análise das comunicações” (Bardin, 2016, p. 37). A análise de conteúdo é composta por três fases: a pré-análise; a exploração do material e o tratamento dos resultados obtidos, que inclui processos de inferência e de interpretação.

A **pré-análise** é uma etapa crucial para definir o foco e os objetivos da análise, envolvendo a leitura “flutuante” de artigos para a organização dos dados, com o propósito de formar o conjunto de informações que será utilizado como base para a

pesquisa. A abordagem inicial de leitura é "flutuante", pois envolve o processo de se familiarizar com os documentos sob análise e absorver o conteúdo de forma a permitir que as impressões e direções se manifestem. Com o tempo, essa leitura gradualmente se aprofunda, guiada por hipóteses emergentes, pela aplicação de teorias relevantes ao material e pelo uso potencial de técnicas previamente empregadas em materiais semelhantes (Bardin, 2016).

Nesta pesquisa, a pré-análise concentrou-se na revisão de estudos anteriores (seção 1.1), com o objetivo de identificar lacunas de pesquisa, formular a pergunta de partida e as perguntas secundárias e definir o objetivo geral. Portanto, os documentos recuperados na revisão de literatura foram utilizados como fonte de informação teórica, permitindo a compreensão do estado da arte dos estudos desenvolvidos na área, das teorias, dos conceitos e das abordagens existentes no campo da CI.

Durante a etapa de **exploração do material**, é definido o conjunto de materiais que será explorado em maior detalhe, visando identificar as unidades de registro e unidades de contexto. Uma unidade de registro é considerada como uma unidade a ser codificada, abrangendo elementos como temas, palavras ou frases (Bardin, 2016).

Nesta pesquisa, a exploração do material foi norteada pelos três objetivos específicos, que procuram dar respostas às três questões secundárias formuladas na introdução.

Com o primeiro objetivo específico "identificar e mapear os principais problemas que têm guiado as pesquisas sobre PLN e *chatbots* na área de recuperação de informação", almejamos identificar e compreender os desafios recorrentes e as lacunas existentes no cenário atual de RI, elucidando questões que demandam atenção especial e investigação mais aprofundada. A análise desses problemas se revela crucial para fundamentar de maneira sólida as estratégias e soluções propostas ao longo desta pesquisa, contribuindo para a construção de um conhecimento sólido sobre o panorama atual da recuperação de informação em repositórios institucionais. Essa abordagem investigativa não apenas delineará o estado atual dos desafios enfrentados, mas também proporcionou uma base robusta para a subsequente análise das soluções propostas, constituindo, assim, um elemento essencial para investigações futuras no campo.

Com o segundo objetivo específico "analisar as soluções propostas em cada estudo específico, destacando as principais abordagens sugeridas para aprimorar a recuperação de informação" procuramos não apenas identificar, mas compreender as

estratégias e os métodos propostos pela literatura analisada. Ao extrair e sintetizar as abordagens mais proeminentes, conseguimos identificar padrões e tendências que têm sido destacados como eficazes na melhoria do processo de recuperação de informação. A análise minuciosa destas soluções propostas serviu como alicerce para a posterior caracterização das aplicações específicas, fornecendo *insights* valiosos sobre como essas abordagens podem ser aplicadas ou adaptadas em contextos específicos de repositórios institucionais.

Com o terceiro objetivo específico “caracterizar as aplicações específicas de cada estudo no sentido de aprimorar a recuperação de informação em repositórios institucionais, destacando as contribuições práticas e potenciais benefícios para o processo de busca e recuperação”, buscamos entender a diversidade de abordagens adotadas e identificar as contribuições práticas e os potenciais benefícios inerentes a cada estudo. A análise cuidadosa dessas aplicações específicas será fundamental para discernir a adaptabilidade e a eficácia dessas propostas em contextos práticos de repositórios institucionais. Além disso, a identificação dessas contribuições práticas proporcionou uma compreensão mais holística das implicações e do impacto potencial de cada estudo na otimização do processo de busca e recuperação de dados. Ao final desta etapa, será possível extrair *insights* valiosos que contribuirão para o desenvolvimento de recomendações sólidas e aplicáveis, orientando a efetiva implementação dessas abordagens em ambientes reais de recuperação de informação em repositórios na CI.

Para alcançar o objetivo da pesquisa, utilizou-se a técnica de análise categorial temática, um componente do método de análise de conteúdo, conforme delineado por Bardin (2016). A análise categorial foi utilizada para o tratamento dos resultados obtidos, incluindo processos de inferência e de interpretação.

As categorias (Quadro 1) foram estabelecidas com base no espaço quadridimensional em que a relevância pode ser analisada em sistemas de recuperação (Mizzaro, 1998). Segundo o autor, em todos os processos de recuperação de informação, o objetivo é encontrar documentos relevantes, portanto, a relevância é um dos conceitos centrais de recuperação da informação. Assim, a relevância é um ponto a ser considerado em um espaço de quatro dimensões, formado por: recursos de informação, representação do problema do usuário, tempo e componentes.

**Quadro 1:** Categorias de análise temática dos estudos

<b>Categoria</b>	<b>Subcategoria</b>	<b>Observações</b>
Recursos de informação	Documento	Nesta subcategoria foram considerados problemas e propostas relacionadas a tipos de documentos e seus suportes (texto, áudio, vídeo, imagem); a unidades de documento e sua complexidade para efeitos de representação e recuperação (por exemplo, livro e suas seções ou e-mails com anexos); a processamento, preservação e acesso, etc.
	Representação do documento	Nesta subcategoria foram considerados problemas e propostas relacionadas a indexação, indexação automática, operações com documentos para ranqueamento e recuperação ( <i>tokenização</i> , eliminação de <i>stop-words</i> , normalização, <i>stemming</i> e lematização), metadado assunto, etc.
	Informação	Nesta subcategoria foram considerados problemas e propostas relacionadas a dados, informação e conhecimento (por exemplo, gestão de dados, ciclo de vida de dados, leis de proteção de dados, gestão da informação e do conhecimento em organizações, uso estratégico de informação, tecnologias de gestão de dados e informações, etc.).
Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)	Relevance Feedback, expansão da consulta e serviços de recomendação	Nesta subcategoria foram considerados problemas e propostas relacionadas aos diferentes tipos de ambiguidade, principalmente a sinonímia (por exemplo, <i>relevance feedback</i> com ajuste da <i>query</i> a partir dos documentos recuperados ou uso de tesouros para expansão de consultas do usuário); às medidas de avaliação de sistemas de recuperação, como precisão e revocação.
	Interação e IA	Nesta subcategoria foram considerados problemas e propostas relacionadas à projeção de sistemas de recuperação da informação baseados em tarefas, que permitem uma maior interação do usuário com o sistema, como <i>chatbots</i> ; à modelagem de tarefas e contextos em recuperação da informação, com aprendizado de máquinas; ao uso do texto completo e da linguagem natural; à estruturas de IA conversacionais, que podem ser utilizadas ou integradas aos Assistentes, aproveitando modelos de linguagem para entender e auxiliar os usuários.

**Fonte:** Adaptado com base em Mizzaro (1998); Manning, Raghavan, Schutze (2008) e Baeza-Yates; Ribeiro-Neto (2013); Tseng et al. (2020).

Na **primeira dimensão** - recursos de informação, a relevância é uma relação de três entidades: a) documento – entidade física que o usuário deseja recuperar; b) substituto – entidade que representa o documento e facilita a busca do usuário, por exemplo, título, resumo ou palavras-chave e c) informação – conteúdo do documento ou entidade não física, que é apropriado pelo usuário ao ler um documento. Logo, a

relevância na dimensão recursos de informação considera o conjunto formado por documento, representação do documento e informação do documento. A representação do documento é menos informativa do que o próprio documento e o documento, por sua vez, é menos informativa em relação ao seu conteúdo (Mizzaro, 1998).

A maioria das abordagens sobre organização da informação e do conhecimento na CI recai sobre a dimensão recursos de informação. A análise documentária, enquanto atividade metodológica de análise, síntese e representação da informação, tem como propósito construir linguagens para indexação, armazenamento e recuperação da informação. Os tesouros, as taxonomias, os sistemas de classificação e outras linguagens documentárias reúnem a maioria das características de relevância na dimensão recursos de informação.

A dimensão recursos de informação foi considerada primeira categoria de análise, formada pelas subcategorias documento, representação do documento e informação. Assim, a identificação dos principais problemas que têm guiado as pesquisas sobre PLN e *chatbots* na área de recuperação de informação e a caracterização da sua aplicação em repositórios institucionais (objetivos específicos da pesquisa) foi feita inicialmente nesta dimensão ou categoria.

Na **segunda dimensão** - representação do problema do usuário, a relevância decorre da interação entre o usuário e o sistema. Nesse contexto, é importante considerar que o usuário tem uma necessidade real de informação, com base na qual constrói a necessidade percebida de informação, uma representação mental do problema que motiva o processo de busca. Em seguida, o usuário representa a sua necessidade percebida de informação através de uma solicitação, que é representada por uma *string* de busca ou consulta. Nesta dimensão, a relevância é uma relação de quatro entidades: necessidade real de informação, necessidade percebida de informação, solicitação e *query*. A *string* de busca ou *query* é menos informativa do que a solicitação. A solicitação é menos informativa do que a necessidade percebida de informação que, por sua vez, é menos informativa em comparação com a necessidade real de informação (Mizzaro, 1998).

Alguns problemas relacionados à dimensão representação do problema do usuário são discutidos nas áreas de comportamento informacional e competência em informação na CI e não fazem parte da nossa análise. Neste estudo consideramos, por exemplo, a importância de modelos de recuperação baseados em tarefas, como

*chatbots*, que permitem uma maior interação entre o usuário e o sistema e o uso da linguagem natural. Neste contexto, as subcategorias *relevance feedback*, expansão da consulta e serviços de recomendação e interação e IA (Tseng et al. 2020) foram acrescentadas à categoria representação do problema do usuário para analisar, principalmente, estudos que discutem problemas ou propõem soluções relacionados a PLN e interação em sistemas de recuperação como *chatbots*.

As **dimensões tempo e componentes** não foram consideradas na pesquisa devido à natureza específica do estudo, que se concentra em aspectos técnicos e estruturais, e não no comportamento dos usuários. Como o objetivo não envolve o estudo de usuários, não foi possível analisar essas dimensões. A relevância pode ser influenciada por diversos fatores, incluindo o processo de busca e o estado cognitivo de cada usuário, os quais não seriam adequadamente avaliados no âmbito desta pesquisa. A dimensão – tempo considera que a relevância varia com o tempo. Por exemplo, a necessidade percebida pode ser modificada com o tempo e influenciar a solicitação e *string* de busca. A relevância pode ser influenciada, entre outros fatores, pelo processo de busca e pelo estado cognitivo de cada usuário. A dimensão – componentes considera que a relevância é influenciada pelo tópico – área de interesse do usuário, pela tarefa – atividade que o usuário pretende executar, pelo contexto em que o usuário efetua a busca e avaliação dos resultados, pela novidade, entre outros aspectos.

Por último, Bardin (2016) considera que a etapa final de **tratamento dos resultados obtidos e interpretados** compreende o processamento dos resultados alcançados e a subsequente interpretação. Envolve a transformação dos dados brutos em resultados compreensíveis e significativos. Essa fase é onde os dados passam por codificação, categorização, análise e interpretação mais aprofundada. Os resultados desta pesquisa foram submetidos a análises estatísticas que conferem significado e validade, facilitando a inferência e a compreensão dos dados, conforme Apêndice A.

## 5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Os critérios foram aplicados criteriosamente para garantir a precisão e relevância na seleção de informações pertinentes ao tema proposto, consolidando assim a pesquisa em uma abordagem metodológica robusta e alinhada aos objetivos delineados.

Após a análise minuciosa dos artigos e coleta de dados, abrangendo títulos, problemas, objetivos e resultados, possibilitou-se uma investigação sobre a aplicabilidade desses estudos em repositórios institucionais. A partir dessa avaliação, foram identificadas as categorias e subcategorias às quais cada estudo está associado, conforme apresentado na tabela 3.

A implementação de *chatbots* em repositórios institucionais oferece uma série de benefícios significativos para melhorar a experiência do usuário e otimizar a recuperação de informações. Essas aplicações abrangem diversas áreas de interação e funcionalidade, proporcionando uma gama de serviços valiosos aos usuários. Um dos principais usos dos *chatbots* em repositórios institucionais é fornecer assistência instantânea ao usuário. Esses assistentes virtuais podem responder a perguntas comuns, fornecer orientações sobre como utilizar o repositório, auxiliar na navegação e ajudar os usuários a encontrar os recursos que estão procurando. Isso contribui para uma experiência mais eficiente e satisfatória para os usuários, facilitando o acesso às informações desejadas.

Além disso, os *chatbots* podem oferecer uma experiência de busca mais personalizada aos usuários, compreendendo suas necessidades individuais e fornecendo resultados mais relevantes com base em suas consultas e preferências. Isso ajuda a otimizar o processo de busca e aumentar a satisfação do usuário, garantindo que eles encontrem rapidamente o que estão procurando.

Outra aplicação importante dos *chatbots* em repositórios institucionais é o envio de notificações e alertas aos usuários. Esses assistentes virtuais podem informar os usuários sobre novos documentos adicionados ao repositório, atualizações relevantes em áreas de interesse específicas ou eventos relacionados à pesquisa institucional. Isso mantém os usuários atualizados e engajados com o conteúdo do repositório.

Os *chatbots* também podem tornar a interação com o repositório mais dinâmica e envolvente, permitindo que os usuários realizem ações diretamente por meio do chat, como fazer *download* de documentos, salvar pesquisas, compartilhar conteúdo

e até mesmo interagir com outros usuários. Isso cria uma experiência mais interativa e colaborativa para os usuários, promovendo o engajamento e a participação ativa no repositório.

Por fim, os chatbots também podem ser utilizados para coletar *feedback* dos usuários sobre a usabilidade do repositório, a eficácia das ferramentas de busca, a relevância dos resultados e outras áreas de melhoria. Essas informações são valiosas para aprimorar continuamente o repositório e garantir que ele atenda às necessidades e expectativas dos usuários de forma eficaz.

Contudo essa metodologia possibilitou uma análise mais detalhada dos desafios enfrentados pelos repositórios institucionais e das soluções potenciais derivadas dos estudos examinados. Ao categorizar os estudos de acordo com suas áreas temáticas e objetivos, foi viável estabelecer conexões significativas entre a pesquisa acadêmica e sua aplicação prática nos repositórios institucionais.

O papel fundamental do PLN e dos *chatbots* na facilitação da recuperação de informações nos repositórios institucionais é inegável. Ao compreender e interpretar consultas em linguagem natural, essas tecnologias simplificam a interação entre usuários e repositórios, tornando o processo de busca por informações mais intuitivo e eficiente.

A integração do PLN e dos *chatbots* nos repositórios institucionais não apenas proporciona benefícios tangíveis, mas também representa um avanço considerável na recuperação e uso da informação dentro desses ambientes. É essencial examinar mais detalhadamente como cada um desses aspectos contribui para aprimorar a experiência do usuário. Uma das melhorias mais significativas é a compreensão aprimorada das consultas. Isso se refere à capacidade aprimorada dos sistemas de recuperação de informações em entender e interpretar consultas formuladas em linguagem natural pelos usuários.

Na Tabela 3, serão apresentados alguns estudos analisados nos resultados, destacando os problemas de pesquisa identificados, suas possíveis soluções e como essas soluções foram aplicadas em repositórios institucionais. Esses estudos ajudarão a fornecer *insights* valiosos sobre o impacto dos *chatbots* na melhoria da experiência do usuário e na otimização da recuperação de informações em contextos institucionais.

**Tabela 3:** Alguns estudos analisados nos resultados

<b>Ano</b>	<b>Estudo</b>	<b>Problema de pesquisa</b>	<b>Possível Solução</b>	<b>Aplicação em Repositório Institucional</b>
2019	The state-of-the-art in expert recommendation systems	<p>O principal problema abordado no estudo é a análise e avaliação dos sistemas de recomendação de especialistas, incluindo suas técnicas, conjuntos de dados, métodos de avaliação e desafios enfrentados por esses sistemas.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	Uma possível solução proposta pelo estudo é fornecer uma visão abrangente e crítica do estado-da-arte desses sistemas, destacando suas vantagens e desvantagens, explorando métricas de avaliação e delineando direções futuras de pesquisa para superar os desafios existentes na área.	Essas abordagens podem ser aplicadas em repositórios institucionais para aprimorar a identificação e recomendação de especialistas em áreas específicas do conhecimento. Ao implementar as técnicas discutidas, os repositórios podem melhorar a descoberta de especialistas, facilitando a colaboração acadêmica e promovendo a troca de conhecimento entre pesquisadores. A utilização de métricas de avaliação adequadas também é crucial para medir a eficácia desses sistemas, contribuindo para a melhoria contínua de suas funcionalidades.
2019	Accuracy evaluation of methods and techniques in Web-based question answering systems: a survey	<p>O principal problema abordado neste estudo é a avaliação da precisão de métodos e técnicas em sistemas de resposta a perguntas baseados na Web. Os pesquisadores procuram identificar quais métodos e técnicas são mais eficazes na melhoria da precisão das respostas fornecidas por esses sistemas. A questão central é</p>	Uma possível solução proposta pelo estudo é realizar uma avaliação abrangente e comparativa dos métodos e técnicas utilizados em sistemas de resposta a perguntas na Web. Os pesquisadores buscam identificar as combinações de métodos que resultam em maior precisão e eficácia nesses sistemas. Além	No contexto de repositórios institucionais, os resultados deste estudo podem ser aplicados para melhorar a precisão das respostas fornecidas pelos sistemas de questionamento e resposta internos. Ao selecionar e implementar métodos e técnicas identificados como os mais eficazes, os repositórios podem

		<p>como avaliar e comparar a precisão de diferentes abordagens em sistemas de resposta a perguntas, a fim de identificar as mais eficazes e informar o desenvolvimento futuro dessas tecnologias.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>disso, eles têm como objetivo desenvolver um sistema de resposta a perguntas baseado nos métodos e técnicas identificados como os mais eficazes, a fim de demonstrar a diferença entre a seleção dos melhores métodos e técnicas em comparação com outros métodos disponíveis.</p>	<p>oferecer aos usuários respostas mais precisas e eficientes. Isso não só aumenta a utilidade dos repositórios institucionais como fontes confiáveis de informação, mas também contribui para uma melhor experiência do usuário ao buscar informações específicas dentro desses repositórios. Portanto, a aplicação dos resultados do estudo pode levar a uma melhoria significativa na qualidade do serviço oferecido pelos repositórios institucionais.</p>
2019	Automated feedback generation for formal manufacturing rule extraction	<p>O principal problema abordado neste estudo é a complexidade da extração automatizada e precisa de regras de fabricação a partir de textos não estruturados. A validação cuidadosa do texto é crucial devido a possíveis informações irrelevantes ou expressões semanticamente inválidas que podem prejudicar a extração correta das regras desejadas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p>	<p>Uma possível solução proposta pelo estudo é o desenvolvimento de um método que ofereça feedback automatizado para validar e aprimorar textos de entrada não estruturados. Esse método permite aos usuários modificar o texto de entrada sem a necessidade de aprender linguagens estruturadas formais, superando as limitações associadas a textos que demandam validação.</p>	<p>No contexto de repositórios institucionais, a aplicação desse estudo pode auxiliar na validação e melhoria de textos não estruturados relacionados à fabricação. Ao utilizar o método proposto, os repositórios podem oferecer aos usuários uma ferramenta automatizada que fornece feedback relevante para aprimorar a qualidade e precisão dos textos armazenados. Isso facilita a extração correta das regras de fabricação, tornando a informação mais acessível e útil para os usuários que consultam o repositório em busca de conhecimento sobre processos de fabricação. Assim, o estudo</p>

		<b>Subcategoria:</b> Relevance Feedback		contribui para aprimorar a eficiência na gestão de informações relacionadas à fabricação em repositórios institucionais.
2020	A New Hybrid Improved Method for Measuring Concept Semantic Similarity in WordNet	<p>O principal problema abordado neste estudo é a necessidade de aprimorar a medição da similaridade semântica entre conceitos em WordNet, especialmente considerando a distinção de similaridades entre nós de um mesmo nível.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Informação</p>	Uma possível solução proposta pelo estudo é a introdução de um novo método híbrido aprimorado para medir a similaridade semântica entre conceitos, superando as limitações dos métodos anteriores, como o método de Wu e Palmer.	No contexto de repositórios institucionais, a aplicação desse estudo pode melhorar significativamente a precisão na recuperação de informações e no gerenciamento do conhecimento. Ao implementar o novo método proposto, os repositórios podem facilitar uma análise mais aprofundada da similaridade semântica entre os conceitos armazenados, permitindo uma organização mais eficiente e uma recuperação mais precisa de informações relevantes. Isso contribui para uma melhor utilização e aproveitamento dos recursos disponíveis nos repositórios institucionais, beneficiando pesquisadores, profissionais e estudantes que dependem dessas informações para suas atividades acadêmicas e de pesquisa.
2021	Hierarchical Matching Network for Multi-Turn Response Selection	O principal problema abordado neste estudo é a seleção apropriada de respostas em chatbots baseados em recuperação, especialmente em	Uma possível solução proposta pelo estudo é o desenvolvimento de um modelo hierárquico de correspondência, denominado Hierarchical	No contexto de repositórios institucionais, a aplicação desse estudo pode melhorar significativamente a recuperação de informações e a

	in Retrieval-Based Chatbots	<p>cenários de diálogos multi-turnos. Os métodos existentes enfrentam desafios em capturar informações de correspondência local e global entre respostas e contextos de conversação em diferentes níveis, levando a uma seleção de respostas menos eficaz.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>Matching Network (HMN), que pode capturar informações de correspondência em diferentes níveis - nível de palavra e nível de enunciado - para melhorar a seleção de respostas em diálogos multi-turnos. Isso permite uma interação mais eficaz entre o <i>chatbot</i> e o usuário, garantindo respostas mais relevantes e contextuais.</p>	<p>interação com os usuários. Ao implementar o modelo HMN, os repositórios podem facilitar uma análise mais aprofundada da similaridade semântica entre os conceitos armazenados, permitindo uma organização mais eficiente e uma recuperação mais precisa de informações relevantes. Isso contribui para uma melhor utilização e aproveitamento dos recursos disponíveis nos repositórios institucionais, beneficiando pesquisadores, profissionais e estudantes que dependem dessas informações para suas atividades acadêmicas e de pesquisa.</p>
2022	Natural Language Semantic Answering Applied to Medicinal Plant and Coronavirus	<p>O principal problema abordado no estudo é a necessidade de criar um sistema eficaz de resposta a perguntas em linguagem natural sobre plantas medicinais e coronavírus. Esse sistema deve ser capaz de interpretar perguntas em espanhol e fornecer respostas precisas e não triviais com base em um conhecimento estruturado e dedutivo.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de</p>	<p>A possível solução apresentada é o desenvolvimento do Sistema PryRe, que integra PLN, ontologias e métodos de recuperação de informações. O PryRe é projetado para interpretar perguntas em espanhol, transformá-las em triplas semânticas e realizar análises em uma ontologia específica do domínio de interesse. Ao utilizar esse sistema, é possível fornecer respostas mais precisas e relevantes às consultas dos</p>	<p>O Sistema PryRe pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações. Ao integrar essa abordagem nos repositórios, é possível oferecer respostas mais precisas e relevantes às consultas dos usuários, facilitando a busca por informações específicas dentro do contexto institucional. Isso pode aprimorar a eficiência e a usabilidade dos repositórios, tornando a busca por</p>

		informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA	usuários sobre plantas medicinais e coronavírus.	informações mais intuitiva e eficaz para os usuários.
2022	On the Effectiveness of Pre-Trained Language Models for Legal Natural Language Processing: An Empirical Study	O principal problema abordado no estudo é a avaliação da eficácia de modelos de linguagem pré-treinados em tarefas de PLN natural no contexto jurídico, especialmente em comparação com métodos tradicionais. O estudo visa entender como esses modelos se saem em relação a abordagens convencionais e identificar suas vantagens e limitações em aplicações jurídicas.  <b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação	A possível solução apresentada é o uso de modelos de linguagem pré-treinados específicos para o domínio jurídico em tarefas de PLN. O estudo sugere que esses modelos podem oferecer um desempenho superior em comparação com abordagens tradicionais, proporcionando uma forma mais eficaz de lidar com informações legais e facilitando tarefas como classificação de documentos, resposta a perguntas e resumos automáticos.	A aplicação desses modelos em repositórios institucionais pode melhorar significativamente a recuperação de informações legais. Ao integrar esses modelos nos sistemas de indexação e recuperação de dados jurídicos, os repositórios podem oferecer aos usuários acesso mais rápido e preciso a informações relevantes. Isso não apenas aumenta a eficiência e a produtividade dos profissionais jurídicos, mas também melhora a prestação de serviços jurídicos aos clientes e à sociedade em geral.
2022	Natural Language Processing Applied to Forensics Information	O principal problema abordado no estudo é como aplicar técnicas avançadas de PLN, como reconhecimento de	A possível solução proposta é o desenvolvimento de modelos de PLN aprimorados por arquiteturas de redes neurais e	Essas técnicas podem ser aplicadas em repositórios institucionais para otimizar a gestão de informações, acelerar

	<p>Extraction With Transformers and Graph Visualization</p>	<p>entidades nomeadas e extração de relações, para a extração de informações em investigações forenses digitais. Esse problema envolve lidar com a complexidade e a heterogeneidade dos dados digitais encontrados nessas investigações, visando automatizar a análise e interpretação eficazes dessas informações.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>transformadores. Esses modelos são projetados para automatizar a análise de extração de informações em diferentes contextos, incluindo forense digital, investigação digital, análise de big data e inteligência empresarial. A combinação dessas técnicas pode melhorar significativamente a eficiência e a precisão na extração de informações, permitindo uma análise mais rápida e precisa dos dados digitais.</p>	<p>processos de análise e fortalecer a segurança cibernética. Ao integrar modelos aprimorados de PLN em sistemas de indexação e recuperação de dados, os repositórios podem automatizar a análise de grandes volumes de dados heterogêneos, fornecendo aos usuários acesso mais rápido e preciso a informações relevantes. Além disso, a utilização de grafos interativos para visualização dos dados extraídos facilita a compreensão e interpretação dos resultados, promovendo uma interação mais eficaz com o repositório institucional. Essa aplicação é essencial para instituições que lidam com grandes volumes de dados digitais e buscam fortalecer sua capacidade de análise e proteção de informações sensíveis.</p>
2022	<p>Legal Information Retrieval systems: State-of-the-art and open issues</p>	<p>O principal problema abordado no estudo é a identificação dos desafios e questões em aberto relacionados aos sistemas de recuperação de informações legais. Isso inclui aspectos como classificação de documentos legais, extração de informações relevantes,</p>	<p>A possível solução proposta é fornecer uma visão abrangente do estado atual da recuperação de informações legais, destacando tecnologias e abordagens recentes utilizadas nesse campo. Além disso, o estudo visa identificar e discutir questões em aberto e desafios</p>	<p>Esses insights podem ser aplicados em repositórios institucionais para melhorar a recuperação de informações legais armazenadas nessas plataformas. O uso de tecnologias como inteligência artificial, processamento de linguagem natural e ontologias</p>

		<p>compreensão semântica e robustez dos sistemas de recuperação de informações legais. Os autores buscam compreender as necessidades reais dos profissionais do direito ao lidar com a busca e extração de informações em documentos legais.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>enfrentados na recuperação de informações legais, com o objetivo de orientar futuras pesquisas e desenvolvimentos. Os autores oferecem insights valiosos para profissionais do direito, pesquisadores e desenvolvedores interessados em aprimorar os sistemas de recuperação de informações legais.</p>	<p>pode aumentar a eficiência na busca e extração de informações relevantes em documentos legais. Isso beneficia instituições ao facilitar o acesso a dados jurídicos importantes, agilizar a pesquisa e análise de casos, e garantir uma recuperação mais precisa e contextualizada de informações legais para profissionais do direito que utilizam os repositórios institucionais.</p>
2022	Dealing with textual noise for robust and effective BERT re-ranking	<p>O principal problema abordado no estudo é o impacto do ruído textual na eficácia e robustez dos modelos de reclassificação BERT em tarefas de recuperação de informações. Os autores investigam como perturbações textuais imperceptíveis podem afetar a qualidade da classificação dos modelos BERT, especialmente em cenários do mundo real onde o texto ruidoso é comum.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de</p>	<p>A possível solução proposta pelo estudo é a introdução do modelo De-Ranker, projetado para lidar de forma mais eficaz com o ruído textual e melhorar o desempenho em comparação com os modelos BERT convencionais. O De-Ranker é treinado para minimizar a distância entre o texto ruidoso e sua versão limpa original, permitindo uma classificação mais precisa e robusta mesmo em presença de ruído textual.</p>	<p>Esse estudo pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações legais armazenadas nessas plataformas. Ao considerar a presença comum de ruído textual em documentos acadêmicos e científicos, a implementação de modelos como o De-Ranker pode melhorar a precisão e relevância dos resultados de busca em repositórios institucionais. Isso beneficia os usuários ao garantir uma recuperação de informações</p>

		<p>informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>		<p>mais eficaz e precisa, facilitando o acesso a conteúdo relevante nos repositórios institucionais.</p>
2022	<p>Building a Technology Recommender System Using Web Crawling and Natural Language Processing Technology</p>	<p>O principal problema abordado neste estudo é a dificuldade enfrentada pelos especialistas em determinado domínio em acompanhar as tecnologias recentes de forma eficiente e eficaz. Além disso, uma vez que a informação sobre uma tecnologia é encontrada, é desafiador e demorado compreendê-la e avaliá-la de forma aproximada para decidir se merece uma investigação mais detalhada.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>A possível solução proposta é a automatização da tarefa de acompanhar as tecnologias recentes, utilizando uma combinação de web crawling, web scraping e sumarização automática de texto com tecnologia de processamento de linguagem natural. O objetivo é desenvolver um protótipo de sistema de recomendação de tecnologia, capaz de fornecer recomendações sobre tecnologias em tendência para especialistas em determinados propósitos de aplicação.</p>	<p>Esse estudo pode ser aplicado em repositórios institucionais para automatizar a busca e sumarização de informações relevantes sobre tecnologia, facilitando a descoberta de novas tecnologias e auxiliando na tomada de decisões. A integração da folksonomia poderia enriquecer o processo de recomendação, melhorando a precisão das recomendações feitas pelo sistema. Além disso, a implementação de chatbots poderia facilitar ainda mais o acesso às recomendações de tecnologia, melhorando a experiência de busca e acesso ao conhecimento sobre tecnologia em repositórios institucionais.</p>
2022	<p>A disponibilização de vocabulário controlado aos usuários para a recuperação da informação</p>	<p>O principal problema abordado neste estudo é a forma como os vocabulários controlados estão sendo apresentados nos catálogos online das bibliotecas universitárias brasileiras e o</p>	<p>Uma possível solução proposta é melhorar a exposição e destaque dos vocabulários controlados nos catálogos online, tornando-os mais visíveis e acessíveis aos</p>	<p>Em um repositório institucional, essa abordagem pode ser aplicada implementando vocabulários controlados no sistema de organização da informação. Os vocabulários</p>

		<p>impacto dessa disponibilização na recuperação da informação pelos usuários. Muitas vezes, esses recursos não são apresentados de forma perceptível para os usuários comuns, o que dificulta sua utilização e prejudica a eficácia e precisão das pesquisas realizadas nos catálogos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>usuários. Além disso, recomenda-se oferecer tutoriais e treinamentos para os usuários, visando aprimorar suas habilidades de busca e recuperação de informações. Essas práticas podem contribuir para tornar os vocabulários controlados mais compreensíveis e promover a autonomia dos usuários na recuperação da informação.</p>	<p>devem ser disponibilizados de forma visível e acessível aos usuários, possivelmente integrados à interface de busca. Além disso, é importante oferecer tutoriais sobre seu uso e garantir uma apresentação clara e compreensível dos termos. Isso ajudará os usuários a realizar buscas mais eficientes e precisas no repositório, promovendo sua autonomia e melhorando sua experiência de busca e acesso ao conhecimento acadêmico e científico disponível.</p>
--	--	---	---	--

2023	<p>Defining Semantically Close Words of Kazakh Language with Distributed System Apache Spark</p>	<p>O principal problema abordado no estudo é a necessidade de definir palavras semanticamente próximas na língua cazaque para melhorar a precisão e eficiência dos sistemas de recuperação de informações. Isso é crucial para facilitar a busca por documentos relacionados e relevantes, especialmente em um contexto em que os recursos linguísticos são limitados.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>Uma possível solução proposta é utilizar métodos de similaridade semântica e algoritmos distribuídos, como o Apache Spark, para identificar palavras semanticamente relacionadas no idioma cazaque. Isso envolve representar as palavras em forma de vetores, calcular a similaridade entre termos e desenvolver um método distribuído para executar tarefas de comparação em paralelo. Essa abordagem pode reduzir significativamente o tempo de cálculo necessário para identificar pares de palavras semanticamente próximas, melhorando assim o desempenho em tarefas de recuperação de informações.</p>	<p>Em um repositório institucional, essa solução pode ser aplicada implementando o método proposto para melhorar a recuperação de informações. Ao integrar o Apache Spark com o sistema de busca do repositório, é possível aprimorar a precisão na identificação de documentos relacionados, facilitando a identificação de conteúdo relevante para os usuários. Além disso, a abordagem distribuída pode acelerar o processo de comparação de palavras, tornando a busca mais eficiente e permitindo uma análise mais abrangente do conteúdo disponível no repositório. Essa melhoria na precisão e eficiência da busca pode melhorar significativamente a experiência do usuário e aumentar a utilidade do repositório como fonte confiável de informações.</p>
------	--	---	---	--

**Fonte:** elaborada pela autora

O estudo "The state-of-the-art in expert recommendation systems" de 2019 investiga a análise e avaliação dos sistemas de recomendação de especialistas. Esta pesquisa se insere na categoria de Recurso de Informação, ao oferecer uma revisão abrangente e crítica do estado atual desses sistemas, destacando suas características, métodos de avaliação e desafios enfrentados. A subcategoria relevante é Documento, uma vez que o estudo constitui uma revisão completa da literatura existente sobre sistemas de recomendação de especialistas.

O principal objetivo deste estudo é fornecer insights sobre os sistemas de recomendação de especialistas, identificando suas vantagens, desvantagens e propondo direções futuras de pesquisa para superar os desafios existentes na área. Ele busca consolidar o conhecimento atual sobre esses sistemas, oferecendo uma compreensão aprofundada do estado-da-arte nesse campo.

Os resultados da pesquisa englobam uma revisão detalhada dos sistemas de recomendação de especialistas, uma análise das vantagens e desvantagens de cada abordagem, além de descrever métricas de avaliação empregadas e visões sobre desafios futuros. Esses resultados constituem um recurso valioso para pesquisadores e profissionais interessados nesse campo, ajudando a orientar pesquisas futuras e a melhorar a eficácia dos sistemas de recomendação de especialistas.

No contexto de repositórios institucionais, as abordagens discutidas no estudo podem ser aplicadas para aprimorar a identificação e recomendação de especialistas em áreas específicas do conhecimento. Ao implementar essas técnicas, os repositórios podem facilitar a descoberta de especialistas, promovendo a colaboração acadêmica e a troca de conhecimento entre pesquisadores. A utilização de métricas de avaliação adequadas também é fundamental para medir a eficácia desses sistemas, contribuindo para a melhoria contínua de suas funcionalidades e garantindo uma recomendação mais precisa e relevante de especialistas nos repositórios institucionais. Assim, a aplicação dos resultados do estudo pode aumentar significativamente o valor e a utilidade desses repositórios como fontes confiáveis de informação especializada.

O estudo "Accuracy evaluation of methods and techniques in Web-based question answering systems: a survey" de 2019 aborda a avaliação da precisão de métodos e técnicas em sistemas de resposta a perguntas baseados na web. Este trabalho se enquadra na categoria de Recursos de Informação, uma vez que se concentra na avaliação e comparação de diferentes abordagens para melhorar a

precisão das respostas fornecidas por esses sistemas. A subcategoria é relevante a Representação do documento, pois o estudo examina métodos e técnicas específicas utilizadas em sistemas de resposta a perguntas.

O objetivo principal deste estudo é avaliar quais métodos e técnicas são mais eficazes na melhoria da precisão das respostas em sistemas de resposta a perguntas na web. Os pesquisadores buscam identificar combinações de métodos que resultem em maior precisão e eficácia nesses sistemas. Além disso, eles têm como meta desenvolver um sistema de resposta a perguntas baseado nos métodos e técnicas identificados como os mais eficazes, a fim de demonstrar sua superioridade em relação a outros métodos disponíveis.

Os resultados destacam a importância da avaliação e comparação de métodos e técnicas para aprimorar a precisão das respostas em sistemas de resposta a perguntas baseados na web. Identificar as melhores combinações pode levar a um aumento significativo na eficácia desses sistemas. Além disso, o desenvolvimento de um sistema baseado nos métodos e técnicas mais eficazes demonstrou uma melhoria substancial em comparação com sistemas de linha de base existentes.

No contexto de repositórios institucionais, os resultados deste estudo podem ser aplicados para melhorar a precisão das respostas fornecidas pelos sistemas de questionamento e resposta internos. Ao selecionar e implementar métodos e técnicas identificados como os mais eficazes, os repositórios podem oferecer aos usuários respostas mais precisas e eficientes. Isso não só aumenta a utilidade dos repositórios institucionais como fontes confiáveis de informação, mas também contribui para uma melhor experiência do usuário ao buscar informações específicas dentro desses repositórios. Portanto, a aplicação dos resultados do estudo pode levar a uma melhoria significativa na qualidade do serviço oferecido pelos repositórios institucionais.

A aplicação dos métodos e técnicas identificados como os mais eficazes neste estudo pode melhorar significativamente a capacidade dos *chatbots* de oferecer respostas precisas e eficientes às perguntas dos usuários. Por exemplo, ao empregar algoritmos de classificação e recuperação de informações, os *chatbots* podem identificar os documentos mais relevantes em resposta a perguntas específicas, garantindo uma resposta direcionada e útil.

Além disso, ao treinar os *chatbots* com base nos modelos de resposta a perguntas desenvolvidos neste estudo, é possível aprimorar sua compreensão e

capacidade de fornecer respostas relevantes. Isso inclui a utilização de técnicas avançadas de PLN para entender a intenção do usuário e extrair informações pertinentes dos documentos armazenados no repositório.

A integração dos *chatbots* com os resultados deste estudo proporciona uma experiência de usuário mais satisfatória nos repositórios institucionais. Os usuários podem fazer perguntas de maneira natural e receber respostas precisas e contextualizadas, agilizando assim o acesso às informações desejadas. Isso fortalece a utilidade e eficácia dos repositórios como fontes de informação confiáveis e acessíveis, contribuindo para uma interação mais eficiente e produtiva com os dados armazenados.

O estudo "Automated feedback generation for formal manufacturing rule extraction" de 2019, está diretamente relacionado à categoria de Representação do problema do usuário e à subcategoria de *Relevance Feedback*. A categoria de Representação do problema do usuário é apropriada porque o estudo se concentra no desenvolvimento de um método para lidar com a complexidade da extração automatizada de regras de fabricação a partir de textos não estruturados. A utilização de técnicas tradicionais de PLN enfrenta limitações ao lidar com essas complexidades, devido à necessidade de validação cuidadosa do texto de entrada para garantir a extração correta das regras desejadas. Portanto, o estudo visa superar essas limitações, desenvolvendo um método que ofereça feedback automatizado para validar e aprimorar os textos de entrada, facilitando assim a extração eficaz e precisa das regras de fabricação formais.

A subcategoria de *Relevance Feedback* é relevante porque o estudo se propõe a desenvolver um método que forneça feedback automatizado para validar e aprimorar textos de entrada não estruturados. Esse feedback permite aos usuários modificar o texto de entrada sem a necessidade de aprender linguagens estruturadas formais, superando as limitações associadas a textos que demandam validação. Dessa forma, o método proposto visa oferecer um feedback relevante para ajustes que facilitem a extração correta das regras de fabricação desejadas.

Os resultados do estudo demonstram a eficácia do método de geração automatizada de *feedback* para extração de regras de fabricação formal. Esse método ajudou os usuários a identificar a necessidade de validação do texto de entrada, proporcionando feedback relevante para ajustes. Cerca de 30% dos casos que exigiam validação receberam *feedback* útil, resultando na extração correta das regras

de fabricação desejadas após a modificação do texto. Todos os casos relevantes que receberam *feedback* alcançaram a extração correta das regras. Esses resultados destacam a utilidade e viabilidade do método proposto na validação e aprimoramento de textos para uma extração precisa de regras de fabricação formal.

No contexto dos repositórios institucionais, a aplicação desse estudo pode auxiliar na validação e melhoria de textos não estruturados relacionados à fabricação. Ao utilizar o método proposto, os repositórios podem oferecer aos usuários uma ferramenta automatizada que fornece *feedback* relevante para aprimorar a qualidade e precisão dos textos armazenados. Isso facilita a extração correta das regras de fabricação, tornando a informação mais acessível e útil para os usuários que consultam o repositório em busca de conhecimento sobre processos de fabricação. Assim, o estudo contribui para aprimorar a eficiência na gestão de informações relacionadas à fabricação em repositórios institucionais.

Embora o estudo se concentre na geração automatizada de *feedback* para extração de regras de fabricação formal, ele pode ser relacionado à folksonomia devido à sua relevância na organização e recuperação de informações. A folksonomia é um método colaborativo de categorização de informações, no qual os usuários atribuem *tags* ou palavras-chave aos recursos digitais com base em sua própria interpretação e contexto. Nesse sentido, a geração automatizada de *feedback* pode ser vista como uma forma de facilitar a participação dos usuários na organização e aprimoramento dos textos não estruturados.

Ao fornecer *feedback* automatizado para validar e aprimorar os textos de entrada, o método proposto no estudo permite que os usuários modifiquem e refinem os textos de acordo com suas necessidades e entendimento do conteúdo. Essa abordagem reflete aspectos da folksonomia, onde a colaboração dos usuários é fundamental para a categorização e organização dos recursos digitais.

Além disso, assim como na folksonomia, onde as *tags* atribuídas pelos usuários refletem suas perspectivas individuais e contribuem para uma organização flexível e dinâmica da informação, o *feedback* automatizado fornecido pelo método proposto no estudo permite uma adaptação contínua dos textos de entrada com base nas necessidades e *feedback* dos usuários. Isso resulta em uma melhoria contínua da qualidade e precisão das informações, tornando a extração de regras de fabricação mais eficaz e alinhada com as expectativas dos usuários.

Portanto, embora o estudo se concentre em aspectos técnicos de PLN e extração de regras de fabricação formal, sua aplicação e resultados podem ser relacionados aos princípios e práticas da folksonomia, especialmente no que diz respeito à participação dos usuários na organização e aprimoramento das informações. Essa intersecção entre a geração automatizada de feedback e a folksonomia destaca a importância da colaboração e do envolvimento dos usuários na gestão e organização de informações, tanto em contextos técnicos quanto em ambientes de colaboração mais amplos.

O estudo "A New Hybrid Improved Method for Measuring Concept Semantic Similarity in WordNet" de 2020 aborda a necessidade de aprimorar a medição da similaridade semântica entre conceitos em *WordNet*, especialmente considerando a distinção de similaridades entre nós de um mesmo nível. Essa pesquisa se encaixa na categoria de Recurso de Informação, pois propõe um método aprimorado para medir a similaridade semântica entre conceitos, visando melhorar a qualidade das informações disponíveis. A subcategoria relevante é Informação, pois o estudo busca aprimorar a precisão e relevância das informações disponíveis em *WordNet*.

O principal objetivo do estudo é propor um novo método híbrido para medir a similaridade semântica entre conceitos, superando as limitações de métodos anteriores, como o método de Wu e Palmer. A pesquisa busca resolver as limitações identificadas nesse método, oferecendo uma abordagem mais eficaz e precisa para medir a similaridade entre conceitos em *WordNet*.

Os resultados da pesquisa demonstraram a eficácia do novo método proposto em distinguir com precisão as similaridades entre nós, tanto de um nó para diferentes nós do mesmo nível quanto em outros contextos semânticos. Isso sugere que o método híbrido pode ser uma ferramenta valiosa para medir a similaridade semântica de conceitos de forma mais precisa e abrangente.

No contexto de repositórios institucionais, a aplicação desse estudo pode melhorar significativamente a precisão na recuperação de informações e no gerenciamento do conhecimento. Ao implementar o novo método proposto, os repositórios podem facilitar uma análise mais aprofundada da similaridade semântica entre os conceitos armazenados, permitindo uma organização mais eficiente e uma recuperação mais precisa de informações relevantes. Isso contribui para uma melhor utilização e aproveitamento dos recursos disponíveis nos repositórios institucionais,

beneficiando pesquisadores, profissionais e estudantes que dependem dessas informações para suas atividades acadêmicas e de pesquisa.

O estudo realizado em 2021 sobre o “Hierarchical Matching Network for Multi-Turn Response Selection in Retrieval-Based Chatbots”, o qual introduz o Modelo de Rede de Correspondência Hierárquica (HMN). Esse modelo propõe uma abordagem inovadora para aprimorar a seleção de respostas em *chatbots*, visando aumentar a eficácia da interação com os usuários. Ao aplicar esse modelo em repositórios institucionais, é possível otimizar a busca e recuperação de dados, fornecendo respostas mais relevantes e contextuais.

Este estudo está intimamente ligado à categoria de Representação do problema do usuário, concentrando-se na análise e compreensão da linguagem humana em diálogos entre *chatbots* e usuários. O HMN é projetado para capturar informações de correspondência em diferentes níveis - palavra e enunciado - a fim de melhorar a seleção de respostas em diálogos multi-turnos. Isso implica compreender não apenas o significado das palavras individualmente, mas também a estrutura e o contexto dos enunciados em uma conversa, essenciais para uma interação eficaz.

Além disso, este estudo está diretamente relacionado à subcategoria de Interação e IA, uma vez que busca melhorar a interação entre o *chatbot* e o usuário. Ao desenvolver um modelo mais eficaz de seleção de respostas, o objetivo é proporcionar uma experiência de interação mais natural, personalizada e eficiente. Isso requer compreender as necessidades e intenções do usuário ao longo de uma conversa, permitindo uma resposta mais adequada e contextualizada.

Assim, ao aplicar técnicas avançadas de PLN e focar na melhoria da interatividade do usuário, este estudo contribui para avanços significativos na área de PLN. Especificamente, no contexto de sistemas de conversação baseados em recuperação, como os *chatbots*, onde a seleção precisa de respostas é crucial para uma experiência satisfatória do usuário.

Como já foi apresentado anteriormente por Zeroual e Lakhouaja (2018), os avanços observados refletem a importância do PLN na capacitação dos *chatbots* para compreender e gerar respostas em linguagem natural de maneira eficaz. O propósito fundamental do PLN é permitir que esses sistemas de conversação analisem as mensagens dos usuários, discernindo a intenção subjacente às perguntas e fornecendo respostas pertinentes e coesas. Adicionalmente, o PLN desempenha um papel crucial no aprimoramento da interação entre os *chatbots* e os usuários,

conferindo às conversas um caráter mais natural e significativo. Esta afirmação é sustentada por evidências empíricas que demonstram a eficácia do PLN na melhoria da comunicação entre humanos e máquinas, promovendo uma interação mais fluida e satisfatória.

Outro estudo relevante é o "Natural Language Semantic Answering Applied to Medicinal Plant and Coronavirus" de 2022, que apresenta o desenvolvimento do Sistema PryRe. Esse sistema representa uma contribuição significativa ao utilizar tanto o PLN quanto ontologias para responder a perguntas específicas. Sua integração em repositórios institucionais pode ter um impacto positivo na organização e compreensão dos conteúdos disponíveis, facilitando assim a recuperação de informações relevantes para os usuários.

Este estudo está intrinsecamente ligado à categoria de Representação do problema do usuário, pois concentra-se na análise e compreensão da linguagem humana em perguntas relacionadas a plantas medicinais e coronavírus. O objetivo central é o desenvolvimento de um sistema de resposta a perguntas em linguagem natural, o PryRe, que utiliza propriedades semânticas e análise de ontologias para oferecer respostas precisas baseadas em um conhecimento estruturado e dedutivo.

Além disso, ele se enquadra na subcategoria de Interação e IA, pois busca aprimorar a interação entre os usuários e o sistema PryRe. Ao fornecer respostas relevantes e precisas com base nas perguntas formuladas pelos usuários, utilizando uma ontologia específica do domínio, o estudo visa tornar essa interação mais natural e eficaz. Isso implica compreender as necessidades e intenções dos usuários ao formular suas perguntas, proporcionando respostas que vão além do trivial.

Ao integrar técnicas avançadas de PLN, ontologias e métodos de recuperação de informações, o Sistema PryRe contribui significativamente para avanços na área de PLN, especialmente no contexto de sistemas de resposta a perguntas. Sua aplicação em repositórios institucionais pode melhorar consideravelmente a recuperação de informações, fornecendo respostas mais precisas e relevantes às consultas dos usuários, tornando a interação com esses sistemas mais intuitiva e eficaz. Isso é crucial na contemporaneidade, onde a quantidade exponencial de informações torna essencial a implementação de sistemas eficientes de organização, como a folksonomia.

O estudo de 2022 intitulado "On the Effectiveness of Pre-Trained Language Models for Legal Natural Language Processing: An Empirical Study" se enquadra na

categoria Representação do problema do usuário. A justificativa para essa classificação reside na abordagem do problema de pesquisa e nos objetivos propostos. O estudo investiga a eficácia de modelos de linguagem pré-treinados em tarefas específicas de PLN no domínio jurídico, com o intuito de oferecer *insights* sobre como esses modelos podem atender às necessidades reais de informação dos usuários nesse contexto.

A categoria é apropriada porque o estudo concentra-se em compreender e resolver um problema específico enfrentado pelos usuários, ou seja, a eficácia na utilização de modelos de linguagem pré-treinados para processar informações jurídicas. A subcategoria *Relevance Feedback*, expansão da consulta e serviços de recomendação é relevante porque o estudo busca identificar as vantagens e limitações desses modelos em relação às demandas reais por informações no campo jurídico, contribuindo assim para aprimorar a forma como as informações legais são acessadas e utilizadas.

A aplicação prática dos resultados desse estudo pode ter um impacto significativo na melhoria dos sistemas de indexação e recuperação de informações legais em repositórios institucionais. A implementação bem-sucedida desses modelos pode agilizar o processo de busca e recuperação de dados legais, fornecendo aos usuários acesso mais rápido e preciso a informações relevantes. Isso não apenas aumenta a eficiência e a produtividade dos profissionais jurídicos, mas também promove uma melhor prestação de serviços jurídicos aos clientes e à sociedade em geral.

Portanto, a integração de técnicas avançadas de PLN e modelos de linguagem pré-treinados na indexação e recuperação de informações jurídicas representa um passo importante para avançar na pesquisa e na prática do Direito. Essa abordagem inovadora tem o potencial de transformar fundamentalmente a forma como os profissionais jurídicos acessam e utilizam informações legais, proporcionando benefícios significativos para toda a comunidade jurídica.

O estudo "Natural Language Processing Applied to Forensics Information Extraction With Transformers and Graph Visualization" de 2022 sobre o PLN aplicado à forense digital está intrinsecamente vinculado à categoria de Representação do problema do usuário e à subcategoria de Interação e IA. A escolha dessas categorias e subcategorias é justificada pela essência do problema de pesquisa e pelos objetivos delineados pelos autores. A categoria de Representação do problema do usuário é

apropriada dado que o estudo concentra-se no desenvolvimento e aplicação de técnicas avançadas de PLN, tais como o reconhecimento de entidades nomeadas e a extração de relações, para a obtenção de informações em investigações forenses digitais. Essas técnicas são fundamentais para lidar com a complexidade e a heterogeneidade dos dados digitais encontrados nessas investigações, permitindo a análise e interpretação eficazes dessas informações.

A subcategoria de Interatividade do Usuário também é relevante, uma vez que o estudo almeja aprimorar a interação entre os usuários e os sistemas de análise de dados forenses. Ao desenvolver modelos aprimorados por meio de arquiteturas de redes neurais e transformadores, os autores têm como objetivo automatizar a análise de extração de informações em diferentes contextos forenses e de investigação digital. Isso não apenas aumenta a eficiência da análise de grandes volumes de dados, mas também facilita a compreensão e interpretação desses dados pelos usuários, promovendo, assim, uma interação mais eficaz e produtiva com os sistemas de análise.

Os resultados do estudo demonstraram que a combinação dessas técnicas avançadas de PLN, aprimoradas por meio de arquiteturas de transformadores, pode aprimorar significativamente a eficiência e a precisão na extração de informações em investigações forenses digitais. Tais resultados indicam não apenas o potencial dessas técnicas para fortalecer a segurança cibernética e combater crimes digitais, mas também sua aplicabilidade em outras áreas, como análise de Big Data e inteligência empresarial.

A contribuição desse estudo para um repositório institucional seria significativa em termos de otimização da gestão de informações e fortalecimento da segurança cibernética. Ao aplicar técnicas avançadas de PLN e arquiteturas de transformadores na extração de informações em investigações forenses digitais, o estudo oferece uma abordagem inovadora para lidar com grandes volumes de dados heterogêneos de forma automatizada e eficiente. Essa automatização da análise de dados permite que um repositório institucional seja capaz de processar e interpretar uma ampla gama de informações, incluindo documentos digitais, registros de atividades e comunicações eletrônicas, de maneira rápida e precisa.

Isso não apenas agiliza os processos de busca e recuperação de informações relevantes, mas também fortalece a segurança cibernética ao identificar padrões e anomalias que possam indicar atividades suspeitas ou ameaças à segurança. Além

disso, a utilização de grafos interativos para visualização dos dados extraídos proporciona uma representação visual clara e intuitiva das relações entre as entidades e conceitos identificados nas investigações forenses. Isso facilita a compreensão e interpretação dos resultados pelos usuários, permitindo uma interação mais eficaz com o repositório institucional.

Portanto, ao integrar essas técnicas avançadas de PLN e transformadores em um repositório institucional, o estudo contribui para a melhoria da eficiência na gestão de informações, aprimoramento da segurança cibernética e facilitação da interação dos usuários com os dados armazenados no repositório. Essa contribuição é essencial para instituições que lidam com grandes volumes de dados digitais e buscam fortalecer sua capacidade de análise e proteção de informações sensíveis.

O estudo de 2022 sobre “Legal Information Retrieval systems: State-of-the-art and open issues” está diretamente ligado à categoria de Representação do problema do usuário e à subcategoria de Interação e IA. A escolha dessas categorias e subcategorias é embasada na essência do problema de pesquisa e nos objetivos delineados pelos autores.

A categoria de Representação do problema do usuário é apropriada porque o estudo busca compreender e abordar os desafios enfrentados pelos profissionais do direito ao buscar e extrair informações de documentos legais. Os autores discutem questões não resolvidas, como a classificação de documentos, a extração de informações relevantes, a compreensão semântica e a robustez dos sistemas de recuperação de informações legais, refletindo as necessidades reais dos usuários.

A subcategoria de Interação e IA também é relevante, já que o estudo procura fornecer uma visão abrangente do estado atual da recuperação de informações legais e identificar lacunas existentes nesse campo. Os autores não apenas discutem os desafios enfrentados, mas também oferecem insights valiosos para orientar futuras pesquisas e desenvolvimentos, demonstrando uma preocupação genuína em atender às necessidades práticas dos profissionais do direito.

Os resultados do estudo incluem uma análise detalhada do estado atual dos sistemas de recuperação de informações legais, destacando as tecnologias e abordagens mais recentes utilizadas nesse campo. Além disso, os autores discutem os desafios e questões em aberto, bem como as tendências emergentes, como o uso de IA e PLN, essenciais para informar profissionais do direito, pesquisadores e desenvolvedores sobre as melhores práticas e tecnologias disponíveis.

No contexto dos repositórios institucionais, a aplicação desses insights pode melhorar significativamente a recuperação de informações legais armazenadas nessas plataformas. O uso de tecnologias avançadas, como IA e PLN, pode agilizar a busca e extração de informações relevantes em documentos legais, beneficiando instituições ao facilitar o acesso a dados jurídicos importantes. Isso fortalece a eficiência na pesquisa e análise de casos, garantindo uma recuperação mais precisa e contextualizada de informações legais para os profissionais do direito que utilizam os repositórios institucionais.

No contexto dos *chatbots*, os resultados e análises desse estudo podem ser fundamentais para aprimorar a capacidade desses sistemas de conversação automatizada em lidar com consultas legais de forma eficiente e precisa. Ao entender os desafios e questões em aberto nos sistemas de recuperação de informações legais, os desenvolvedores de *chatbots* podem adaptar suas soluções para oferecer respostas mais relevantes e contextualizadas aos usuários. Por exemplo, os *chatbots* podem ser treinados com base nos padrões de busca identificados no estudo, permitindo que ofereçam orientações mais precisas sobre onde encontrar informações legais específicas dentro de um repositório institucional. Além disso, o estudo pode inspirar o desenvolvimento de *chatbots* jurídicos mais sofisticados, capazes de interpretar nuances semânticas e responder a consultas complexas dos usuários.

No que diz respeito aos repositórios institucionais, as descobertas desse estudo podem fornecer orientações valiosas para aprimorar a eficácia na gestão de informações legais nessas plataformas. Ao identificar as lacunas e desafios na recuperação de informações legais, as instituições podem direcionar seus esforços para melhorar a organização e acessibilidade dos dados armazenados. Por exemplo, os resultados do estudo podem motivar a implementação de técnicas avançadas de PLN para indexação e categorização automatizada de documentos legais em repositórios institucionais. Isso facilitaria a busca e recuperação de informações relevantes para profissionais do direito e outros usuários que dependem desses recursos para suas atividades.

Assim, ao relacionar esse estudo tanto aos *chatbots* quanto aos repositórios institucionais, é possível promover avanços significativos na forma como as informações legais são acessadas, compreendidas e utilizadas, beneficiando tanto os profissionais do direito quanto o público em geral.

O estudo de 2022 intitulado "FabNER: information extraction from manufacturing process Science domain literature using named entity recognition" está diretamente associado à categoria de Recursos de informação e à subcategoria de Documento. Essa categorização é apropriada, pois o estudo concentra-se no desenvolvimento de um modelo de Reconhecimento de Entidades Nomeadas (NER) para extrair informações em larga escala da literatura científica sobre processos de fabricação. O NER é uma técnica vital dentro do campo de PLN, que visa identificar entidades nomeadas em textos não estruturados, como nomes de pessoas, locais, organizações e termos específicos de domínio, sendo essencial para a extração precisa e relevante de informações.

A subcategoria de Documento também é pertinente, uma vez que o estudo aborda os desafios associados à extração de informações específicas relacionadas à fabricação a partir de textos científicos. A aplicação de técnicas de NER enfrenta ambiguidades e desafios na identificação e categorização de entidades em um contexto específico, destacando a importância de abordagens avançadas para lidar com essas questões.

Os resultados da pesquisa revelaram que o modelo de NER desenvolvido foi eficaz na extração de informações relevantes da literatura científica sobre processos de fabricação, alcançando uma precisão geral de 88%. Além disso, o estudo demonstrou a aplicabilidade prática do modelo em dois casos de uso específicos, incluindo a identificação de tópicos de artigos. Isso evidencia a utilidade e viabilidade do modelo de NER para a extração de informações em larga escala e sua aplicabilidade prática na análise de conteúdo científico.

No âmbito dos repositórios institucionais, a implementação desse estudo pode aprimorar a indexação e recuperação de informações relacionadas à fabricação. Ao incorporar o modelo de NER desenvolvido, os repositórios podem automatizar a categorização de documentos científicos sobre fabricação, facilitando a busca e organização de conteúdo para os usuários. Isso não apenas aumenta a eficiência na descoberta de conhecimento dentro do domínio da fabricação, mas também beneficia pesquisadores, profissionais e estudantes que dependem desses repositórios para acessar informações relevantes e atualizadas sobre processos de fabricação.

O estudo "Dealing with textual noise for robust and effective BERT re-ranking" de 2022 investiga o impacto do ruído textual na eficácia e robustez dos modelos de reclassificação Bidirectional Encoder Representations from Transformers (BERT) em

tarefas de recuperação de informações. Esta pesquisa se enquadra na categoria de Representação do problema do usuário, pois explora como os modelos de PLN, especificamente os modelos BERT, lidam com o ruído textual em documentos. A subcategoria relevante é *Relevance Feedback*, expansão da consulta e serviços de recomendação, uma vez que o estudo aborda os desafios associados à presença de ruído textual e como isso pode afetar a eficácia dos modelos de reclassificação BERT em recuperar informações relevantes.

O objetivo principal do estudo é avaliar como os modelos de reclassificação BERT se comportam em cenários do mundo real, nos quais o ruído textual é comum. Os autores conduzem experimentos no conjunto de dados Microsoft Machine Reading Comprehension (MS MARCO) para investigar se os modelos BERT conseguem manter um bom desempenho mesmo quando confrontados com textos ruidosos. Além disso, propõem um modelo chamado De-Ranker, projetado para lidar de forma mais eficaz com o ruído textual e melhorar o desempenho em comparação com os modelos BERT convencionais.

Os resultados da pesquisa destacam que os modelos BERT re-ranker são impactados negativamente pela presença de ruído textual, resultando em uma degradação significativa na eficácia da classificação. No entanto, ao introduzir o modelo De-Ranker, observou-se uma melhoria de desempenho em comparação com os modelos BERT tradicionais. O De-Ranker foi capaz não apenas de lidar com ruído textual sintético, mas também demonstrou melhorias na capacidade de generalização entre domínios.

No contexto de repositórios institucionais, a aplicação desse estudo pode ser significativa. A implementação de modelos como o De-Ranker pode melhorar a precisão e relevância dos resultados de busca em repositórios institucionais, considerando a presença comum de ruído textual em documentos acadêmicos e científicos. Essa abordagem pode mitigar os efeitos negativos do ruído textual, garantindo uma recuperação de informações mais eficaz e precisa para os usuários que acessam esses repositórios em busca de conteúdo relevante. Assim, ao enfrentar os desafios do ruído textual, os modelos como o De-Ranker podem melhorar a experiência do usuário e a utilidade dos repositórios institucionais como fontes confiáveis de informação.

No estudo "Building a Technology Recommender System Using Web Crawling and Natural Language Processing Technology" de 2022, a abordagem adotada visa

abordar a dificuldade enfrentada pelos especialistas em acompanhar as tecnologias recentes de forma eficiente e eficaz. O foco está na compreensão dessas tecnologias para decidir se merecem investigações mais detalhadas. Essa pesquisa se enquadra na categoria de Recurso de Informação, oferecendo uma solução para melhorar a descoberta de conhecimento sobre tecnologias emergentes. A subcategoria relevante é Informação, visto que o estudo visa automatizar a extração e a sumarização de informações relevantes sobre tecnologia.

O objetivo principal deste estudo é explorar a combinação de web crawling, web scraping e sumarização automática de texto com tecnologia de PLN. O propósito é desenvolver um protótipo de sistema de recomendação de tecnologia, que visa fornecer recomendações sobre tecnologias em tendência para especialistas em determinados propósitos de aplicação. Os resultados da pesquisa indicaram que, embora fosse possível desenvolver um programa para auxiliar na busca e sumarização de informações sobre tecnologia, o programa não alcançou resultados satisfatórios em termos de qualidade das recomendações. Isso se deveu principalmente à extração relativamente aleatória de texto das fontes de conteúdo, resultando em resumos pouco organizados e que não correspondiam bem aos termos de pesquisa.

No contexto de repositórios institucionais, a aplicação desse estudo pode automatizar a busca e a sumarização de informações relevantes sobre tecnologia, facilitando a descoberta de novas tecnologias e auxiliando na tomada de decisões. Essa abordagem torna a recuperação de informações mais eficiente e eficaz para os usuários dos repositórios, melhorando sua experiência de busca e acesso ao conhecimento sobre tecnologia.

A integração da folksonomia nesse contexto poderia enriquecer o processo de recomendação de tecnologias. Ao permitir que os usuários contribuam com *tags* e metadados relevantes para descrever as tecnologias, a folksonomia poderia melhorar a precisão das recomendações feitas pelo sistema. Isso ocorreria ao aumentar a relevância e a contextualização das informações extraídas sobre as tecnologias, tornando mais fácil para os especialistas encontrar e compreender os detalhes que influenciam suas decisões de investigação.

Além disso, a implementação de *chatbots* poderia facilitar ainda mais o acesso às recomendações de tecnologia. Os *chatbots* poderiam interagir diretamente com os usuários, compreendendo suas necessidades e fornecendo recomendações

personalizadas com base em suas consultas. Ao utilizar tecnologias PLN, os chatbots seriam capazes de entender as solicitações dos usuários de forma mais intuitiva, melhorando assim a experiência de busca e facilitando a descoberta de novas tecnologias.

Em repositórios institucionais, a aplicação desse estudo, juntamente com a integração da folksonomia e dos *chatbots*, pode automatizar a busca e a sumarização de informações relevantes sobre tecnologia. Isso facilitaria a descoberta de novas tecnologias e auxiliaria na tomada de decisões, tornando a recuperação de informações mais eficiente e eficaz para os usuários dos repositórios. Assim, a combinação dessas abordagens pode melhorar significativamente a experiência dos usuários ao buscar e acessar conhecimento sobre tecnologia em repositórios institucionais.

O estudo "A disponibilização de vocabulário controlado aos usuários para a recuperação da informação" de 2022 analisa a presença e a disposição de vocabulários controlados nos catálogos online de bibliotecas universitárias brasileiras, com o objetivo de verificar seu impacto na recuperação da informação pelos usuários. Essa pesquisa se insere na categoria de Representação do problema do usuário, pois aborda a organização e disponibilização de recursos para facilitar o acesso à informação. A subcategoria relevante é Interação e IA, já que propõe melhorias na forma como os vocabulários controlados são apresentados aos usuários.

O estudo investiga como os vocabulários controlados são apresentados nos catálogos online das bibliotecas universitárias, buscando aprimorar sua disponibilização e promover a autonomia dos usuários na recuperação da informação. Além disso, visa contribuir para o aprimoramento das estratégias de busca dos usuários, facilitando o acesso aos vocabulários controlados nos sistemas de informação das bibliotecas.

Os resultados da pesquisa ressaltam a importância de tornar os vocabulários controlados mais visíveis e compreensíveis para os usuários, visando melhorar a eficácia e precisão das pesquisas nos catálogos online das bibliotecas. Recomenda-se uma exposição e destaque mais efetivos desses vocabulários, acompanhados da oferta de tutoriais e treinamentos para os usuários, com o intuito de aprimorar suas habilidades de busca e recuperação de informações.

No contexto dos repositórios institucionais, a aplicação desse estudo pode ser feita através da implementação de vocabulários controlados, disponibilizando-os de

forma visível e acessível aos usuários. Além disso, é importante oferecer tutoriais sobre seu uso e garantir uma apresentação clara e compreensível. Essas práticas contribuirão para uma busca mais eficiente de informações nos repositórios, promovendo a autonomia dos usuários e melhorando sua experiência de busca acadêmica e científica.

Esses estudos destacam a diversidade de abordagens e aplicações que têm sido exploradas na área de Recuperação de Informação, demonstrando o contínuo esforço dos pesquisadores em desenvolver soluções inovadoras para os desafios enfrentados nesse campo. Ao integrar essas contribuições à pesquisa mais ampla em repositórios institucionais, podemos avançar ainda mais na melhoria da busca e recuperação de informações em diferentes domínios e ambientes institucionais.

O estudo "Defining Semantically Close Words of Kazakh Language with Distributed System Apache Spark" de 2023 aborda a questão de definir palavras semanticamente próximas na língua cazaque, com o propósito de melhorar a precisão e a eficiência dos sistemas de recuperação de informações. Nesse contexto, a pesquisa explora como a utilização de modelos de incorporação de palavras e algoritmos distribuídos, como o Apache Spark, pode auxiliar na identificação de palavras semanticamente relacionadas no idioma cazaque. Esta pesquisa se enquadra na categoria de Recurso de Informação e na subcategoria de Documento.

O objetivo principal do estudo é determinar palavras semanticamente próximas na língua cazaque, empregando métodos de similaridade semântica para melhorar o desempenho em tarefas de recuperação de informações. Para alcançar esse objetivo, a pesquisa propõe representar as palavras em forma de vetores, calcular a similaridade semântica entre termos e desenvolver um método distribuído que execute tarefas de comparação em paralelo, utilizando o Apache Spark. Essa abordagem visa reduzir o tempo de cálculo necessário para identificar pares de palavras semanticamente próximas.

Os resultados da pesquisa englobam o desenvolvimento de um novo conjunto de dados de pares de palavras semanticamente próximas na língua cazaque, a proposição e implementação de um método para encontrar palavras semanticamente similares, a avaliação do desempenho em termos de complexidade temporal e a melhoria do tempo de computação para muitos pares de palavras por meio de uma abordagem paralelizada. Adicionalmente, o estudo propõe um pipeline de processamento para identificar documentos semanticamente similares em idiomas

com recursos linguísticos limitados. Um pipeline consiste em uma série de etapas conectadas entre si, as quais processam os dados de entrada com o intuito de alcançar um objetivo específico de maneira automatizada e eficiente. Amplamente utilizado em áreas como PLN e aprendizado de máquina, o pipeline visa otimizar fluxos de trabalho complexos. Esses resultados demonstram a aplicabilidade do método proposto na língua cazaque e destacam a importância da representação vetorial e da comparação distribuída de palavras para tarefas de PLN.

No contexto dos repositórios institucionais, a aplicação desse estudo pode ser significativa para melhorar a recuperação de informações. Ao implementar o método proposto, os repositórios podem aprimorar a precisão na busca por documentos relacionados, facilitando a identificação de conteúdo relevante e aprimorando a experiência do usuário ao encontrar informações específicas dentro do repositório. Além disso, a abordagem distribuída pode acelerar o processo de comparação de palavras, tornando a busca mais eficiente e permitindo uma análise mais abrangente do conteúdo disponível nos repositórios institucionais.

O Apache Spark é uma ferramenta de processamento distribuído projetada para lidar com grandes volumes de dados linguísticos. Por meio de algoritmos eficientes, o Spark consegue identificar relações semânticas entre palavras e conceitos. Isso contribui significativamente para melhorar a precisão na identificação de palavras semanticamente próximas na língua cazaque. É importante ressaltar que o Apache Spark não é restrito a nenhum idioma específico, incluindo o cazaque. Na verdade, trata-se de uma estrutura de computação distribuída amplamente versátil, aplicável em diversos contextos e idiomas, inclusive o português. Sua capacidade de processar grandes volumes de dados em paralelo e executar análises complexas torna-o uma escolha popular em diferentes domínios, independentemente do idioma em questão.

No contexto de um repositório institucional, a aplicação do Apache Spark pode ser ainda mais significativa. Integrando essa tecnologia com *chatbots*, por exemplo, é possível oferecer aos usuários uma experiência de busca mais inteligente e personalizada. Os *chatbots* podem ser treinados para entender as consultas dos usuários de forma mais precisa e fornecer recomendações de conteúdo mais relevantes com base nas relações semânticas identificadas pelo Apache Spark.

Além disso, o uso do Apache Spark pode acelerar o processamento de consultas e análise de dados, tornando a busca por informações mais eficiente e

rápida. Isso não só melhora a experiência do usuário, mas também aumenta a utilidade e relevância do repositório institucional como uma fonte confiável de informações linguísticas e culturais.

A síntese das contribuições identificadas na literatura revela oportunidades significativas para o avanço do conhecimento, especialmente na interseção entre a CI e a IA. A pesquisa sobre o uso de *chatbots* para auxiliar a recuperação da informação em repositórios institucionais é crucial nesse contexto. Os *chatbots*, agentes virtuais capazes de interagir de forma natural com os usuários, oferecem assistência e respondem a consultas de maneira eficiente. No campo da recuperação da informação, essas ferramentas têm o potencial de facilitar o acesso e a busca por conteúdo nos repositórios institucionais.

Ao integrar *chatbots* nos repositórios institucionais, é possível aumentar a eficiência na recuperação da informação, tornando o processo mais intuitivo e acessível para os usuários. Isso contribui para a disseminação e uso efetivo do conhecimento armazenado nos repositórios, beneficiando tanto pesquisadores quanto o público em geral.

Portanto, a pesquisa sobre o uso de *chatbots* para auxiliar a recuperação da informação em repositórios institucionais representa uma área promissora de estudo, que combina os avanços em CI e IA para melhorar a acessibilidade e usabilidade de recursos de informação essenciais.

O Apêndice A apresenta um levantamento e análise dos resultados de artigos selecionados. O leitor pode esperar encontrar uma lista de artigos relevantes, cada um detalhando seu título, a problemática de pesquisa abordada, os objetivos do estudo, os resultados obtidos e as possíveis aplicações em repositórios institucionais. Ademais, cada artigo foi analisado para identificar sua categoria e subcategoria. Esses dados fornecem uma visão detalhada das pesquisas conduzidas nessa área, permitindo ao leitor compreender melhor as tendências, desafios e oportunidades relacionadas à integração de tecnologias de IA, como *chatbots* e PLN, em ambientes de recuperação da informação, indexação e gestão de repositórios institucionais.

Após a análise dos resultados, foi possível criar a Tabela 3, que sintetiza as possíveis aplicações de *chatbots* em repositórios institucionais. Essa tabela apresenta um panorama das diversas maneiras pelas quais *chatbots* podem ser utilizados para melhorar a eficiência e a acessibilidade desses repositórios, abordando desde a

assistência na navegação e recuperação de informação até o suporte técnico e a interação personalizada com os usuários.

**Tabela 3:** Síntese de possíveis aplicações de *chatbots* em repositórios institucionais

<b>Suporte ao Usuário</b>	Responder a perguntas frequentes sobre o funcionamento do repositório
	Auxiliar os usuários na navegação e busca de informações
	Fornecer orientações sobre como depositar ou acessar conteúdo
	Responder a perguntas frequentes e ajudar a solucionar problemas técnicos
<b>Recomendações Personalizadas</b>	Sugerir documentos com base nas preferências e histórico de busca do usuário
	Oferecer recomendações de pesquisa e leitura relacionadas aos interesses do usuário
	Maximizar a utilidade do repositório ao ajudar os usuários a descobrir novos conteúdos
<b>Assistência na Recuperação de Informações</b>	Auxiliar na elaboração de consultas de busca mais eficazes
	Facilitar a descoberta de documentos relevantes por meio de interações conversacionais
<b>Notificações e Atualizações</b>	Informar os usuários sobre novos conteúdos adicionados ao repositório
	Enviar lembretes sobre datas de vencimento de prazos ou eventos importantes
<b>Auxílio na Busca Avançada</b>	Auxiliar os usuários na realização de buscas complexas e avançadas
	Fornecer sugestões de palavras-chave e estratégias de refinamento de consulta
<b>Integração com Sistemas de Gerenciamento</b>	Integração com sistemas de gerenciamento de conteúdo para atualização automática de informações
	Facilitar o acesso e a interação com o repositório diretamente através de plataformas de mensagens
<b>Suporte Multilíngue e Acessibilidade</b>	Oferecer suporte em diferentes idiomas para atender a uma base de usuários diversificada
	Garantir acessibilidade para usuários com necessidades especiais, como leitores de tela
	Coletar feedback dos usuários para melhorar continuamente o serviço do repositório

<b>Análise de Dados e Feedback do Usuário</b>	Analisar padrões de uso e tendências de busca para otimizar a experiência do usuário
<b>Educação e Treinamento</b>	Oferecer tutoriais interativos sobre como usar efetivamente o repositório
	Treinar usuários em técnicas avançadas de busca e recuperação de informações

**Fonte:** elaborada pela autora

A partir desses resultados demonstrados, é possível refletir sobre a importância de repositórios baseados em tarefas, conforme apresentado por Mizzaro (1998). Nesses ambientes, ao invés de apenas armazenar os documentos ou dados, eles são estruturados para atender tarefas específicas dos usuários. Essas tarefas são traduzidas para *Concrete Characteristics of Documents* (CCD), que são conjuntos de informações relacionadas à tarefa ou ainda, de maneira mais factível, em *Abstract Characteristics of Documents* (ACD), que serão posteriormente transformadas em CCD em um nível maior de abstração.

Um possível conjunto sugerido pelo autor de ACD seria composto por: compreensibilidade - quantos dos documentos recuperados são fáceis de se ser compreendidos; atualidade - quão recente são os documentos recuperados; quantidade - quanto de informação o usuário necessita; linguagem - qual é o idioma dos documentos recuperados; fertilidade - quanto os documentos recuperados serão úteis para encontrar outros documentos na mesma temática.

Esse grupo de ACD seria posteriormente transformado em CCD a partir dos seguintes aspectos: tipo de documento, características do documento como aplicação ou biografia utilizada, número de páginas, data de publicação e idioma (Mizzaro, 1998).

Essas características apresentam as informações de maneira intuitiva nos processos de busca do usuário e fazem com que os resultados sejam mais próximos daquilo que o usuário busca, como o exemplo citado pelos autores, em que um usuário busca especificamente documentos do tipo livros, mas a base de dados recupera na busca também artigos e anais de eventos, isso significa que apenas  $\frac{1}{3}$  dos documentos recuperados seriam aquilo que o usuário busca (Mizzaro, 1998).

Nesse sistema, há ainda a possibilidade da interação do usuário para manipular os resultados. Os documentos mais relevantes são apresentados como pontos e o usuário pode lê-los e movimentá-los para um espaço relevante e essa manipulação

permite ao sistema identificar feedbacks sobre a performance do sistema (Mizzaro, 1998).

Um exemplo para o uso desse sistema é para buscas específicas, como a explicitada no referido artigo de Vajjala et al. (2020), sobre a possibilidade de cozinheiros buscarem por receitas a partir de suas preferências, o que não seria atendido pelos sistemas tradicionais.

Para solucionar questões como essas, os autores sugerem o uso do *Diagflow*, um agente conversacional desenvolvido pelo Google, que compreende e gera experiências em linguagem natural. A partir da definição de entidades - que poderiam ser ingredientes, nível de calorias, tempo de preparo, etc. - e posteriormente, treinar o *chatbot* a partir de questões específicas como “quero uma sobremesa de baixas calorias e vegana”.

Após esse treinamento, o *Diagflow* consegue identificar as preferências do usuário para recuperar uma resposta correta e, além disso, manterá sempre o ambiente de chat aberto, para que possa compreender ainda melhor as necessidades do usuário. Dessa forma, é possível construir um chatbot para uso em sistemas de recuperação da informação.

Outra estrutura que pode ser utilizada para essa construção é a Rasa (<https://rasa.community/>), uma plataforma de código aberto disponível para implementar e gerenciar *chatbots* ou agentes conversacionais. Ela é composta por *Natural Language Understanding* (NLU), que é a parte que lida com a compreensão em linguagem natural, extraindo intenções e entidades das mensagens dos usuários e a Rasa Core, responsável pela gestão do diálogo, definindo o que o *chatbot* deve dizer na continuidade da conversa.

O Rasa possui como principais funcionalidades treinamento de modelos *machine learning* para interpretação de mensagens; personalização da plataforma, permitindo que os desenvolvedores criem *chatbots* de acordo com o uso de seus sistemas; cursos para aprendizado sobre a ferramenta e inteligência artificial; uma comunidade engajada para troca de informações.

Dessa forma e, considerando todo o exposto dos artigos recuperados, sugere-se a construção de repositórios orientados a tarefas e não apenas depósitos de documentos e, para construção de *chatbots* para efetiva aplicação das possibilidades na Tabela 3, o uso de ferramentas como o *Diagflow* ou a plataforma Rasa.

## CONSIDERAÇÕES FINAIS

A pesquisa se concentrou em explorar como o PLN e os *chatbots* podem ser aplicados para melhorar a recuperação da informação em repositórios institucionais. Os objetivos específicos incluíram uma análise detalhada dos problemas fundamentais que têm orientado as pesquisas nesse campo, a avaliação das soluções propostas em cada estudo e a identificação das aplicações específicas voltadas para aprimorar a recuperação de informação, destacando as contribuições práticas e os potenciais benefícios do processo de busca e recuperação de informação.

A pesquisa foi desenvolvida com a intenção de fornecer *insights* valiosos e contribuições significativas para avançar o conhecimento e as práticas na área da recuperação da informação, especialmente em repositórios institucionais. A metodologia adotada foi criteriosa, garantindo a precisão e relevância na seleção das informações pertinentes ao tema, consolidando a pesquisa em uma abordagem robusta alinhada aos objetivos delineados.

As tendências identificadas nas pesquisas analisadas revelaram um cenário promissor no uso de PLN e *chatbots* para a recuperação da informação em repositórios institucionais. A integração dessas tecnologias visa aprimorar a interação dos usuários com os repositórios, simplificando o processo de busca e tornando-o mais eficiente e intuitivo.

Um aspecto relevante identificado foi a melhoria na compreensão das consultas realizadas pelos usuários. Os sistemas baseados em PLN e *chatbots* estão cada vez mais aptos para interpretar consultas em linguagem natural, fornecendo respostas mais precisas e relevantes de acordo com as necessidades dos usuários.

No contexto dos repositórios institucionais, a integração dessas tecnologias não apenas proporciona benefícios práticos, como uma recuperação de informações mais eficiente, mas também representa um avanço significativo na acessibilidade e usabilidade dos recursos de informação disponíveis, beneficiando tanto os pesquisadores quanto o público em geral.

Desta forma, a implementação de *chatbots* em repositórios institucionais traz benefícios significativos, como a melhoria da experiência do usuário, respostas rápidas e precisas, automatização de tarefas repetitivas, personalização das recomendações, redução da carga de trabalho e otimização dos recursos. Os *chatbots* oferecem uma interface interativa e amigável, estão disponíveis a qualquer momento,

fornece respostas instantâneas, automatiza tarefas rotineiras, aprende preferências individuais, reduz a carga de trabalho dos funcionários e contribui para a eficiência operacional. A presença de *chatbots* nos repositórios institucionais moderniza os serviços, agiliza o acesso às informações e melhora a interação entre usuários e documentos, trazendo benefícios tanto para os usuários quanto para a instituição.

A ideia de uma maior interação e iteração nos processos de busca em repositórios institucionais, facilitado, por exemplo, pelas técnicas de feedback de relevância e recomendação, permite-nos perceber algumas vantagens do uso de *chatbots* e da linguagem natural. Alguns termos decorrentes de variantes de língua e que requerem normalização, como *ônibus*, *ómnibus* e *autocarro*, podem ser facilmente tratados como equivalentes, diminuindo alguns problemas atuais das folksonomias. Word2Vec é um modelo de PLN que é usado para aprender incorporações de palavras a partir de grandes conjuntos de dados. Ao contrário dos modelos tradicionais que procuram determinar a semântica das palavras com base na sua ocorrência em frases, o Word2Vec utiliza outras propriedades, como proximidade entre palavras, para determinar o contexto. Na CI, por exemplo, a ocorrência isolada dos termos “ciência” e “informação” tem uma semântica diferente para efeitos de recuperação. Porém, a proximidade de ocorrência é frequentemente utilizada para se referir ao campo ou área de estudos “Ciência da Informação”. A proximidade de ocorrência entre termos pode ser facilmente interpretada pelos *chatbots*, de modo a enriquecer a semântica nos processos de busca em repositórios institucionais e diminuir alguns problemas atuais de ambiguidade.

A pesquisa tem uma importante contribuição social, pois melhora a acessibilidade e usabilidade dos recursos de informação, promovendo a disseminação ampla do conhecimento científico, como por exemplo, guiando os usuários na navegação do repositório, fornecendo respostas rápidas e precisas a perguntas específicas e adaptando-se às necessidades dos usuários ao longo do tempo. Isso contribui para a democratização do acesso à informação e para o avanço da educação e pesquisa em diversas áreas do conhecimento.

Para trabalhos futuros, e também para expandir o campo da CI relacionada ao uso de *chatbots* na recuperação da informação em repositórios institucionais, há várias oportunidades de pesquisa a serem exploradas. Uma área promissora de estudo seria a investigação de técnicas avançadas de PLN para aprimorar a

capacidade dos *chatbots* de compreender consultas complexas e contextos específicos. Isso poderia incluir o desenvolvimento de algoritmos mais sofisticados para interpretar nuances semânticas e contextuais, aumentando assim a precisão das respostas fornecidas pelos *chatbots*.

Desta forma as pesquisas futuras poderiam se concentrar na personalização da experiência do usuário com os *chatbots* na recuperação da informação. Isso envolveria a adaptação do estilo de interação do *chatbot*, a personalização das respostas com base nas preferências e histórico de busca do usuário, bem como a incorporação de mecanismos de recomendação personalizada de conteúdo relevante nos repositórios institucionais.

## REFERÊNCIAS

- ADAMOPOULOU, E.; MOUSSIADES, L. **Chatbots: History, technology, and applications**. Machine Learning with Applications. v. 2, 2020. <https://doi.org/10.1016/j.mlwa.2020.100006>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666827020300062>. Acesso em: 20 abr. 2022.
- AQUINO, Maria Clara. A folksonomia como hipertexto potencializador de memória coletiva: um estudo dos links e das tags no de.licio.us e no Flickr. **Liinc em Revista**. Rio de Janeiro, v. 4, n. 2, 2008. Disponível em: <https://revista.ibict.br/liinc/article/view/3159/2828>. Acesso em: 06 jul. 2023.
- ALI, M. Y.; NAEEM, S. B.; BHATTI, R. Artificial intelligence tools and perspectives of university librarians: an overview. **Business Information Review**, v. 37, n. 3, p. 116-124, 2020. Disponível em: <https://doi.org/10.1177/0266382120952016>. Acesso em: 26 jun. 2023.
- ALLEN, J. **Natural Language Understanding**. 2. ed. Redwood City, CA: The Benjamin/Cummings Pub. Co., 1995.
- ANELLI, V. W. et. al. Anna: a Virtual Assistant to interact with Puglia Digital Library (Discussion Paper). **SEBD**, Itália, jun. p. 16-19, 2019. Disponível em: <https://ceur-ws.org/Vol-2400/paper-30.pdf>. Acesso em: 26 jun. 2023.
- ARAÚJO JÚNIOR, Rogério Henrique de. **Precisão no processo de busca e recuperação da informação**. Brasília: Thesaurus, 2007.
- ARAKAKI, F. A.; SIMIONATO, A. C.; SANTOS, P. L. V. A. C. Catalogação e tecnologia: interseções com a Web Semântica. **Inf. Prof.**, Londrina, v. 6, n. 2, p. 03 – 19, jul./dez. 2017. Disponível em: <http://dx.doi.org/10.5433/2317-4390.2017v6n2p03>. Acesso em: 11 maio 2022.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676: Método para Análise de Documentos – Determinação de Seus Assuntos e Seleção de Termos de Indexação**. Rio de Janeiro, 1992.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação da informação: conceitos e tecnologias das máquinas de busca**. 2. ed. Porto Alegre: Bookman, 2013.
- BAGCHI, M. Conceptualising a Library Chatbot using Open Source Conversational Artificial Intelligence. **Journal of Library & Information Technology**, Índia, v. 40, n. 6, p. 329-333, nov. 2020. Disponível em: <https://drdo.gov.in/drdo/sites/default/files/inline-files/document3.pdf>. Acesso em: 26 jun. 2023.
- BARDAN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2016.
- BARRETO, A. A. Uma história da ciência da informação. In: Lidia Maria Batista

Brandão Toutain. (Org.). **Para entender a ciência da informação**. Salvador: EDUFBA, 2007. p. 13-34. Disponível em: <https://repositorio.ufba.br/bitstream/ufba/145/1/Para%20entender%20a%20ciencia%20da%20informacao.pdf>. Acesso em: 07 jul. 2023.

BATISTA, C. P. **A utilização de chatbot no atendimento em suporte técnico de primeiro nível**. 2021. 85 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Sistemas de Informação e Gestão do Conhecimento da Universidade FUMEC, Belo Horizonte. 2020. Disponível em: <https://repositorio.fumec.br/xmlui/handle/123456789/913>. Acesso em: 26 jun. 2023.

CARVALHO, R. C. de. **Chatbot aplicado à recuperação de informação: um modelo orientado a metadados**. Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências Marília, 2022. 214 f. Marília. 2022. Disponível em: <https://repositorio.unesp.br/server/api/core/bitstreams/68a4207e-25d6-4c9d-a4bb-0177d817f045/content>. Acesso em: 26 jun. 2023.

CHOUKIMATH, P. A.; SHIVARAMA, J.; GUJRAL, G. Perceptions and prospects of artificial intelligence technologies for academic libraries: an overview of global trends. **12º CALIBRE Internacional-2019**, KIIT, Bhubaneswar, Odisha, 28 a 30 de novembro de 2019. Disponível em: <https://ir.inflibnet.ac.in/handle/1944/2337>. Acesso em: 26 jun. 2023.

CONEGLIAN, C. S. **Recuperação da informação com abordagem semântica utilizando linguagem natural: a inteligência artificial na Ciência da Informação**. 2020. Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, Marília, 2020. Disponível em: [https://repositorio.unesp.br/bitstream/handle/11449/193051/coneglian\\_cs\\_dr\\_mar.pdf?sequence=3&isAllowed=y](https://repositorio.unesp.br/bitstream/handle/11449/193051/coneglian_cs_dr_mar.pdf?sequence=3&isAllowed=y). Acesso em: 07 jun. 2023.

CORRÊA, R. F.; SANTOS, R. F. dos. Análise das definições de folksonomia: em busca de uma síntese. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.23, n.2, p.1-32, abr./jun. 2018. Disponível em: <https://scielo.br/j/pci/a/wq76G649MfqdWHWtQkwwgGB/?format=pdf&lang=pt>. Acesso em: 14 fev. 2024.

COZMAN, F. G.; NERI, H. O que, afinal, é Inteligência Artificial? In: COZMAN, F. G.; PLONSKI, G. A.; NERI, H. (Org.) **Inteligência artificial: avanços e tendências**. São Paulo: Instituto de Estudos Avançados, 2021. Disponível em: <https://www.livrosabertos.abcd.usp.br/portaldelivrosUSP/catalog/download/650/579/2181?inline=1>. Acesso em: 13 fev. 2024.

CROW, R. The Case for Institutional Repositories: A SPARC Position Paper, **The Scholarly Publishing & Academic Resources Coalition**. Washington: SPARC, ago. 2002. Disponível em: [https://ils.unc.edu/courses/2014\\_fall/inls690\\_109/Readings/Crow2002-CaseforInstitutionalRepositoriesSPARCPaper.pdf](https://ils.unc.edu/courses/2014_fall/inls690_109/Readings/Crow2002-CaseforInstitutionalRepositoriesSPARCPaper.pdf). Acesso em: 16 fev. 2024.

CROWSTON, Kevin; ALLEN, Eileen E.; HECKMAN, Robert. Using natural language processing technology for qualitative data analysis. **International Journal of Social**

**Research Methodology**, London, v. 15, n. 6, p. 523-543, 2012. Disponível em: <https://doi.org/10.1080/13645579.2011.625764>. Acesso em: 13 fev. 2024.

DRAKE, M. Institutional Repositories: Hidden Treasures. **Searcher**, v. 12, n. 5, may. 2004. Disponível em: <http://www.infotoday.com/SEARCHER/may04/drake.shtml>. Acesso em: 16 fev. 2024.

ERICKSON, M.; KIM, P. Can chatbots work well with knowledge management systems? **Issues In Information Systems**, v. 21, n. 4, p. 53-58, 2020. Disponível em: [https://iacis.org/iis/2020/4\\_iis\\_2020\\_53-58.pdf](https://iacis.org/iis/2020/4_iis_2020_53-58.pdf). Acesso em: 26 jun. 2023.

EVENS, M. W. **Relational Model of the Lexicon**: Representing Knowledge in semantic networks. New York: Cambridge University Press, 1992.

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://www.scielo.br/j/pci/a/rPpchWXW8kKL8tYQ36tJH4w/?format=pdf&lang=pt>. Acesso em: 22 maio 2023.

FUJITA, M. S. L. Política de indexação para bibliotecas: funções e finalidades. In: FUJITA, M. S. L (Org.). **Política de indexação para bibliotecas**: elaboração, avaliação e implantação. Marília: Oficina Universitária, 2016. p. 15-19. Disponível em: [https://www.marilia.unesp.br/Home/Publicacoes/politicas-de-indexacao-para-bibliotecas\\_ebook.pdf](https://www.marilia.unesp.br/Home/Publicacoes/politicas-de-indexacao-para-bibliotecas_ebook.pdf). Acesso em: 20 fev. 2024.

FUJITA, M. S. L.; RIBAS, R.; RODRIGUES, M.; SILVEIRA, T. Metadados de assuntos de teses e dissertações em repositórios: estudo exploratório sobre controle de vocabulário. **Em Questão**, Porto Alegre, v. 28, n. 4, p. 123710, 2022. DOI: 10.19132/1808-5245284.123710. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/123710>. Acesso em: 22 maio 2023.

FUJITA, M. S. L.; RUBI, M. P.; BOCCATO, V. R. C. O contexto sociocognitivo do catalogador em bibliotecas universitárias: perspectivas para uma política de tratamento da informação documentária. **DataGramZero**, v. 10, n. 2, 2009. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/6489>. Acesso em: 20 fev. 2024.

FUJITA, M. S. L.; SANTOS, L. B. P. D. Política de indexação em bibliotecas universitárias: estudo diagnóstico e analítico com pesquisa participante. **Transinformação**, v. 28, n. 1, p. 59-76, 2016. Disponível em: <https://www.scielo.br/j/tinf/a/K6wKdXKG3pbPVkcLChr7FtL/?lang=pt&format=pdf>. Acesso em: 20 fev. 2024.

FUJITA, M.; TARTAROTTI, M. Análise de palavras-chave da produção científica de pesquisadores: o autor como indexador. **Informação & Informação**, Londrina, v. 25, n. 3, p. 332-374, 2020. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/41866/pdf>. Acesso em: 15 nov. 2023

FUJITA, M. S. L.; TOLARE, J. B. Vocabulários controlados na representação e recuperação da informação em repositórios brasileiros. **Informação & Informação**, Londrina, v. 24, n. 2, p. 93-125, 2019. Disponível em: <https://doi.org/10.5433/1981-8920.2019v24n2p93>. Acesso em: 04 jul. 2023.

FUSCO, E. **Modelos conceituais de dados como parte do processo da catalogação**: perspectiva de uso dos FRBR no desenvolvimento de catálogos bibliográficos digitais. 2010. 249f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2010. Disponível em: [https://bdtd.ibict.br/vufind/Record/UNSP\\_b78ec55c8f43e99cba493e395c228722](https://bdtd.ibict.br/vufind/Record/UNSP_b78ec55c8f43e99cba493e395c228722). Acesso em: 19 fev. 2024.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo : Atlas, 2002.

GIL LEIVA, I. Aspectos conceituais da indexação. In: FUJITA, M. S. L.; GIL LEIVA, I. G. (ed.). **Política de Indexação**. São Paulo: Cultura Acadêmica, p. 31-106, 2012. Disponível em: [https://www.marilia.unesp.br/Home/Publicacoes/politica-de-indexacao\\_ebook.pdf](https://www.marilia.unesp.br/Home/Publicacoes/politica-de-indexacao_ebook.pdf). Acesso em: 20 fev. 2024.

GOMES, R. F. **A relevância da política de indexação para a representação e recuperação da informação**: uma revisão sistemática. 2019. Dissertação (Mestrado) – Programa de Pós Graduação em Gestão e Organização do Conhecimento, Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: <https://repositorio.ufmg.br/handle/1843/34289>. Acesso em: 19 fev. 2024.

GUIMARÃES, J. A. C. As políticas de indexação como elemento para a gestão do conhecimento nas organizações. In: VIDOTTI, S. A. B. G. (Coord.). **Tecnologia e conteúdos informacionais**: abordagens teóricas e práticas. São Paulo: Polis, 2004. Disponível em: <https://abecin.org.br/wp-content/uploads/2021/03/Tecnologia-e-conteudos.pdf>. Acesso em: 19 fev. 2024.

GUIMARÃES, L. J. B. L. S. **Chatbot em contexto**: design de experiência do usuário aplicado à recuperação da informação no catálogo de teses e dissertações da CAPES. 2022. Tese (doutorado) - Universidade Federal de Minas Gerais, Escola de Ciência da Informação, Belo Horizonte, 2022. Disponível em: <https://repositorio.ufmg.br/handle/1843/50840>. Acesso em: 26 jun. 2023.

GUNDAKANAL, S. S.; KADDIPUJAR, M. Intelligent Libraries: new horizons with Artificial Intelligence. **Journal of Information Management and Educational Technology**. Índia, v. 3. n. 2/3, p. 8-17, aug/dec. 2019. Disponível em: [http://dspace.rii.res.in/bitstream/2289/7478/1/2019\\_JIMET\\_V3\\_p8-17.pdf](http://dspace.rii.res.in/bitstream/2289/7478/1/2019_JIMET_V3_p8-17.pdf). Acesso em: 26 jun. 2023.

HOCKX-YU, H. Digital preservation in the context of institutional repositories, **Program: Electronic Library and Information Systems**, Reino Unido, v. 40, n. 3, p. 232-243, jul. 2006. Disponível em: <https://doi.org/10.1108/00330330610681312>. Acesso em: 16 fev. 2024.

IDEMUDIA, B. E.; MAKINDE, B. Artificial Intelligence in Libraries: Prospect and Challenges for Nigerian Academic Libraries. **Journal of Library and Information Science**, Nigeria, v. 24, n. 2, dec. 2022. Disponível em: <https://cjol.org/index.php/cjol/article/view/11>. Acesso em: 26 jun. 2023.

INGWERSEN, P. **Information retrieval interaction**. Los Angeles: Taylor Graham Publishing, 1992.

JELUSIC, S.; STRICEVIC, I. **A Librarian's Guide on How to Publish**. Sawston: Chandos Publishing, 2011.

LAKATOS, E. M.; MARCONI, M. de A. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas 2003.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2.ed. Brasília: Briquet de Lemos, 2004.

LANCASTER, F. W. **Information retrieval systems**. 2. ed. New York: Wiley, 1978.

LANCASTER, F. W. **Vocabulary control for information retrieval**. 2. ed. Arlington: Information Resources Press, 1986.

LEITE, F. C. L.; COSTA, S. Repositórios institucionais como ferramentas de gestão do conhecimento científico no ambiente acadêmico. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.11, n.2, p. 206-219, mai./ago. 2006. Disponível em: <https://www.scielo.br/j/pci/a/xHsy3pkHDq3w6Sm3PLvPRVL/?format=pdf&lang=pt>. Acesso em: 05 maio 2022.

LEVY, P. **As tecnologias da inteligência: o futuro do pensamento na era da informática**. 2. Ed. São Paulo: Editora 34, 2010.

LIMA, C. O. de A. **A influência dos bots em processos informacionais: limitações, benefícios e desdobramentos**. 2022. Tese (doutorado) -Universidade Federal de Pernambuco, Centro de Artes e Comunicação, Departamento De Ciência Da Informação, Recife, 2022. Disponível em: <https://repositorio.ufpe.br/handle/123456789/49148>. Acesso em: 19 fev. 2024.

LIMA, G. N. B. O.; FUJITA, M. S. L.; REDIGOLO, F. M. A importância do contexto para a indexação. **Ponto de Acesso**, v. 15, n. 3, 2021. Disponível em: <https://doi.org/10.9771/rpa.v15i3.47469>. Acesso em: 26 jun. 2023.

LOPES, I. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ciência da Informação**, Belo Horizonte, v. 31, n. 1, p. 41-52, 2002. Disponível em: <https://revista.ibict.br/ciinf/article/%20view/976>. Acesso em: 15 nov. 2023.

LOPEZ YEPEZ, J. (ED.). **Fundamentos de informação e documentação**. Madrid: EDEMA, 1989. Recensão de: SILVA, L. A. G. da. **Ci. Inf.**, Brasília, v. 20, n. 1, p. 95-

97, jan./jun. 1991. Disponível em: <https://revista.ibict.br/ciinf/article/view/426/426>. Acesso em: 15 nov. 2023.

LYNCH, C. A. Institutional Repositories: essential infrastructure for scholarship in the Digital Age. **ARL Bimonthly Report**, n. 226, p. 1-7, feb. 2003. Disponível em: <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>. Acesso em: 16 fev. 2024.

MCKIE, I. A. S.; NARAYAN, B. Enhancing the Academic Library Experience with Chatbots: An Exploration of Research and Implications for Practice. **Journal of the Australian Library and Information Association**, Londres, v. 68, n. 3, jul. p. 268-277, 2019. Disponível em: <https://doi.org/10.1080/24750158.2019.1611694>. Acesso em: 26 jun. 2023.

MANNING, C. D.; RAGHAVAN, P.; SCHUTZE, H. **Introduction to information retrieval**. Cambridge: Cambridge University Press, 2008.

MARCIAL, V. F., GOMES, L. I. E.; MARQUES, M. B. Perspetiva teórica e metodológica em sistemas de informação complexos. **Páginas a&b: Arquivos & Bibliotecas**, Porto, v. 3, n. 4, p. 3-21, 2015. Disponível em: <https://ojs.letras.up.pt/index.php/paginasaeb/article/view/995>. Acesso em: 09 fev. 2024.

MARTINEZ, A. R. Natural language processing. **Wiley Interdisciplinary Reviews: Computational Statistics**, Hoboken, v. 2, n. 3, p. 352-357, 2010. Disponível em: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wics.76>. Acesso em: 09 fev. 2024.

MARTINS, A. L. Potenciais aplicações da Inteligência Artificial na Ciência da Informação. *Inf. Inf.*, Londrina, v. 15, n. 1, p. 1 - 16, jul./jun. 2010. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/3882>. Acesso em: 09 fev. 2024.

MATHES, A. **Folksonomies**: cooperative classification and communication through shared metadata. 2004. Disponível em: <https://adammathes.com/academic/computer-mediated-communication/folksonomies.html>. Acesso em: 16 nov. 2023.

MUKHERJEE, S.; PATRA, S. K. Chatbots: a review of their potential applications in library services. **Qeios**, Itália, maio 2023. Disponível em: <https://doi.org/10.32388/R3USN5>. Acesso em 19 set. 2023.

MIZZARO, S. How many relevances in information retrieval? **Interacting with Computers**, Reino Unido, v.10, n.3, jun.1998, p.303-320. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0953543898000125#>. Acesso em: 13 out. 2020.

MONTEREI, R. C. **Perspectivas do uso do aprendizado de máquina em bibliotecas**: uma revisão sistemática de literatura. 2022. 145f. Dissertação (Mestrado) - Faculdade de Ciência da Informação da Universidade de Brasília.

Brasília. 2022. Disponível em: <http://icts.unb.br/jspui/handle/10482/45416>. Acesso em: 26 jun. 2023.

MOOERS, C.N. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, v. 2, n. 1 p. 20-32, 1951. Disponível em: <https://doi.org/10.1002/asi.5090020107>. Acesso em: 20 set. 2023.

MOURA, M. A.; SILVA, A. P.; AMORIM, V. R. (2002). A concepção e o uso das linguagens de indexação face às contribuições da semiótica e da semiologia. **Informação & Sociedade**, João Pessoa, v. 12, n. 1. 2002. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/160>. Acesso em: 18 fev. 2024.

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. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/50756>. Acesso em: 22 abr. 2023.

NUNES, C. O. I. Algumas considerações acerca da ausência de políticas de indexação em bibliotecas brasileiras. **BIBLOS - Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 16, p. 55-61, 2004. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/22625>. Acesso em: 20 fev. 2024.

PEREIRA, S. L. **Introdução à Inteligência Artificial**. São Paulo: IME-USP, 2007. Disponível em: <https://www.ime.usp.br/~slago/IA-introducao.pdf>. Acesso em: 13 fev. 2024.

PINKOSOVA, Z.; MCGEOWN, W. J.; MOSHFEGHI, Y. Moderating effects of self-perceived knowledge in a relevance assessment task: An EEG study, **Computers in Human Behavior Reports**, Canadá, v. 11, ago. 2023. Disponível em: <https://doi.org/10.1016/j.chbr.2023.100295>. Acesso em: 01 mar. 2024.

PINTO, V. B. Indexação documentária: uma forma de representação do conhecimento registrado. **Revista de Letras**, [S. l.], v. 1, n. 22, 2016. Disponível em: <http://periodicos.ufc.br/revletras/article/view/2175>. Acesso em: 20 fev. 2024.

RUBI, M. P. Os princípios da política de indexação na análise de assunto para catalogação: especificidade, exaustividade, revocação e precisão na perspectiva dos catalogadores e usuários. In: FUJITA, M. S. L. et al (org.). **A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias: um estudo de observação do contexto sociocognitivo com protocolos verbais**. São Paulo: Cultura Acadêmica, 2009. Disponível em: <https://static.scielo.org/scielobooks/wcvbc/pdf/boccatto-9788579830150.pdf>. Acesso em: 20 fev. 2024.

RUBI, M. P. Política de Indexação. In: GIL LEIVA, I.; FUJITA, M. S. L. (ed.). **Política de Indexação**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012. Disponível em: [https://www.marilia.unesp.br/Home/Publicacoes/politica-de-indexacao\\_ebook.pdf](https://www.marilia.unesp.br/Home/Publicacoes/politica-de-indexacao_ebook.pdf). Acesso em: 20 fev. 2024.

RUBIN, R. E.; RUBIN, R. G. **Foundations of library and information Science**. 5. ed. Chicago: ALA Neal-Schuman, 2020.

RUMSEY, S. The purpose of institutional repositories in UK higher education: a repository manager's view. **International Journal of Information Management**, Reino Unido, v. 26, n. 3, p. 181-186, 2006. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0268401206000041>. Acesso em: 16 fev. 2024.

RUSSELL, S. J.; NORVIG, P. **Artificial intelligence: a modern approach**. Malaysia: Pearson, 2016.

SALCEDO, D. A.; BEZERRA, V. C. A. Encontro e descoberta da informação em ambientes digitais. **Páginas a&b: Arquivos & Bibliotecas**, Porto, n.13, p.142-155, 2020. Disponível em: <https://doi.org/10.21747/21836671/pag13a10>. Acesso em: 18 fev. 2024.

SANTARÉM SEGUNDO, J. E. **Representação Iterativa: um modelo para repositórios digitais**. 2010. 224 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2010. Disponível em: <https://repositorio.unesp.br/handle/11449/103346>. Acesso em: 26 jun. 2023.

SANTAREM SEGUNDO, J. E.; VIDOTTI, S. A. B. G. Representação Iterativa e folksonomia assistida para repositórios digitais. **Liinc em Revista**, Rio de Janeiro, v.7, n.1, mar. p. 283-300, 2011. Disponível em: <https://revista.ibict.br/liinc/article/view/3300/2916>. Acesso em: 27 fev. 2024.

SANTOS, H. P. Etiquetagem e Folksonomia: o usuário e sua motivação para organizar e compartilhar informação na web 2.0. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.1, n.2, p.91-104, abr./jun. 2013. Disponível em: <https://www.scielo.br/j/pci/a/Cy4m3vbMRQvWSCRQMr9B49k/?format=pdf&lang=pt>. Acesso em: 27 fev. 2024.

SANTOS, R. F. D.; CORRÊA, R. F. A folksonomia e a representação colaborativa da informação em ambientes digitais. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 8, n. 1, 2015. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/119561>. Acesso em: 06 jul. 2023.

SANTOS, R. F. dos; ALBUQUERQUE, M. E. B. C. de. **PontodeAcesso**, Salvador, v. 15, n. 3, p. 381-397, dez. 2021. Disponível em: DOI: <https://doi.org/10.9771/rpa.v15i3.47473>. Acesso em: 27 fev. 2024.

SANTOS, R. F. D.; CARVALHO, T. F.; RODRIGUES, K. O. Folksonomia em repositórios digitais: análise da produção científica nacional e internacional = folksonomy in digital repositories: analysis of national and international scientific production. **Revista Bibliomar**, São Luís, v. 20, n. 1, p. 57-79, 2021. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/161109>. Acesso em: 08 ago. 2023.

SANTOS, R. F. D. Indexação em repositórios digitais: uma abordagem sobre o metadado assunto da biblioteca digital de monografias da UFRN. **Revista Informação na Sociedade Contemporânea**, Natal, Número especial, p. 1-22, 2017. Disponível em: <https://doi.org/10.21680/2447-0198.2017v1n0ID12279>. Acesso em: 16 nov. 2023.

SARACEVIC, T.; EMPTY, Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, 1996. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/22308>. Acesso em: 06 jul. 2023.

SARACEVIC, T. Relevance: A review of the literature and a framework for thinking on the notion in information Science. Part II: nature and manifestations of relevance. **Journal of the American Society for Information Science and Technology**, Romania, v. 58, n. 13, p. 1915–1933, 2007. Disponível em: <https://doi.org/10.1002/asi.20682>. Acesso em: 14 jul. 2023.

SAYÃO, L. F. O papel dos repositórios digitais na curadoria de dados de pesquisa. In: VECHIATO, F. et al. (Orgs.) **Repositórios digitais: teoria e prática**. Curitiba: EDUTFPR, 2017. p. 145-165. Disponível em: <file:///C:/Users/Samara/Downloads/repositoriosdigitasteoriapratica.pdf>. Acesso em: 14 jul. 2023.

SILVA, R. **Batendo um papo com a informação**: o uso dos chatbots para a recuperação da informação e a contribuição da Ciência da Informação nesse processo. 2020. 125 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Ciência da Informação - Escola de Comunicações e Artes / Universidade de São Paulo. São Paulo. 2020. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27163/tde-10032021-013140/en.php>. Acesso em: 26 jun. 2023.

SILVA, R. E. da; SANTOS, P. L. V. A. da C.; FERNEDA, E. Modelos de recuperação de informação e web semântica: a questão da relevância. **Informação & Informação**, Londrina, v. 18, n. 3, p. 27-44, 2013. Disponível em: <https://doi.org/10.5433/1981-8920.2013v18n3p27>. Acesso em: 16 fev. 2024.

SILVA, R. C; BRITO, J. F. Proposta de um manual de indexação para bibliotecas universitárias. **Informação@Profissões**, v. 7, n. 1, jan./jun. p. 92-113, 2018. Disponível em: <https://doi.org/10.5433/2317-4390.2018v7n1p92>. Acesso em 18 fev. 2024.

SOPRANO, M. et al. Cognitive biases in fact-checking and their countermeasures: a review. *Information Processing & Management*, v. 61, n. 3, 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0306457324000323>. Acesso em 1 abr. 2024.

STRASSMANN, P. A. **The politics of information management**: policy guidelines. Connecticut: Information Economics Press, 1994.

TELLA, A.; AJANI, Y. A. Robots and public libraries. **Library Hi Tech News**. v. 39, n. 7, p.15-18, jul. 2022. Disponível em: <https://doi.org/10.1108/LHTN-05-2022-0072>. Acesso em: 26 jun. 2023.

TSENG, Yuen-Hsien et al. Implementation and Evaluation of a Retrieval-based Chinese Humor Chatbot. **Journal of Library and Information Science**, v. 18, n.2, 2020. Disponível em: <https://jlis.lis.ntu.edu.tw/files/journal/j51-4.pdf>. Acesso em: 14 mar. 2023.

TSIMARA, I.; OIKONOMOU, A.; KATSIRIKOU, A. Applications of Artificial Intelligence Technology in the Field of Scientific Information. **Qualitative and Quantitative Methods in Libraries (QQML)**, v. 12, n.1, p. 153-166, 2023. Disponível em: [file:///C:/Users/Samara/Downloads/1219QQML\\_Journal\\_2023\\_153+-166%20\(1\).pdf](file:///C:/Users/Samara/Downloads/1219QQML_Journal_2023_153+-166%20(1).pdf). Acesso em: 26 jun. 2023.

VAJJALA S. et al. **Practical natural language processing**. Sebastopol, CA: O'Reilly Media, Inc., 2020.

VIERA, A. F. G.; GARRIDO, I. dos S. Folksonomia como uma estratégia para recuperação colaborativa da informação. **DataGramZero - Revista de Ciência da Informação**, Rio de Janeiro, v. 12, n. 2, abr. 2011. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/7336>. Acesso em: 16 nov. 2023.

WARNER, S.; BEKAERT, J.; LAGOZE, J.; LIU, X.; PAYETTE, S.; VAN DE SOMPEL, H. Pathways: Augmenting interoperability across scholarly repositories. **International Journal on Digital Libraries**, [S. l.], v. 1, n. 5, out. 2005. Disponível em: <https://arxiv.org/pdf/cs/0610031.pdf>. Acesso em: 16 fev. 2024.

ZEROUAL, Imad; LAKHOAJA, Abdelhak. Data science in light of natural language processing: an overview. **Procedia Computer Science**, [S. l.], v. 127, p. 82–91, 2018. Disponível em: <https://doi.org/10.1016/j.procs.2018.01.101>. Acesso em: 16 fev. 2024.

## APÊNDICE A – Levantamento e análise dos resultados

Ano	Título do Artigo	Problemática de pesquisa	Objetivos	Resultados	Aplicações em repositórios Institucionais
2019	A Customised Grammar Framework for Query Classification	<p>A pesquisa aborda a problemática da classificação de consultas em categorias específicas, visando aprimorar o desempenho por meio de um novo framework de gramática personalizado. Esse framework supera as limitações dos métodos convencionais, como a abordagem de "saco de palavras", ao considerar a estrutura sintática do texto. O objetivo é investigar como a informação prévia sobre o problema e o conhecimento especializado do domínio podem contribuir para soluções confiáveis em problemas de classificação de texto, especialmente em contextos ambíguos, como consultas de busca na web.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p>	<p>A pesquisa tem como objetivo desenvolver e aplicar um framework de gramática personalizado para a classificação de consultas de acordo com a intenção do usuário, utilizando as categorias propostas por Broder. Os objetivos específicos incluem investigar o impacto dos diferentes níveis de detalhe das informações do domínio na performance de classificação, comparar algoritmos de aprendizado de máquina para essa tarefa e avaliar o desempenho da abordagem proposta em relação a outras abordagens do estado da arte. A pesquisa busca demonstrar como a estrutura da consulta e as categorias sintáticas do domínio podem influenciar positivamente a precisão da classificação de consultas.</p>	<p>Os resultados da pesquisa indicam que o framework de gramática personalizado proposto para a classificação de consultas com base na intenção do usuário superou abordagens tradicionais, como bag-of-words e dicionários/lexicons. A inclusão da estrutura sintática do texto, aliada a categorias sintáticas específicas do domínio, resultou em uma classificação mais precisa e eficaz das consultas. A comparação entre diferentes algoritmos de aprendizado de máquina para a classificação da intenção do usuário destacou a eficácia do framework em relação a outras abordagens do estado da arte. Esses resultados ressaltam a importância de considerar a estrutura do texto e as características específicas do domínio para aprimorar a classificação de consultas, contribuindo para a relevância dos resultados de busca para os usuários.</p>	<p>A aplicação da pesquisa em repositórios institucionais pode otimizar a organização e recuperação de informações. A implementação do framework de gramática personalizado na classificação de consultas melhora a precisão na identificação da intenção do usuário, tornando o sistema mais eficiente e relevante. A consideração da estrutura sintática do texto e categorias sintáticas específicas do domínio contribui para uma classificação refinada, facilitando a recuperação de informações em ambientes desafiadores. A comparação de algoritmos de aprendizado de máquina possibilita a escolha da abordagem mais eficaz, promovendo a usabilidade e utilidade dos repositórios institucionais na disseminação e acesso a informações acadêmicas e científicas.</p>

2019	A multi-approach to community question answering	<p><b>Subcategoria:</b> Interação e IA</p> <p>Este estudo aborda os desafios enfrentados na resposta eficaz a perguntas em um contexto de comunidade, especialmente na Community Question Answering (CQA) para a língua árabe. Identificando problemas como desigualdade nos conjuntos de dados, extensão das perguntas e respostas, tópicos técnicos, escassez de recursos linguísticos para o árabe e a presença de palavras latinas, transliterações e ruídos. Essas questões resultam em baixos valores de mAP e precisão em comparação com a CQA em inglês. Além disso, destaca a complexidade frequente das perguntas, a natureza técnica dos tópicos, respostas extensas com possíveis referências, e o desafio adicional de respostas de spam que introduzem informações irrelevantes. Para a língua árabe, são mencionados obstáculos como textos não vocalizados, morfologia complexa e falta de recursos linguísticos. O estudo ressalta a necessidade de abordagens inovadoras para aprimorar a qualidade e eficácia das respostas na CQA em árabe.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de</p>	<p>O objetivo da pesquisa é desenvolver e avaliar um conjunto de modelos e abordagens para melhorar a precisão e eficácia na resposta a perguntas em um ambiente de comunidade, focando especificamente na Community Question Answering (CQA) para a língua árabe. Os autores buscam aprimorar a seleção, pontuação e classificação de respostas dentro de threads de texto em linguagem natural, utilizando técnicas de aprendizado profundo e transfer learning.</p>	<p>Os resultados da pesquisa demonstraram que o modelo baseado na abordagem de Indexação Semântica Latente superou os modelos de ponta existentes para a tarefa de Community Question Answering (CQA) em língua árabe. Além disso, o uso de transfer learning para alimentar a camada de embeddings de vários modelos de deep learning mostrou que abordagens em pares superam seus equivalentes pontuais. Todos os modelos propostos foram avaliados na tarefa Semeval 2017 Subtask D e alcançaram desempenho de ponta ou próximo a ele.</p>	<p>Os modelos e abordagens desenvolvidos podem ser utilizados para melhorar a eficiência na busca e recuperação de informações dentro desses repositórios. Ao implementar sistemas de Community Question Answering (CQA), as instituições podem facilitar a interação dos usuários com o conteúdo dos repositórios, permitindo que façam perguntas e recebam respostas relevantes de forma mais rápida e eficaz. Isso pode contribuir para a disseminação do conhecimento e a colaboração entre os membros da comunidade acadêmica que utilizam os repositórios institucionais.</p>
------	--	--	--	---	---

		<p>informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>			
<p>2019</p>	<p>A New Digital Signal Processing Based Model With Multi-Aspect Term Frequency for Information Retrieval</p>	<p>A pesquisa aborda a necessidade de explorar novas categorias de frameworks e modelos para a recuperação de informações. Embora existam modelos bem-sucedidos na área de recuperação de informações, como os modelos de espaço vetorial, modelos probabilísticos e modelos estatísticos de linguagem, há uma demanda por abordagens inovadoras que possam melhorar o desempenho dos sistemas de recuperação de informações. Assim, o estudo busca investigar e desenvolver um novo modelo baseado em processamento digital de sinais com foco na frequência de termos em múltiplos aspectos para aprimorar a recuperação de informações.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo da pesquisa é propor e desenvolver um novo modelo de recuperação de informações baseado em processamento digital de sinais, denominado DSP-MATF, que incorpora um método eficaz de ponderação de termos do modelo de Frequência de Termos em Múltiplos Aspectos (MATF) do Modelo de Espaço Vetorial (VSM), com o intuito de aprimorar a eficiência e precisão dos sistemas de recuperação de informações.</p> <p>O modelo DSP-MATF é um modelo de recuperação de informações baseado em processamento digital de sinais que incorpora um método eficaz de ponderação de termos do modelo de Frequência de Termos em Múltiplos Aspectos (MATF) do Modelo de Espaço Vetorial (VSM). Ele representa cada termo de consulta como diferentes categorias de sinais envelopados por sete funções de kernel, visando melhorar a eficiência e precisão dos sistemas de recuperação de informações.</p>	<p>O resultado da pesquisa foi a proposição e desenvolvimento de um novo modelo de recuperação de informações baseado em processamento digital de sinais, o DSP-MATF. Este modelo incorpora um método eficaz de ponderação de termos do modelo MATF do VSM e representa cada termo de consulta como diferentes categorias de sinais envelopados por sete funções de kernel. Os experimentos realizados demonstraram que o DSP-MATF superou os modelos de referência em termos de métricas variadas, como MAP, P@10 e P@20, evidenciando sua eficácia e vantagens em relação aos modelos existentes.</p>	<p>Ao utilizar o DSP-MATF, os repositórios institucionais podem aprimorar a precisão e eficiência na busca e recuperação de documentos acadêmicos, pesquisas científicas e outros tipos de conteúdo. Isso pode facilitar a localização de informações relevantes para pesquisadores, professores, estudantes e demais usuários dos repositórios, tornando a busca por conteúdo mais eficaz e satisfatória ao aplicar o modelo DSP-MATF em repositórios institucionais pode contribuir para a organização e categorização mais eficiente dos documentos, auxiliando na classificação e indexação adequada do conteúdo disponível. Isso pode resultar em uma experiência de busca mais intuitiva e personalizada para os usuários, promovendo a descoberta e acesso mais fácil a informações relevantes dentro do repositório.</p>

2019	A novel unsupervised corpus-based stemming technique using lexicon and corpus statistics	<p>O problema de pesquisa abordado neste artigo é a necessidade de desenvolver um método de stemming totalmente não supervisionado que possa lidar com diferentes idiomas e complexidades linguísticas. O objetivo é criar uma técnica que utilize tanto características lexicais quanto estatísticas do corpus para agrupar palavras morfologicamente relacionadas, visando melhorar o desempenho em tarefas de recuperação de informações e análise de texto.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo deste estudo é desenvolver um algoritmo de stemming multiuso que possa ser aplicado não apenas em tarefas de recuperação de informações, mas também em atividades não tradicionais, como classificação de texto, análise de sentimentos e remoção de inflexões. A proposta é utilizar características baseadas em lexicon e estatísticas do corpus para descobrir palavras morfologicamente relacionadas e, assim, melhorar o desempenho em diferentes idiomas, incluindo aqueles com morfologia complexa.</p>	<p>Os resultados da pesquisa indicam que o método de stemming proposto superou significativamente outros stemmers baseados em regras e em corpus em tarefas de recuperação de informações em quatro idiomas (Inglês, Marathi, Húngaro e Bengali) usando coleções de testes padrão. Houve melhorias notáveis na precisão média em todos os idiomas testados em comparação com a recuperação baseada em palavras e outros stemmers. Além disso, o algoritmo proposto também se destacou em tarefas de classificação de texto e remoção de inflexões, demonstrando sua eficácia como uma ferramenta multiuso em várias aplicações de processamento de linguagem natural.</p>	<p>O estudo propõe um método de stemming que pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações e a classificação de documentos. Ao utilizar características lexicon e estatísticas do corpus, o algoritmo pode identificar palavras morfologicamente relacionadas, facilitando a busca por conteúdo relevante e a organização eficiente dos documentos no repositório.</p>
2019	A Sequential Matching Framework for Multi-Turn Response Selection in Retrieval-Based Chatbots	<p>A problemática abordada neste estudo é a seleção de respostas em conversas de múltiplas rodadas em chatbots baseados em recuperação. O desafio consiste em combinar candidatos de resposta com o contexto da conversa, reconhecendo partes importantes do contexto e modelando as relações entre as falas ao longo da interação. Os métodos de correspondência existentes podem perder informações cruciais nos</p>	<p>Desenvolver um novo framework de correspondência sequencial para a seleção de respostas em conversas de múltiplas rodadas em chatbots baseados em recuperação. Este framework visa superar as limitações dos métodos existentes, permitindo que as informações importantes nos contextos das conversas sejam adequadamente consideradas durante a seleção de respostas e modelando as relações entre as falas ao longo da interação. Além disso, o</p>	<p>O estudo apresentou um novo framework de correspondência sequencial para a seleção de respostas em chatbots de múltiplas rodadas. Os modelos propostos, como a Rede Convolutacional Sequencial e a Rede de Atenção Sequencial, superaram significativamente os modelos existentes, alcançando um novo patamar de eficácia na correspondência entre contextos e respostas. A análise de ablação e visualização dos</p>	<p>O estudo sobre o framework de correspondência sequencial para seleção de respostas em chatbots de conversas de múltiplas rodadas pode ser aplicado em repositórios institucionais para melhorar a assistência ao usuário, a recuperação de informações, o suporte à pesquisa, o engajamento do usuário e a análise de dados. Essas aplicações visam aprimorar a interação dos usuários com os repositórios, oferecendo uma</p>

		<p>contextos das conversas, o que motivou a proposta de um novo framework de correspondência sequencial que visa capturar e utilizar de forma eficaz essas informações importantes para a seleção de respostas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>estudo busca melhorar a compreensão e a eficácia do processo de correspondência entre contextos e respostas em chatbots, com foco em capturar e alavancar informações relevantes nos contextos para a seleção de respostas</p>	<p>componentes dos modelos destacou sua capacidade de capturar e utilizar informações importantes nos contextos para a seleção de respostas, demonstrando tanto eficácia quanto interpretabilidade.</p>	<p>experiência mais personalizada, eficiente e interativa na busca e no acesso a informações relevantes.</p>
<p>2019</p>	<p>A universal information theoretic approach to the identification of stopwords</p>	<p>O documento aborda a problemática relacionada à identificação e remoção de palavras semanticamente não informativas, conhecidas como "stopwords", em corpora de texto. A falta de consenso sobre quais palavras devem ser excluídas como stopwords gera debates contínuos na comunidade de especialistas. A abordagem tradicional de listas de stopwords manualmente curadas apresenta desafios, pois não pode ser facilmente generalizada entre diferentes domínios de conhecimento ou idiomas. A escassez de estudos sistemáticos sobre o impacto da remoção de stopwords no desempenho de algoritmos destaca a necessidade de uma</p>	<p>O objetivo do documento é apresentar uma abordagem baseada em teoria da informação para a identificação de stopwords em corpora de texto. Essa abordagem visa superar as limitações das listas de stopwords manualmente curadas, oferecendo uma solução automatizada e mais eficaz para identificar palavras semanticamente não informativas. O objetivo é melhorar o desempenho de algoritmos de processamento de linguagem natural, como modelos de tópicos, ao remover stopwords de forma mais precisa e generalizável entre diferentes domínios e idiomas.</p>	<p>O resultado do estudo apresentado no documento demonstra que a abordagem baseada em teoria da informação para a identificação de stopwords supera outras heurísticas tradicionais de stopwords. A remoção de stopwords de acordo com essa abordagem resultou em uma redução substancial no tamanho do corpus, sem uma diminuição significativa na precisão, especialmente para frações menores de dados removidos. Além disso, a abordagem de stopwords baseada em informação teórica mostrou ser a mais eficaz em diferentes corpora de texto de diversos domínios e idiomas, destacando sua capacidade de</p>	<p>O estudo sobre identificação de stopwords com base em teoria da informação pode ser aplicado em repositórios institucionais para melhorar a qualidade dos dados, facilitar a recuperação da informação, classificar documentos, analisar tópicos e minerar dados. A remoção de stopwords pode aprimorar a organização dos conteúdos, a precisão da recuperação da informação e a compreensão dos padrões nos documentos, contribuindo para uma gestão mais eficaz e análises mais profundas nos repositórios institucionais.</p>

		<p>abordagem mais fundamentada e automatizada para identificar palavras não informativas em corpora de texto.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>		<p>generalização e desempenho superior em comparação com outras listas de stopwords tradicionais.</p>	
2019	<p>Accuracy evaluation of methods and techniques in Web-based question answering systems: a survey</p>	<p>A problemática abordada neste estudo está relacionada à avaliação da precisão de métodos e técnicas em sistemas de resposta a perguntas baseados na Web. Os pesquisadores buscam identificar quais métodos e técnicas são mais eficazes na melhoria da precisão das respostas fornecidas por esses sistemas. A questão central é como avaliar e comparar a precisão de diferentes abordagens em sistemas de resposta a perguntas, a fim de identificar as mais eficazes e informar o desenvolvimento futuro dessas tecnologias.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo deste estudo é avaliar os métodos e técnicas utilizados em sistemas de resposta a perguntas baseados na Web, visando melhorar a precisão das respostas fornecidas por esses sistemas. Os pesquisadores buscam identificar as combinações de métodos que resultam em maior precisão e eficácia nos sistemas de resposta a perguntas na Web. Além disso, o estudo tem como objetivo desenvolver um sistema de resposta a perguntas baseado em métodos e técnicas identificados como os mais eficazes, a fim de demonstrar a diferença entre a seleção dos melhores métodos e técnicas em comparação com outros métodos disponíveis.</p>	<p>Os resultados deste estudo indicam que a avaliação e comparação de métodos e técnicas em sistemas de resposta a perguntas baseados na Web são essenciais para melhorar a precisão das respostas fornecidas por esses sistemas. Identificar as melhores combinações de métodos e técnicas pode resultar em um aumento significativo na eficácia dos sistemas de resposta a perguntas. Além disso, o desenvolvimento de um sistema de resposta a perguntas baseado nos métodos e técnicas mais eficazes identificados no estudo demonstrou uma melhoria substancial em relação aos sistemas de linha de base existentes. Esses resultados destacam a importância da seleção cuidadosa de métodos e técnicas para melhorar a precisão e eficácia dos sistemas de resposta a perguntas na Web.</p>	<p>O estudo sobre a avaliação de métodos e técnicas em sistemas de resposta a perguntas baseados na Web pode ser aplicado em repositórios institucionais para melhorar a precisão das respostas, desenvolver sistemas de QA internos, selecionar métodos e técnicas adequados, realizar avaliações regulares e aprimorar a eficácia dos sistemas de busca e recuperação de informações. Essa aplicação dos resultados do estudo pode contribuir para oferecer aos usuários respostas mais precisas e eficientes nos repositórios institucionais.</p>

2019	<p>An Improved Word Representation for Deep Learning Based NER in Indian Languages</p>	<p>O problema central de pesquisa aborda aprimorar a representação de palavras para o Reconhecimento de Entidades Nomeadas (NER) em idiomas indianos. O foco é capturar informações sintáticas, semânticas e morfológicas de maneira eficaz, visando aprimorar a precisão e o desempenho geral do sistema de extração de entidades nesses idiomas. O problema específico considera a complexidade morfológica e aglutinativa dos idiomas indianos, juntamente com a falta de recursos disponíveis para o desenvolvimento de sistemas de NER eficazes. A pesquisa propõe uma solução baseada em técnicas de aprendizado profundo, utilizando uma representação combinada de palavras, incluindo embeddings em nível de caractere, palavra e afixo, para enfrentar os desafios únicos apresentados por esses idiomas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo do estudo é desenvolver um sistema de extração de entidades nomeadas (NER) mais eficaz para idiomas indianos, utilizando uma abordagem baseada em técnicas de aprendizado profundo. O objetivo principal é melhorar a representação de palavras para o NER em idiomas indianos, capturando informações sintáticas, semânticas e morfológicas de maneira mais precisa e abrangente.</p>	<p>Os resultados apresentados na pesquisa demonstraram uma melhoria significativa na precisão do NER em comparação com abordagens tradicionais, destacando a eficácia da representação de palavras aprimorada. A inclusão de representações em nível de caractere contribuiu para capturar informações sintáticas, semânticas e morfológicas, resultando em um desempenho superior na identificação de entidades como nome, localização, organização, evento, entre outros, em textos indianos. O modelo proposto superou modelos existentes em termos de acurácia e eficiência, validando a importância da abordagem de aprendizado profundo para o NER em idiomas indianos.</p>	<p>A proposta de utilizar embeddings em nível de caractere, palavra e afixo em repositórios institucionais visa otimizar a extração de informações de documentos acadêmicos. Essa abordagem permite a identificação eficaz de entidades nomeadas, como autores, instituições e locais, facilitando a recuperação de informações relevantes e a categorização de documentos. A aplicação de técnicas avançadas de Reconhecimento de Entidades Nomeadas (NER) pode aprimorar os sistemas de recomendação de conteúdo, tornando a organização e recuperação de informações mais eficientes e precisas em repositórios acadêmicos.</p>
------	--	---	--	---	--

2019	Automated feedback generation for formal manufacturing rule extraction	<p>O estudo aborda a complexidade da extração automatizada e precisa de regras de fabricação a partir de textos não estruturados. A validação cuidadosa do texto é crucial devido a possíveis informações irrelevantes ou expressões semanticamente inválidas que podem prejudicar a extração correta das regras desejadas. Dada a limitação das técnicas tradicionais de Processamento de Linguagem Natural (NLP) para lidar com essas complexidades, o objetivo da pesquisa é desenvolver um método que forneça feedback automatizado para a validação e melhoria dos textos de entrada. Isso visa facilitar a extração eficaz e precisa das regras de fabricação formais.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback</p>	<p>O estudo visa desenvolver um método que ofereça feedback automatizado para validar e aprimorar textos de entrada não estruturados, facilitando assim a extração precisa e eficaz de regras de fabricação formais. Em contraste com as técnicas tradicionais de Processamento de Linguagem Natural (NLP), esse método permite que os usuários modifiquem o texto de entrada sem a necessidade de aprender linguagens estruturadas formais, superando as limitações associadas a textos que demandam validação.</p>	<p>O estudo revelou a eficácia do método de geração automatizada de feedback para extração de regras de fabricação formal. Este método ajudou os usuários a identificar a necessidade de validação do texto de entrada, proporcionando feedback relevante para ajustes. Cerca de 30% dos casos que exigiam validação receberam feedback útil, resultando na extração correta das regras de fabricação desejadas após a modificação do texto. Todos os casos relevantes que receberam feedback alcançaram a extração correta das regras. Esses resultados destacam a utilidade e viabilidade do método proposto na validação e aprimoramento de textos para uma extração precisa de regras de fabricação formal.</p>	<p>O estudo sobre geração automatizada de feedback para extração de regras de fabricação formal pode ser aplicado em repositórios institucionais para auxiliar na validação e melhoria de textos não estruturados relacionados à fabricação. Ao utilizar o método proposto, os repositórios podem oferecer aos usuários uma ferramenta automatizada que fornece feedback relevante para aprimorar a qualidade e precisão dos textos armazenados. Isso facilita a extração correta das regras de fabricação, tornando a informação mais acessível e útil para os usuários que consultam o repositório em busca de conhecimento sobre processos de fabricação.</p>
2019	Avaliação das etapas de pré-processamento e de treinamento em algoritmos de classificação de textos no contexto da	<p>A problemática abordada no estudo é a dificuldade em eliminar as questões de aprendizado e compreensão do modelo para classificações de sentimentos que envolvem</p>	<p>O objetivo do estudo é avaliar quantitativamente as etapas de pré-processamento e treinamento de um classificador de textos no processo de classificação automática de</p>	<p>Os resultados do estudo indicaram que um tratamento mais refinado no vocabulário utilizado no pré-processamento dos textos melhorou significativamente os resultados</p>	<p>O estudo sobre a avaliação das etapas de pré-processamento e treinamento em algoritmos de classificação de textos pode ser aplicado em repositórios institucionais para melhorar a</p>

	<p>recuperação da informação</p>	<p>características subjetivas da expressão humana, como ironia ou sarcasmo, tornando a interpretação e classificação desses textos mais desafiadoras. Além disso, a variação na expressão de sentimentos em diferentes contextos e a relevância dos termos utilizados no pré-processamento dos textos também são questões identificadas.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>dados não estruturados, utilizando Redes Neurais Artificiais do tipo Multi-Layer Perceptron (MLP) como algoritmo classificador. O foco está na avaliação de sentimentos em textos escritos em língua inglesa, utilizando informações linguísticas contextuais. O estudo busca contribuir com pesquisas relacionadas à aprendizagem de máquina voltada para os processos de organização e recuperação da informação.</p>	<p>do processo de classificação, chegando a uma melhoria de até 30% nas métricas utilizadas para avaliar a classificação. A exclusão de termos pouco relevantes para a classificação dos sentimentos expressos pelos usuários resultou em melhores resultados. Além disso, foi observado que termos de comunicação humana que envolvem ironia, figuras de linguagem ou expressões idiomáticas regionalizadas podem interferir negativamente no processo de aprendizado do classificador, dificultando a correta classificação dos textos.</p>	<p>classificação automática de documentos não estruturados, como artigos, relatórios e teses. Utilizando técnicas avançadas de processamento de linguagem natural e treinamento de algoritmos de classificação, é possível aprimorar a organização e recuperação da informação nesses repositórios, facilitando a busca e acesso aos conteúdos relevantes para os usuários. A identificação e classificação de sentimentos expressos nos textos podem auxiliar na análise de feedbacks, avaliações e opiniões presentes nos documentos, contribuindo para uma melhor compreensão e utilização dos recursos disponíveis nos repositórios institucionais.</p>
<p>2019</p>	<p>Biomedical Term Extraction: NLP Techniques in Computational Medicine</p>	<p>O estudo aborda a extração de termos biomédicos em espanhol, árabe e japonês, destacando desafios como variação na inflexão de termos, escrita árabe e segmentação de palavras em japonês. A falta de recursos em NLP médico para o árabe é mencionada. O trabalho inclui a criação de um corpus multilíngue, anotação morfológica, extração de termos e desenvolvimento de ferramentas online para consulta de termos médicos. Essa pesquisa destaca a necessidade de abordagens adaptadas para</p>	<p>O objetivo da pesquisa é desenvolver e aplicar técnicas de Processamento de Linguagem Natural (PLN) para extrair e gerenciar termos biomédicos em diferentes idiomas, como espanhol, árabe e japonês. Isso envolve a criação de um corpus multilíngue de textos médicos, a anotação morfológica desses textos, a extração de listas de termos e o desenvolvimento de ferramentas online para consulta e esclarecimento de termos médicos. O estudo visa superar os desafios específicos</p>	<p>O resultado da pesquisa foi a criação de um corpus multilíngue de textos médicos em espanhol, árabe e japonês, a anotação morfológica desses textos, a extração de listas de termos biomédicos e o desenvolvimento de ferramentas online para consulta e esclarecimento de termos médicos. Foram identificados desafios específicos relacionados à variação linguística e às características únicas dos termos médicos em cada idioma, como a inflexão em espanhol, a escrita árabe e a segmentação</p>	<p>Utilizando técnicas de Processamento de Linguagem Natural (PLN), pode ser aplicado em repositórios institucionais para melhorar a indexação e recuperação de informações. A criação de um corpus multilíngue de textos, anotação morfológica, extração de termos e o desenvolvimento de ferramentas online podem auxiliar na organização e busca eficiente de conteúdo. Essas técnicas podem aprimorar a precisão e relevância dos resultados de pesquisa em repositórios institucionais, beneficiando</p>

		<p>lidar com a diversidade linguística e desafios únicos dos termos médicos em diferentes idiomas.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>apresentados pela diversidade linguística e pelas características únicas dos termos médicos em cada idioma, contribuindo para a eficácia na extração e gerenciamento desses termos no contexto biomédico.</p>	<p>de palavras em japonês. A pesquisa contribuiu para a aplicação de técnicas de PLN na extração e gerenciamento eficaz de termos biomédicos em diferentes idiomas, fornecendo recursos valiosos para profissionais de saúde, terminologistas, tradutores e estudantes de medicina.</p>	<p>profissionais, pesquisadores e estudantes que buscam informações em diferentes idiomas</p>
2019	<p>Boosting Document Retrieval with Knowledge Extraction and Linked Data</p>	<p>O problema de pesquisa abordado neste estudo é avaliar como a extração de conhecimento e a utilização de dados vinculados podem melhorar a recuperação de documentos. A questão de pesquisa central é investigar se a expansão semântica de consultas e documentos, por meio da combinação de dados abertos vinculados e técnicas de extração de conhecimento, pode aprimorar o desempenho da recuperação de documentos em comparação com abordagens tradicionais baseadas apenas em informações textuais.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo deste estudo é avaliar uma abordagem de recuperação de documentos que utiliza a extração de conhecimento e dados vinculados para melhorar o desempenho da recuperação de documentos. Mais especificamente, o estudo visa investigar se a expansão semântica de consultas e documentos, por meio da integração de dados abertos vinculados e técnicas de extração de conhecimento, pode superar significativamente o desempenho da recuperação de documentos em comparação com abordagens convencionais que se baseiam apenas em informações textuais.</p>	<p>Os resultados do estudo demonstraram que a expansão com conteúdo semântico extraído de consultas e documentos possibilitou consistentemente superar o desempenho de recuperação em comparação com abordagens que utilizam apenas informações textuais. Em uma avaliação específica em um conjunto de dados para busca semântica, o sistema KE4IR superou um sistema de busca baseado em ontologias de referência. Além disso, os experimentos validaram a viabilidade da aplicação de técnicas de extração de conhecimento para a recuperação de documentos em grandes coleções.</p>	<p>O estudo sobre a recuperação de documentos com a utilização de conhecimento extraído e dados vinculados pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de informações. Ao integrar técnicas de extração de conhecimento e dados vinculados, os repositórios podem oferecer uma busca mais precisa e abrangente, permitindo aos usuários encontrar documentos relevantes com maior eficiência. Isso pode melhorar a descoberta de conteúdo, facilitar a navegação e promover a interoperabilidade entre diferentes sistemas de informação em repositórios institucionais.</p>
2019	<p>Building and Querying an Enterprise Knowledge Graph</p>	<p>O problema de pesquisa abordado no documento é como construir e consultar um grafo de</p>	<p>O objetivo da pesquisa é desenvolver e apresentar os esforços da Thomson Reuters</p>	<p>Os resultados da pesquisa evidenciam o sucesso na implementação de um grafo de</p>	<p>O estudo sobre a construção e consulta de um grafo de conhecimento empresarial pode</p>

		<p>conhecimento empresarial de forma eficaz e intuitiva, especialmente para usuários não técnicos que não estão familiarizados com linguagens de consulta especializadas, como SPARQL. O objetivo é superar as barreiras entre os usuários e os dados, permitindo que eles encontrem as informações relevantes de maneira eficiente em um mundo de dados complexos e competitivos. A solução proposta envolve a aquisição de dados de várias fontes, a mineração de informações valiosas, a integração desses dados em um grafo de conhecimento e a disponibilização de uma interface de linguagem natural para consulta.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>na construção e consulta de um grafo de conhecimento empresarial. Isso inclui a aquisição de dados de diversas fontes, o desenvolvimento de técnicas de reconhecimento de entidades nomeadas, extração de relações e vinculação de entidades para extrair informações valiosas dos dados, a modelagem e armazenamento dos dados em triplas RDF, e a criação de uma interface de linguagem natural, TR Discover, para permitir que os usuários pesquisem informações no grafo de conhecimento de forma intuitiva. O objetivo final é avaliar a viabilidade e a eficácia do grafo de conhecimento desenvolvido e explorar maneiras de aprimorar os algoritmos de processamento de linguagem natural e expandir a implantação do grafo para mais produtos, com base no feedback dos clientes.</p>	<p>conhecimento empresarial pela Thomson Reuters, em produção por um ano, replicado em dois data centers para alta disponibilidade. O processo contínuo de ingestão, transformação e integração de dados mantém o grafo atualizado, equilibrando a ingestão de dados e a disponibilidade online. A cobertura de domínio continua a expandir-se com o desenvolvimento de novos produtos dependentes do grafo. A eficácia de técnicas como reconhecimento de entidades, extração de relações e vinculação de entidades é demonstrada, assim como a utilidade da interface de linguagem natural TR Discover. O grafo atua como fonte central de dados, conectando informações e beneficiando melhorias em produtos existentes e desenvolvimento de novos.</p>	<p>ser aplicado em repositórios institucionais para melhorar a organização, busca e recuperação de informações. A aquisição de dados de diversas fontes, a mineração de informações valiosas e a integração em um grafo de conhecimento podem ajudar a enriquecer os repositórios com dados estruturados e não estruturados. Além disso, a implementação de uma interface de linguagem natural para consulta pode tornar a busca por informações mais intuitiva e acessível para os usuários, facilitando a descoberta de conhecimento relevante nos repositórios institucionais.</p>
2019	Content-based table retrieval for web queries	<p>A problemática de pesquisa abordada neste estudo é a recuperação de tabelas com base em conteúdo para consultas na web. O desafio principal consiste em encontrar a tabela mais relevante em uma coleção de tabelas, dada uma determinada consulta. Isso</p>	<p>O objetivo da pesquisa é desenvolver abordagens baseadas em classificação e redes neurais para a recuperação de tabelas com base em conteúdo para consultas na web. O foco está em encontrar a tabela mais relevante em uma coleção de</p>	<p>Os resultados da pesquisa demonstram a eficácia das abordagens propostas, tanto baseadas em características cuidadosamente projetadas quanto em arquiteturas de redes neurais, para medir a relevância entre consultas e o conteúdo das tabelas. Foram realizados</p>	<p>O estudo sobre recuperação de tabelas com base em conteúdo para consultas na web pode ser aplicado em repositórios institucionais para melhorar a eficiência e precisão da recuperação de informações. Ao utilizar abordagens baseadas em características e redes</p>

		<p>envolve a representação eficaz de tabelas semi-estruturadas, a compreensão das relações entre o texto não estruturado e o conteúdo das tabelas, e a aplicação de abordagens de classificação e arquiteturas de redes neurais para medir a relevância entre a consulta e o conteúdo da tabela. A pesquisa visa melhorar os sistemas de recuperação de informações existentes e explorar a conexão entre tabelas e texto para avançar no campo de processamento de linguagem natural e recuperação de informações.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>tabelas, dada uma consulta específica, e em explorar a conexão entre texto não estruturado e tabelas semi-estruturadas. Além disso, a pesquisa visa disponibilizar um conjunto de dados aberto que inclui consultas da web e tabelas da Wikipedia, a fim de fornecer recursos valiosos para futuras pesquisas nessa área.</p>	<p>experimentos abrangentes em conjuntos de dados do mundo real e sintéticos, com diversos critérios de avaliação que mostram que a abordagem proposta tem um desempenho comparável ou melhor do que um sistema baseado em características cuidadosamente projetadas. Os resultados também destacam a importância de compreender profundamente o conteúdo das tabelas e da linguagem para obter bons resultados nessa tarefa. Além disso, são apresentados desafios futuros e oportunidades para explorar conexões mais profundas entre tabelas e texto, incentivando o interesse da comunidade em investigar essa área de pesquisa.</p>	<p>neurais para medir a relevância entre consultas e tabelas, é possível aprimorar a busca e recuperação de dados estruturados, como relatórios, estatísticas e informações específicas armazenadas em tabelas. Isso pode facilitar a localização de informações relevantes e a resposta a consultas específicas, contribuindo para uma melhor organização e acesso aos dados nos repositórios institucionais.</p>
<p>2019</p>	<p>Coupled Intrinsic and Extrinsic Human Language Resource-Based Query Expansion</p>	<p>A problemática de pesquisa abordada neste artigo é o desafio do descompasso de vocabulário entre consultas de usuários e documentos relevantes em sistemas de recuperação de informações. Essa discrepância dificulta a formulação de consultas precisas em linguagem natural que estejam alinhadas com o vocabulário dos documentos considerados relevantes para</p>	<p>O objetivo da pesquisa é apresentar um framework de expansão de consulta que capitalize nas características linguísticas intrínsecas das consultas dos usuários e em recursos linguísticos extrínsecos, como ontologias, para a codificação de constituintes da consulta, extração de conceitos de expansão e ponderação de conceitos. A pesquisa visa</p>	<p>O resultado da pesquisa foi a apresentação de um framework de expansão de consulta baseado em recursos linguísticos que combina características linguísticas intrínsecas das consultas dos usuários com recursos linguísticos extrínsecos, como ontologias. Esse framework foi projetado para codificar os constituintes da consulta, extrair conceitos de expansão</p>	<p>O framework de expansão de consulta baseado em recursos linguísticos proposto nesta pesquisa pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações. Ao utilizar características linguísticas intrínsecas das consultas dos usuários e recursos linguísticos extrínsecos, como ontologias, o framework pode ajudar a identificar e incluir conceitos</p>

		<p>um objetivo de busca específico. A solução proposta para essa problemática é a aplicação de processos de expansão de consulta para gerar e integrar termos adicionais a uma consulta inicial, visando garantir a identificação precisa dos principais conceitos da consulta e a inclusão de conceitos de expansão relevantes na consulta enriquecida.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>melhorar o desempenho da recuperação de informações, abordando a problemática do descompasso de vocabulário entre consultas e documentos relevantes. A avaliação empírica do framework proposto é realizada em conjuntos de dados do mundo real para validar a abordagem em comparação com modelos de linguagem unigram, modelo de relevância e uma técnica baseada em dependência sequencial.</p>	<p>relevantes e ponderar esses conceitos de forma a melhorar o desempenho da recuperação de informações. A avaliação empírica do framework foi realizada em conjuntos de dados do mundo real e comparada com modelos de linguagem unigram, modelo de relevância e uma técnica baseada em dependência sequencial, demonstrando a eficácia e relevância do framework proposto em lidar com o desafio do descompasso de vocabulário na recuperação de informações.</p>	<p>relevantes nas consultas, aumentando a precisão e relevância dos resultados recuperados. Isso pode aprimorar a experiência de busca dos usuários em repositórios institucionais, garantindo que suas consultas sejam mais alinhadas com o conteúdo disponível e resultem em uma recuperação mais eficaz de informações pertinentes.</p>
<p>2019</p>	<p>Criminal Event Ontology Population and Enrichment using Patterns Recognition from Text</p>	<p>O problema de pesquisa abordado neste estudo é a extração de eventos criminais e suas causas a partir de textos em espanhol, com foco na população e enriquecimento de uma ontologia para representar esses eventos. O estudo visa desenvolver um processo baseado em reconhecimento de padrões para identificar eventos criminais específicos em manchetes de notícias, com o objetivo de auxiliar na prevenção de crimes e na segurança dos</p>	<p>O objetivo deste estudo é utilizar o reconhecimento de padrões para extrair eventos criminais e suas causas de textos em espanhol, populando e enriquecendo um modelo de ontologia. Especificamente, o estudo visa identificar eventos criminais específicos, como assassinatos, violações e agressões, a partir de manchetes de notícias, e representar esses eventos e suas relações de causa por meio de uma ontologia. O objetivo</p>	<p>O estudo conduziu à criação de uma ontologia abrangendo eventos criminais como assassinato, violação e agressão, acompanhados por entidades e causas correlatas. Desenvolveram-se padrões linguísticos para extrair eventos e relações causais, abrangendo frases verbais, nominais, preposicionais e de conexão. A avaliação da ontologia demonstrou resultados promissores em consistência, eficiência, precisão e</p>	<p>A integração do método de reconhecimento de padrões e ontologia em repositórios institucionais possibilita a automação da extração e representação de informações a partir de textos. Isso simplifica a análise e a tomada de decisões relacionadas ao conteúdo recuperado, proporcionando maior eficiência no processo.</p>

		<p>cidadãos em comunidades de língua espanhola.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>final é contribuir para a prevenção de crimes e a segurança dos cidadãos em comunidades de língua espanhola, fornecendo uma base de conhecimento estruturada e enriquecida com informações relevantes sobre eventos criminais e suas causas.</p>	<p>generalidade. A precisão e recall na extração de eventos criminais e reconhecimento de causas foram satisfatórios. O estudo ressalta a importância da detecção precoce de eventos criminais na sociedade de língua espanhola, sugerindo futuros trabalhos, incluindo a extração automática de atores envolvidos e avaliação através de um sistema de perguntas e respostas baseado na ontologia.</p>	
2019	<p>Criação de um chatbot para atendimento de suporte sistêmico de uma rede de supermercados a partir de uma base de conhecimento registrada no Mantis?</p>	<p>A problemática de pesquisa abordada é a necessidade de responder de forma eficiente e precisa às dúvidas e questionamentos dos usuários, utilizando uma base histórica de atendimentos em uma rede de supermercados, visando atender à demanda diária de suporte a sistemas de forma ágil e eficaz.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo do estudo apresentado é propor e avaliar os retornos de um chatbot às perguntas inseridas em linguagem natural, utilizando uma base de dados de histórico de atendimentos de uma rede de supermercados, com o intuito de fornecer respostas automáticas e eficientes aos usuários, reduzindo o tempo de espera e aumentando a assertividade nas soluções apresentadas.</p>	<p>Os resultados do estudo indicam que o chatbot foi capaz de localizar respostas e consolidar possíveis soluções utilizadas em atendimentos anteriores, que podem ser aplicadas para resolver problemas atuais. A pesquisa demonstrou que os chatbots podem ser uma alternativa viável para atender usuários na prestação de serviços de suporte, proporcionando agilidade e eficiência.</p>	<p>O estudo demonstrou que a utilização de chatbots baseados em bases de conhecimento histórico pode ser uma solução eficaz para responder de forma rápida e precisa às dúvidas dos usuários. Essa abordagem pode ser aplicada em repositórios institucionais para facilitar a recuperação de informações importantes e relevantes, agilizando o acesso a dados e conhecimentos armazenados. Ao implementar sistemas de perguntas e respostas baseados em chatbots, as instituições podem melhorar a eficiência na busca por informações, reduzir o tempo de resposta e aumentar a assertividade nas soluções apresentadas aos usuários.</p>
2019	<p>Detecting new Chinese words from massive domain</p>	<p>O estudo aborda a detecção de novas palavras chinesas em textos extensos, destacando a importância dessa detecção</p>	<p>O objetivo do estudo é desenvolver um método eficaz para detectar novas palavras chinesas em textos de domínio</p>	<p>O estudo apresenta resultados promissores na detecção de novas palavras chinesas em textos de domínio massivo</p>	<p>Os repositórios institucionais podem aprimorar a identificação de novas palavras em textos extensos, beneficiando</p>

	texts with word embedding	<p>para a recuperação de informações textuais. As abordagens tradicionais de segmentação de palavras podem enfrentar desafios na identificação precisa de novas palavras devido à evolução constante da linguagem. A pesquisa propõe um método que combina segmentação de palavras e incorporação de palavras para melhorar a detecção de novas palavras em textos chineses extensos.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	massivo, visando melhorar a precisão e eficiência desse processo em comparação com as abordagens tradicionais de segmentação de palavras. A pesquisa busca superar os desafios enfrentados pela segmentação de palavras convencional na identificação de novas palavras, aproveitando a incorporação de palavras para aprimorar a detecção de novos termos em textos extensos em chinês.	utilizando a abordagem de incorporação de palavras. A combinação de segmentação de palavras e incorporação de palavras demonstrou melhorias significativas na precisão e eficiência da detecção de novas palavras em comparação com métodos tradicionais. A utilização do modelo word2vec para gerar vetores de palavras mostrou-se eficaz na identificação de novas palavras, contribuindo para aprimorar a recuperação de informações textuais em textos extensos em chinês. Os resultados sugerem que a abordagem proposta pode ser uma ferramenta valiosa para lidar com a evolução da linguagem e a detecção de novas palavras em contextos de grande volume de dados.	pesquisadores, estudantes e demais usuários na busca por informações específicas. Essa abordagem contribui para a eficiência e precisão na recuperação de dados em contextos acadêmicos e institucionais.
2019	Discovering composable web services using functional semantics and service dependencies based on natural language requests	O problema de pesquisa abordado no documento é a descoberta de serviços da web componíveis usando semântica funcional e dependências de serviços com base em solicitações em linguagem natural. O foco está em identificar serviços básicos relevantes para uma tarefa específica e em encontrar conjuntos de serviços básicos que possam ser combinados para atender aos requisitos do usuário, tanto para consultas simples quanto para consultas	O objetivo da pesquisa é propor um framework para descobrir conjuntos de serviços componíveis de acordo com os requisitos complexos do usuário. Isso é feito utilizando técnicas de processamento de linguagem natural e semântica para extrair a semântica funcional dos conjuntos de serviços e compreender o contexto do usuário. O foco é encontrar serviços básicos ou composições de serviços que atendam aos requisitos do usuário, seja por meio de	A pesquisa propôs um framework para descobrir serviços web componíveis, utilizando semântica funcional e dependências de serviços baseadas em solicitações em linguagem natural. O framework foi desenvolvido para identificar serviços básicos relevantes e encontrar conjuntos combináveis para atender aos requisitos do usuário. Os resultados indicaram eficácia na descoberta de serviços para consultas simples e complexas, melhorando a precisão na	O estudo sobre a descoberta de serviços da web componíveis usando semântica funcional e dependências de serviços com base em solicitações em linguagem natural pode ser aplicado em repositórios institucionais para melhorar a descoberta e seleção de serviços relevantes para atender às necessidades dos usuários. Ao utilizar técnicas de processamento de linguagem natural e semântica, é possível aprimorar a busca por serviços básicos e composições de

		<p>complexas. A pesquisa visa melhorar a precisão na descoberta de serviços básicos e composições de serviços disponíveis que possam atender aos requisitos do usuário, caso um serviço básico sozinho não seja satisfatório.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>serviços básicos individuais ou por meio de conjuntos de serviços que juntos possam cumprir a tarefa desejada. A pesquisa visa melhorar a precisão na descoberta de serviços básicos e composições de serviços disponíveis, especialmente para consultas complexas.</p>	<p>identificação e composição de serviços disponíveis. Experimentos demonstraram que o método é preciso na geração de modelos de serviços componíveis e escalável para conjuntos de dados maiores.</p>	<p>serviços que atendam aos requisitos específicos dos usuários. Isso pode resultar em uma melhor organização e disponibilização de serviços dentro dos repositórios institucionais, facilitando a integração e composição de serviços para atender às demandas dos usuários de forma mais eficiente e precisa.</p>
<p>2019</p>	<p>Discriminative Feature Spamming Technique for Roman Urdu Sentiment Analysis</p>	<p>A problemática de pesquisa abordada neste estudo está relacionada à análise de sentimentos em texto em Roman Urdu. Alguns dos desafios e complexidades enfrentados incluem a falta de padronização na representação da língua Urdu em script latino, variações na ortografia das palavras, riqueza morfológica da língua Urdu e a presença de características correlacionadas nos dados de texto. Além disso, a questão do "curse of dimensionality" é destacada, referindo-se aos desafios envolvidos na análise de dados de alta dimensionalidade, especialmente em dados textuais, onde a</p>	<p>O objetivo da pesquisa é desenvolver e avaliar uma técnica inovadora de ponderação de termos chamada Discriminative Feature Spamming Technique (DFST) para melhorar a precisão da análise de sentimentos em texto em Roman Urdu. Além disso, o estudo visa explorar a influência de diferentes conjuntos de características e técnicas de ponderação de termos na precisão de classificação de diferentes algoritmos. O trabalho também se propõe a construir o maior conjunto de dados de Roman Urdu já reunido, composto por 11000 avaliações, e a realizar experimentos extensivos para comparar os</p>	<p>Os resultados do estudo demonstraram que a técnica DFST superou um conjunto de técnicas de ponderação de termos testadas ao longo do tempo, provenientes do campo de recuperação de informações. Os experimentos foram realizados no maior conjunto de dados de Roman Urdu já coletado, composto por 11000 avaliações, e mostraram melhorias significativas na precisão da classificação. Além disso, a comparação cruzada entre diferentes técnicas de ponderação de termos revelou que os resultados obtidos com a utilização do DFST foram estatisticamente significativos e superiores às abordagens</p>	<p>O estudo sobre a Discriminative Feature Spamming Technique (DFST) para análise de sentimentos em Roman Urdu oferece potenciais benefícios aplicados em repositórios institucionais, contribuindo para aprimorar a classificação e organização de conteúdo textual. A técnica DFST, ao identificar e priorizar termos discriminativos, pode otimizar a eficiência na recuperação de informações e análise de sentimentos em documentos. A implementação desta abordagem inovadora de ponderação de termos nos repositórios institucionais pode aprimorar a precisão na categorização e busca de</p>

		<p>dimensionalidade aumenta exponencialmente devido à diversidade de palavras e variações de ortografia, exigindo técnicas especializadas de ponderação de características para extrair informações significativas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>resultados da técnica proposta com as abordagens convencionais de ponderação de termos.</p>	<p>anteriores. Esses resultados destacam a eficácia da técnica proposta na análise de sentimentos em texto em Roman Urdu.</p>	<p>documentos, facilitando a recuperação de informações relevantes e a análise de sentimentos em textos de maneira mais eficaz e precisa.</p>
<p>2019</p>	<p>DSPF: A Digital Signal Processing Based Framework for Information Retrieval</p>	<p>O estudo aborda as limitações dos modelos de recuperação de informação baseados em processamento de sinais digitais, enfatizando desafios como a dificuldade na previsão da natureza dos sinais na frequência, a estreiteza das linhas espectrais ao considerar consultas como sinais senoidais e a ausência de estudo sobre a introdução de esquemas de ponderação de termos de outras categorias de modelos de recuperação de informação. Para superar tais obstáculos, os pesquisadores propuseram o Framework de Recuperação de Informação baseado em Processamento de Sinais</p>	<p>O objetivo do estudo é propor um novo Framework de Recuperação de Informação baseado em Processamento de Sinais Digitais (DSPF) que incorpora teorias do campo de processamento de sinais digitais e apresenta dois modelos correspondentes de recuperação de informação baseados em DSP, denominados DSPF-BM25 e DSPF-DLM. Esses modelos buscam melhorar a eficácia dos sistemas de recuperação de informação ao representar cada termo de consulta como um espectro com forma gaussiana e ao ajustar os parâmetros diretamente no domínio da</p>	<p>Os resultados do estudo mostraram que, na maioria dos casos, os modelos propostos (DSPF-BM25 e DSPF-DLM) superaram os baselines fortes em termos de Média de Precisão Média (MAP) e foram comparáveis em termos de P@10 e P@20. Os experimentos realizados demonstraram a eficácia dos modelos propostos em melhorar a classificação e recuperação de documentos em comparação com abordagens tradicionais. Além disso, foram recomendados intervalos de valores de parâmetros para garantir a robustez dos modelos propostos.</p>	<p>O estudo sobre o Framework de Recuperação de Informação baseado em Processamento de Sinais Digitais (DSPF) pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações e documentos. Ao utilizar modelos DSPF-BM25 e DSPF-DLM, é possível melhorar a classificação e ranking de documentos com base na energia dos sinais filtrados, proporcionando uma abordagem mais eficaz e precisa para a busca de informações nos repositórios. Isso pode resultar em uma recuperação mais eficiente e relevante de conteúdo para os usuários dos repositórios institucionais.</p>

		Digitais (DSPF) como uma abordagem inovadora.  <b>Categoria:</b> Recursos de informação  <b>Subcategoria:</b> Representação do documento	frequência, proporcionando maior controle e precisão.		
2019	Empirical evaluation of compounds indexing for Turkish texts	O problema de pesquisa abordado no artigo é a avaliação empírica da indexação de compostos para textos em língua turca. O estudo investiga a eficácia do uso de compostos como termos de índice na recuperação de informações em textos turcos, comparando essa abordagem com métodos tradicionais de indexação por palavras-chave. A pesquisa busca analisar como a indexação de compostos pode melhorar o desempenho da recuperação de informações em textos turcos e identificar os tipos de padrões de compostos mais eficazes para esse fim.  <b>Categoria:</b> Recursos de informação  <b>Subcategoria:</b> Representação do documento	O objetivo do estudo é propor um framework para extração e indexação de compostos turcos, indo além dos métodos tradicionais de indexação por palavras-chave. O estudo visa identificar doze tipos de padrões de compostos turcos e classificá-los em seis categorias, avaliando a eficácia do uso de compostos como termos de índice na recuperação de informações. Além disso, o estudo investiga a eficácia de diferentes estratégias de indexação de compostos, incluindo o uso de um único tipo de composto e a combinação de diferentes tipos de compostos. Os experimentos são realizados sobre o conjunto de dados de teste Milliyet, com o intuito de analisar como o uso de compostos pode impactar positivamente o desempenho da recuperação de informações em textos turcos.	O estudo realizou uma avaliação empírica da indexação de compostos para textos em língua turca, comparando-a com métodos tradicionais de indexação por palavras-chave. O objetivo foi propor um framework para extração e indexação de compostos turcos, identificando padrões eficazes. Os resultados mostraram que o uso de compostos como termos de índice melhorou significativamente o desempenho da recuperação de informações, especialmente em precisão em níveis baixos de recall. A pesquisa destacou a importância de identificar os tipos de compostos mais eficazes e demonstrou a superioridade da abordagem de compostos em relação à indexação por palavras-chave.	A aplicação do estudo sobre indexação em repositórios institucionais visa aprimorar a recuperação de informações. Ao empregar compostos como termos de índice, há um potencial aumento na precisão da busca, especialmente em níveis baixos de recall. Identificar e indexar de maneira eficaz os tipos de compostos contribui para melhorar a organização e a recuperação de documentos nos repositórios, resultando em buscas mais eficientes e precisas para os usuários.
2019	Exploiting named entity recognition	A pesquisa visa explorar como a inclusão de descrições de	O foco está em mitigar o problema de incompatibilidade	O estudo visava melhorar a descoberta de serviços da web	O estudo sobre a melhoria da descoberta de serviços da web

	<p>for improving syntactic-based web service discovery</p>	<p>entidades nomeadas em serviços da web pode aprimorar a capacidade de descoberta desses serviços, especialmente em ambientes baseados em sintaxe. A questão central é como a aplicação de técnicas de reconhecimento de entidades nomeadas pode contribuir para a eficácia e eficiência da descoberta de serviços da web, superando desafios como o problema de vocabulário e a correspondência de termos.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>de termos e melhorar a eficácia dos métodos de descoberta de serviços da web baseados em sintaxe. A pesquisa busca avaliar como a inclusão de informações de entidades nomeadas pode aumentar a precisão e a recuperação de serviços da web, comparando essa abordagem com métodos tradicionais de descoberta de serviços baseados em sintaxe.</p> <p><b>Entidades nomeadas</b> são termos específicos que se referem a entidades do mundo real, como nomes de pessoas, organizações, locais, datas, valores numéricos, entre outros.</p>	<p>através do reconhecimento de entidades nomeadas. Ao expandir automaticamente as descrições dos serviços com informações de entidades nomeadas, houve melhorias significativas nos resultados. A abordagem demonstrou aumentos de até 17,24% na precisão e 15,06% na recuperação dos serviços. Além disso, métricas como NDCG e F-Measure também apresentaram melhorias. A comparação com métodos tradicionais mostrou a eficácia dessa técnica na otimização da descoberta de serviços da web.</p>	<p>através do reconhecimento de entidades nomeadas pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações. Ao expandir automaticamente as descrições dos serviços com informações de entidades nomeadas, é possível melhorar a precisão e a eficiência na busca por conteúdo nos repositórios. Essa abordagem pode ajudar os usuários a encontrar informações relevantes de forma mais eficaz, reduzindo problemas de incompatibilidade de termos e aumentando a recuperação de documentos pertinentes.</p>
<p>2019</p>	<p>Extracting Emotion Causes Using Learning to Rank Methods From an Information Retrieval Perspective</p>	<p>O estudo aborda a desafiadora tarefa de extrair causas de emoção, visando identificar os estímulos por trás de uma emoção observada. O desafio reside na ambiguidade e sutileza das expressões emocionais, exigindo a análise do contexto que envolve as causas e palavras emocionais. A extração de causas de emoção é essencial para uma análise aprofundada das emoções, destacando a necessidade de capturar a relevância entre as cláusulas candidatas e as emoções observadas.</p>	<p>O objetivo do trabalho apresentado é propor um método inovador baseado em aprendizado de classificação para extração de causas de emoção, visando uma análise mais detalhada das emoções. Os autores buscam abordar a tarefa desafiadora de identificar as causas das emoções de forma mais precisa e refinada, utilizando métodos de aprendizado de classificação para classificar as cláusulas candidatas de acordo com sua relevância para as emoções observadas.</p>	<p>Os autores demonstram a eficácia do uso de métodos de aprendizado de classificação para classificar as cláusulas candidatas de acordo com sua relevância para as emoções observadas. Eles também exploram a influência de diferentes modelos de tópicos e a remoção de stopwords na performance da extração de causas de emoção, fornecendo insights valiosos sobre a abordagem proposta.</p>	<p>Ao utilizar técnicas de aprendizado de classificação para identificar as causas das emoções expressas em documentos, os repositórios institucionais podem aprimorar a compreensão do impacto emocional do conteúdo armazenado, isso pode ser útil em áreas como análise de feedback de usuários, avaliação de satisfação do cliente e identificação de tendências emocionais em dados textuais, contribuindo para uma análise mais abrangente e significativa</p>

		<p>considerando a informação emocional transmitida nas cláusulas.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>			do conteúdo disponível nos repositórios institucionais.
2019	Hindi Query Expansion based on Semantic Importance of Hindi WordNet Relations and Fuzzy Graph Connectivity Measures	<p>O problema de pesquisa abordado neste estudo é a expansão de consultas em Hindi com base na importância semântica das relações do Hindi WordNet e nas medidas de conectividade de grafos difusos. O objetivo é melhorar o desempenho da recuperação de informações para consultas em Hindi, considerando a ambiguidade e a complexidade da língua Hindi, bem como a necessidade de fornecer resultados relevantes e precisos para os usuários que buscam informações nesse idioma.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo deste estudo é propor um método de expansão de consultas em Hindi que leve em consideração a importância semântica das relações presentes no Hindi WordNet. O objetivo principal é melhorar o desempenho da recuperação de informações para consultas em Hindi, utilizando medidas de conectividade de grafos difusos para identificar termos adicionais relevantes que possam aprimorar a precisão e a relevância dos resultados obtidos. A pesquisa visa otimizar a compreensão e o processamento de consultas em Hindi, levando em consideração a riqueza semântica do Hindi WordNet e a complexidade da língua Hindi.</p>	<p>A abordagem de expansão de consultas em Hindi, baseada na importância semântica das relações do Hindi WordNet e em medidas de conectividade de grafos difusos, demonstrou melhorias significativas na recuperação de informações em comparação com métodos anteriores. A atribuição de significância às relações semânticas no Hindi WordNet facilitou o agrupamento de palavras relacionadas no grafo do WordNet, resultando em maior precisão e recall nas consultas em Hindi. Os resultados experimentais destacaram a eficácia dessa abordagem, superando métodos existentes e aprimorando a qualidade dos resultados de recuperação de informações no idioma Hindi.</p>	<p>A implementação da importância semântica, por meio das relações do Hindi WordNet e medidas de conectividade de grafos difusos em um repositório institucional, promete aprimorar significativamente a recuperação de informações em Hindi. Essa abordagem visa fornecer resultados de pesquisa mais relevantes e precisos, elevando a experiência de busca dos usuários.</p>
2019	Identification of New Parameters for Ontology Based	<p>O problema de pesquisa abordado neste artigo é a identificação de novos parâmetros para medidas de</p>	<p>O objetivo deste estudo é analisar e identificar novos parâmetros que possam ser utilizados na formulação de</p>	<p>O estudo identifica vários parâmetros que impactam a precisão e eficácia das medidas de similaridade semântica</p>	<p>O estudo sobre a identificação de novos parâmetros para medidas de similaridade semântica baseadas em</p>

	Semantic Similarity Measures	<p>similaridade semântica baseadas em ontologia. Os pesquisadores exploram diferentes abordagens e parâmetros para calcular a similaridade semântica entre palavras, com o objetivo de melhorar a precisão e adaptabilidade dos sistemas de processamento de linguagem natural e recuperação de informações.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>medidas de similaridade semântica baseadas em ontologia. Os pesquisadores buscam aprimorar a avaliação e comparação de medidas de similaridade existentes, bem como fornecer insights para o desenvolvimento de novas medidas que possam ser aplicadas em áreas como mineração de texto, mineração web e recuperação de informações.</p>	<p>baseadas em ontologia. Os pesquisadores sublinham a importância de considerar elementos como conjunto de características, tipo de ontologia, precisão, estrutura de desenvolvimento de ontologia, relações entre palavras e tipo de conjunto de dados. A análise desses parâmetros em diferentes abordagens de medidas de similaridade oferece insights valiosos para o desenvolvimento de novas medidas mais precisas e adaptáveis em diversas aplicações de processamento de linguagem natural e recuperação de informações.</p>	<p>ontologia pode ser aplicado em repositórios institucionais para melhorar a precisão e eficácia da recuperação de informações ao considerar os parâmetros identificados, como conjunto de características, tipo de ontologia, precisão, estrutura de desenvolvimento de ontologia, relações entre palavras e tipo de conjunto de dados, os repositórios institucionais podem aprimorar a busca e organização de informações, facilitando a recuperação de conteúdo relevante para os usuários. Essa abordagem pode contribuir para uma melhor experiência de uso e eficiência na gestão de dados em repositórios institucionais.</p>
2019	Integrating heterogeneous sources for predicting question temporal anchors across Yahoo! Answers	<p>O problema de pesquisa abordado neste artigo é a integração de fontes heterogêneas para prever âncoras temporais de perguntas em fóruns de perguntas e respostas da comunidade, como o Yahoo! Answers. O estudo visa melhorar a precisão e relevância dos resultados de busca ao considerar a temporalidade das perguntas e respostas, permitindo uma apresentação mais dinâmica e contextualizada das informações.</p> <p><b>Categoria:</b> Recursos de informação</p>	<p>O objetivo deste estudo é investigar a complexidade da identificação de âncoras temporais em perguntas de diferentes tópicos em fóruns de perguntas e respostas, como o Yahoo! Answers, tanto para humanos quanto para máquinas. Para isso, os pesquisadores desenvolveram um corpus humano anotado com a participação de até quatorze trabalhadores, baseado exclusivamente no conteúdo do Yahoo! Answers. Além disso, o estudo propõe novas categorias específicas de intervalo de tempo, realiza experimentos para avaliar a eficácia de</p>	<p>O estudo propôs novas categorias de intervalo de tempo para perguntas em fóruns de perguntas e respostas, criando um corpus de alta qualidade com 6683 perguntas rotuladas. Foram realizados experimentos para avaliar características linguísticas e evidências de recursos heterogêneos, como buscas na web. O aprendizado multi-visão mostrou-se eficaz na classificação de perguntas. Informações linguísticas e de snippets da web foram cruciais para prever a temporalidade das perguntas, destacando a importância de características como frases nominais e sujeitos</p>	<p>O estudo sobre a previsão de âncoras temporais em perguntas de fóruns de perguntas e respostas tem como objetivo aplicar essa abordagem em repositórios institucionais, visando melhorar a organização e relevância das informações. A consideração da temporalidade inclui a introdução de novas categorias de intervalo de tempo, junto com características linguísticas e evidências de recursos heterogêneos, com o propósito de aprimorar a classificação e busca de conteúdo nos repositórios. Essa estratégia busca proporcionar uma apresentação mais</p>

		<b>Subcategoria:</b> Representação do documento	diferentes características linguísticas e investiga o impacto do aprendizado multi-visão na classificação de perguntas.	sintáticos. As informações extraídas de buscas na web foram as mais relevantes para prever a temporalidade das perguntas em fóruns de perguntas e respostas.	dinâmica e contextualizada das informações, beneficiando o público-alvo. A previsão de âncoras temporais, ao categorizar o tempo associado às perguntas, emerge como uma ferramenta valiosa para otimizar a busca e classificação de conteúdo em repositórios institucionais, promovendo uma experiência mais eficaz e alinhada às necessidades dos usuários.
2019	Machine Reading Comprehension for Answer Re-Ranking in Customer Support Chatbots	<p>O problema de pesquisa abordado neste artigo é aprimorar a seleção de respostas em chatbots de suporte ao cliente, utilizando técnicas de Machine Reading Comprehension para reclassificar as respostas sugeridas com base na pergunta feita. O objetivo é melhorar a eficácia e a relevância das respostas fornecidas pelos chatbots em interações de suporte ao cliente.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	O objetivo deste estudo é desenvolver um framework de reclassificação de respostas baseado em Machine Reading Comprehension para pares de perguntas e respostas em chatbots de suporte ao cliente. O foco é encontrar a resposta mais adequada para uma pergunta, considerando que cada candidato pode ser produzido por um sistema diferente, como baseado em conhecimento, baseado em regras, redes neurais profundas, recuperação de informações, entre outros. O objetivo final é melhorar a qualidade e a relevância das respostas fornecidas pelos chatbots em cenários de suporte ao cliente.	O estudo revelou que o framework de reclassificação de respostas, fundamentado em Machine Reading Comprehension, apresenta melhorias substanciais em termos de sobreposição de palavras e semântica, tanto para modelos individuais quanto para combinações de modelos. A utilização da abordagem, incluindo amostragem negativa com dados do Twitter Customer Support Dataset, resultou em significativo aprimoramento na qualidade das respostas sugeridas por chatbots de suporte ao cliente. Os experimentos realizados confirmam a eficácia do framework na seleção de respostas mais pertinentes e precisas para as perguntas dos usuários em ambientes de suporte ao cliente.	O estudo propõe um framework de reclassificação de respostas baseado em Machine Reading Comprehension para chatbots de suporte ao cliente. Essa abordagem visa aprimorar a relevância e precisão das respostas em repositórios institucionais, beneficiando usuários em busca de informações. Ao empregar técnicas avançadas de processamento de linguagem natural, o framework contribui para uma experiência do usuário mais eficaz e personalizada ao interagir com sistemas de suporte, elevando a qualidade das respostas fornecidas.

<p>2019</p>	<p>Mining predicate-based entailment rules using deep contextual architecture</p>	<p>O problema de pesquisa abordado neste artigo é a extração de regras de implicação baseadas em predicados a partir de textos, com o objetivo de melhorar a cobertura e a precisão dessas regras para aplicações em tarefas de Processamento de Linguagem Natural. Os pesquisadores propõem o uso da Deep Contextual Architecture (DCA) como uma abordagem inovadora para lidar com esse desafio.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo deste estudo é desenvolver e aplicar a Deep Contextual Architecture (DCA) para extrair regras de implicação baseadas em predicados a partir de textos, visando melhorar a cobertura e a precisão das regras extraídas. A intenção é aprimorar a modelagem do contexto dos predicados, combinando informações estatísticas e semânticas, e demonstrar a eficácia dessa abordagem em comparação com métodos anteriores, alcançando um desempenho superior em termos de métricas de precisão e cobertura.</p>	<p>Os resultados da aplicação da Deep Contextual Architecture (DCA) para extrair regras de implicação baseadas em predicados foram promissores, com a abordagem demonstrando potencial para melhorar a modelagem do contexto dos predicados ao integrar informações estatísticas e semânticas. A DCA superou diversas baselines robustas em desempenho, conforme avaliação em conjuntos de dados públicos. Apesar de alguns erros, como falsos positivos e negativos, a DCA foi eficaz em capturar e discernir a relação de implicação, mostrando-se promissora para lidar com desafios na extração de regras de implicação.</p>	<p>O estudo sobre a Deep Contextual Architecture (DCA) para extrair regras de implicação baseadas em predicados pode ser aplicado em repositórios institucionais para aprimorar a extração de informações e relações semânticas dos textos armazenados ao combinar informações estatísticas e semânticas, a abordagem melhora a modelagem do contexto dos predicados, resultando em uma mineração mais precisa e abrangente de regras de implicação. Isso beneficia a organização e recuperação de informações em repositórios institucionais, tornando a busca e análise de dados mais eficazes e significativas.</p>
<p>2019</p>	<p>MPSC: A Multiple-Perspective Semantics-Crossover Model for Matching Sentences</p>	<p>O problema de pesquisa abordado no artigo é a correspondência de sentenças, com foco na identificação de similaridade semântica em um nível profundo. O objetivo é desenvolver um modelo, denominado Multiple-Perspective Semantics-Crossover (MPSC), que possa avaliar o grau de correspondência entre duas sentenças com base em uma análise semântica avançada.</p>	<p>O objetivo do estudo é propor o modelo MPSC (Multiple-Perspective Semantics-Crossover) para a correspondência semântica entre duas sentenças. Esse modelo visa extrair informações de correspondência das sentenças a partir de diferentes ângulos semânticos, a fim de calcular o grau de correspondência entre elas. O MPSC não apenas captura padrões de correspondência ricos em diferentes níveis, mas também adquire características</p>	<p>Os resultados experimentais demonstraram que o modelo MPSC proposto apresentou uma eficácia superior em comparação com abordagens populares de correspondência semântica. Isso sugere que o MPSC é capaz de capturar padrões de correspondência mais ricos em diferentes níveis e extrair características interativas de várias perspectivas semânticas, o que o torna uma ferramenta eficaz para lidar com tarefas de Processamento de Linguagem Natural, como</p>	<p>O estudo sobre o modelo MPSC para correspondência de sentenças pode ser aplicado em repositórios institucionais para melhorar a correspondência semântica entre diferentes tipos de documentos e informações armazenadas. Ao utilizar o MPSC, é possível aprimorar a precisão na recuperação de informações relevantes, facilitando a busca e a análise de conteúdo nos repositórios. Isso pode resultar em uma melhor organização e acesso aos recursos disponíveis, tornando a</p>

		<p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>interativas de diferentes perspectivas semânticas. Além disso, o modelo é projetado para lidar com questões importantes em campos de Processamento de Linguagem Natural (NLP), como correspondência de informações em recuperação de texto e correspondência de perguntas e respostas em sistemas de Q&amp;A.</p>	<p>correspondência de informações em recuperação de texto e correspondência de perguntas e respostas em sistemas de Q&amp;A.</p>	<p>pesquisa e a navegação mais eficientes para os usuários dos repositórios institucionais.</p>
2019	<p>Multilingual Information Access (MLIA) Tools on Google and WorldCat: Bi/Multilingual University Students' Experience and Perceptions</p>	<p>O problema de pesquisa abordado no estudo é a experiência e percepções de estudantes universitários bi/multilíngues ao utilizar ferramentas de Acesso à Informação Multilíngue (MLIA) em plataformas como Google e WorldCat. O estudo investiga como esses estudantes lidam com a busca de informações em diferentes idiomas e como as ferramentas MLIA podem influenciar sua eficácia na busca de informações.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo do estudo foi examinar como os avanços tecnológicos em processamento de linguagem natural, como a disponibilidade de ferramentas de acesso à informação multilíngue (MLIA), impactam o comportamento de busca de informações de usuários acadêmicos bi/multilíngues. O estudo envolveu a participação de 31 estudantes bi/multilíngues em um experimento controlado em laboratório, onde realizaram tarefas atribuídas no Google e no WorldCat, seguidas por um questionário pós-experimento. O foco era analisar como as ferramentas MLIA afetam a busca de informações desses usuários.</p>	<p>O estudo revelou que 86,7% dos participantes utilizaram ferramentas de acesso à informação multilíngue durante o experimento. Houve uma diversidade significativa no comportamento de busca, mesmo para o mesmo usuário e par de idiomas. A utilização das ferramentas MLIA variou nas tarefas e na abordagem. Além disso, os participantes gastaram menos tempo em suas buscas ao usar as ferramentas MLIA. Os resultados indicam que modelos personalizados de recuperação de informações multilíngues podem ser promissores para representar efetivamente o comportamento de busca de usuários bi/multilíngues.</p>	<p>O estudo sobre o uso de ferramentas de Acesso à Informação Multilíngue (MLIA) por estudantes bi/multilíngues sugere aplicações em repositórios institucionais. A implementação de ferramentas MLIA pode aprimorar a busca de informações em diferentes idiomas, tornando-a mais eficiente e inclusiva. Instituições que mantêm repositórios podem considerar essa abordagem para facilitar o acesso a conteúdos multilíngues. Além disso, a personalização de modelos de recuperação de informações, conforme as preferências dos usuários, pode contribuir para uma experiência de busca mais eficaz e satisfatória nesses repositórios.</p>
2019	<p>Multi-Turn Response Selection for Chatbots With Hierarchical Aggregation Network</p>	<p>O problema de pesquisa abordado neste artigo a seleção de respostas em conversas de múltiplas rodadas para chatbots. Especificamente, o desafio é</p>	<p>O objetivo deste trabalho é propor um modelo de rede de agregação hierárquica de múltiplas representações (HAMR) para melhorar a seleção</p>	<p>Os resultados experimentais indicam que o modelo HAMR supera os modelos de estado-da-arte existentes, como o Deep Attention Matching Network</p>	<p>O estudo propõe um modelo avançado, o HAMR, para aprimorar a seleção de respostas em chatbots de múltiplas rodadas. Essa</p>

	of Multi-Representation	<p>calcular o grau de correspondência entre o contexto e as respostas candidatas, a fim de permitir que o chatbot recupere a resposta mais adequada e a retorne ao usuário.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>de respostas em chatbots de múltiplas rodadas. O modelo HAMR visa alavancar várias representações de contexto e resposta de forma eficaz, combinando-as e realçando informações valiosas para aprimorar o processo de seleção de respostas.</p>	<p>(DAM), em dois conjuntos de dados públicos de seleção de respostas de múltiplas rodadas: o Ubuntu Dialogue Corpus e o Douban Conversation Corpus. Em comparação com o DAM, o HAMR obteve melhor desempenho, com melhorias significativas em métricas como R@1, P@1 e R@10. Esses resultados demonstram que o modelo proposto alcança um novo estado-da-arte nessas tarefas de seleção de respostas.</p>	<p>abordagem pode ser aplicada em repositórios institucionais para melhorar a interação com usuários, oferecendo respostas mais relevantes e precisas com base nos contextos de consulta. Ao utilizar a agregação hierárquica de múltiplas representações, os repositórios institucionais podem otimizar a qualidade das interações, proporcionando uma experiência mais eficaz e satisfatória aos usuários.</p>
2019	Natural language processing applications in library and information science	<p>O estudo busca identificar as aplicações de processamento de linguagem natural (NLP) no campo da ciência da informação e biblioteconomia, bem como analisar a evolução histórica dessas aplicações.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo é fornecer um guia baseado em bibliometria e análises de redes sociais para pesquisadores interessados em estudar NLP nesse contexto, identificando subcampos de LIS onde o NLP pode ser utilizado.</p>	<p>O estudo revela as categorias de assuntos das publicações, palavras-chave frequentemente utilizadas e as relações entre essas palavras. Além disso, classifica os principais periódicos e artigos tematicamente para orientar pesquisadores que pretendem aplicar NLP em suas pesquisas em LIS.</p>	<p>Ao identificar as principais aplicações de NLP em LIS e os periódicos-chave na área, os repositórios podem priorizar a indexação e disponibilização de conteúdos relevantes, facilitando o acesso e a descoberta de informações para os usuários.</p>
2019	PAVAL: a location-aware virtual personal assistant for retrieving geolocated points of interest and location-based services	<p>A lacuna identificada no estudo que motivou a pesquisa sobre o desenvolvimento do assistente virtual PAVAL é a necessidade de um sistema que possa compreender as necessidades de informação dos usuários, incluindo referências</p>	<p>O objetivo do estudo é apresentar o framework PAVAL, um assistente virtual pessoal sensível à localização, projetado para sugerir pontos de interesse locais e serviços com base em consultas em linguagem natural dos usuários.</p>	<p>O resultado da pesquisa foi a implementação e validação do framework PAVAL, um assistente virtual baseado em localização que é capaz de compreender as necessidades de informação dos usuários, estimar múltiplos níveis de</p>	<p>O estudo sobre o framework PAVAL pode ser aplicado em repositórios institucionais para melhorar a busca e recuperação de informações georreferenciadas. A utilização de técnicas de Processamento de Linguagem Natural (NLP) e</p>

		<p>geográficas implícitas em suas consultas em linguagem natural, para sugerir pontos de interesse locais e serviços de forma mais precisa e contextualizada do que as soluções existentes, como assistentes virtuais gerais e sistemas de recomendação tradicionais.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Informação</p>		<p>intenções geográficas em consultas em linguagem natural e fornecer recomendações precisas de pontos de interesse locais e serviços georreferenciados. O PAVAL demonstrou um desempenho superior em comparação com outros assistentes virtuais populares em termos de precisão e recall interpolados, destacando sua eficácia na compreensão e atendimento às necessidades dos usuários de forma contextualizada e georreferenciada.</p>	<p>semântica pode aprimorar a compreensão das necessidades dos usuários e a contextualização das consultas, resultando em recomendações mais precisas de pontos de interesse locais e serviços baseados em localização. Isso pode beneficiar os repositórios institucionais ao tornar a busca por informações mais eficiente e relevante para os usuários, especialmente em ambientes onde a localização é um fator importante.</p>
2019	<p>Query expansion based on clustering and personalized information retrieval</p>	<p>O problema de pesquisa abordado é a dificuldade enfrentada pelos sistemas de recuperação de informações em extrair informações relevantes para os usuários, especialmente quando os dados processados são textos. Isso se deve à natureza complexa das bases de dados de texto. O artigo propõe soluções para melhorar a precisão e o tempo computacional desses sistemas, incluindo otimizações nas expansões de consultas e na reestruturação de conjuntos de dados, visando lidar com a grande quantidade de dados não estruturados disponíveis atualmente.</p> <p><b>Categoria:</b> Recursos de informação</p>	<p>O objetivo é apresentar um modelo formal para organizar dados e um novo algoritmo de busca para navegá-los. O artigo propõe melhorias nas técnicas de expansão de consultas e reestruturação de conjuntos de dados para aprimorar a recuperação de informações, avaliando o impacto dessas otimizações no desempenho do sistema de recuperação de informações. O objetivo final é alcançar altos níveis de precisão e eficiência computacional na recuperação de informações relevantes para os usuários.</p>	<p>Apresenta resultados satisfatórios após a implementação das otimizações propostas, a avaliação formal do impacto de cada otimização demonstrou um aumento significativo no desempenho do sistema de recuperação de informações. Os resultados obtidos mostraram que a precisão mais alta alcançada foi de 96,2% e a recall de 99,2%. Esses resultados indicam que as melhorias nas técnicas de expansão de consultas e reestruturação de conjuntos de dados contribuíram para aprimorar a eficácia do sistema, fornecendo informações mais relevantes para os usuários em um tempo computacional razoável.</p>	<p>O estudo sobre expansão de consultas com base em clustering e recuperação de informações personalizada pode ser aplicado em repositórios institucionais para aprimorar a precisão e eficiência na busca de informações relevantes. Ao utilizar técnicas avançadas de processamento de texto, classificação de termos, clustering de documentos e personalização da recuperação de informações, é possível melhorar a organização e a busca de dados nos repositórios. A implementação dessas estratégias pode resultar em uma recuperação mais precisa e rápida de documentos relevantes para os usuários, contribuindo para uma melhor experiência de busca e acesso à</p>

		<b>Subcategoria:</b> Informação			informação nos repositórios institucionais.
2019	SLTFNet: A spatial and language-temporal tensor fusion network for video moment retrieval	<p>O problema de pesquisa abordado neste trabalho é o da recuperação de momentos temporais em vídeos em resposta a consultas descritivas de linguagem natural. O objetivo é identificar pontos específicos de início e fim dentro de um vídeo que correspondam à descrição fornecida na consulta. O desafio principal é reconhecer e localizar atividades humanas dinâmicas ou interações entre objetos em vídeos longos, levando em consideração a complexidade das cenas e a necessidade de identificar objetos e ações relevantes mencionados na consulta.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>O objetivo deste trabalho é desenvolver um modelo inovador chamado SLTFNet (Spatial and Language-Temporal Tensor Fusion network) para a recuperação de momentos em vídeos, integrando informações espaciais e temporais de linguagem. O foco principal é aprimorar a precisão na localização de segmentos temporais específicos em vídeos em resposta a consultas de linguagem natural, permitindo a identificação de momentos relevantes que correspondam às descrições fornecidas. O SLTFNet visa superar os desafios de reconhecer objetos e interações relevantes, compreender informações cruciais das consultas e enfatizar palavras-chave para a recuperação precisa de momentos desejados nos vídeos.</p>	<p>O estudo propõe o SLTFNet, um modelo inovador para a recuperação de momentos em vídeos, que combina informações espaciais e temporais de linguagem. O SLTFNet superou significativamente métodos basais, melhorando a precisão na localização de segmentos temporais em vídeos em resposta a consultas de linguagem natural. A introdução da rede de atenção espacial e a fusão de tensores foram destacadas como contribuições importantes. Experimentos em conjuntos de dados mostraram melhorias significativas no desempenho, e a disponibilização do conjunto de dados e implementação do modelo para a comunidade foi enfatizada como uma contribuição adicional.</p>	<p>O estudo do SLTFNet pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de informações em vídeos por meio de consultas de linguagem natural. A integração de informações espaciais e temporais de linguagem, juntamente com a atenção espacial e a fusão de tensores, pode melhorar a precisão na localização de momentos específicos em vídeos, facilitando a recuperação de conteúdo relevante para os usuários dos repositórios institucionais. Isso pode tornar a busca por vídeos mais eficiente e precisa, permitindo aos usuários encontrar rapidamente o conteúdo desejado com base em descrições em linguagem natural.</p>
2019	Survey of Temporal Information Extraction	<p>O problema de pesquisa abordado é a extração de informações temporais de documentos escritos em linguagem natural. O estudo investiga métodos, questões e desafios relacionados à extração</p>	<p>O objetivo do estudo é relatar as pesquisas existentes sobre os métodos de extração de informações temporais, discutir questões relacionadas e fornecer direções futuras para a pesquisa nesse campo. O</p>	<p>A pesquisa resultou em uma revisão abrangente das abordagens de extração de informações temporais, incluindo a evolução das linguagens de anotação e tarefas compartilhadas. O</p>	<p>Ao implementar métodos eficazes de extração de dados temporais, os repositórios podem oferecer recursos avançados de busca por datas, eventos e relações temporais, facilitando a localização de</p>

		<p>de informações temporais, com o objetivo de desenvolver técnicas mais eficazes para essa tarefa específica.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>estudo visa identificar abordagens eficazes para extrair diferentes tipos de informações temporais de documentos, a fim de melhorar a compreensão e utilização dessas informações em diversas aplicações, como sistemas de perguntas e respostas, sistemas de recuperação de informações e sistemas de recomendação.</p>	<p>estudo identificou questões de pesquisa, delineou direções futuras no campo e enfatizou a relevância da extração precisa de informações temporais para várias aplicações. Além disso, destacou a importância da avaliação rigorosa dos sistemas dedicados a essa tarefa.</p>	<p>informações relevantes ao longo do tempo, isso pode aprimorar a experiência do usuário, permitindo uma navegação mais eficiente e uma compreensão mais profunda do contexto temporal dos documentos armazenados nos repositórios institucionais.</p>
2019	SWFQA Semantic Web Based Framework for Question Answering	<p>O problema de pesquisa abordado no artigo é a necessidade de desenvolver um framework baseado em web semântica para sistemas de perguntas e respostas que possa processar perguntas dos usuários com contexto apropriado e fornecer respostas precisas e relevantes. O objetivo é superar os desafios de responder a perguntas em linguagem natural, especialmente aquelas que possuem ambiguidade inerente e estão relacionadas a um domínio amplo e não restrito. O artigo propõe um framework que associa contexto às perguntas dos usuários e realiza validação assistida pelo usuário para melhorar a classificação das respostas para interações futuras.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de</p>	<p>O objetivo do artigo é propor um framework baseado em web semântica para sistemas de perguntas e respostas que possa processar perguntas dos usuários, associar contexto apropriado a essas perguntas e fornecer respostas relevantes e precisas em tempo real. O framework visa lidar com a complexidade das perguntas em linguagem natural, especialmente aquelas relacionadas a um domínio amplo e não restrito, como o domínio de e-learning. Além disso, o objetivo é realizar uma avaliação experimental do desempenho do sistema proposto e fornecer uma revisão comparativa com outros frameworks de perguntas e respostas existentes.</p>	<p>Os resultados da pesquisa apresentados no artigo demonstram que o framework proposto, baseado em web semântica para sistemas de perguntas e respostas, obteve resultados promissores em termos de processamento de perguntas dos usuários, associação de contexto apropriado e fornecimento de respostas relevantes e precisas em tempo real. A avaliação experimental do desempenho do sistema, utilizando um conjunto de dados do domínio de e-learning, mostrou que o framework foi capaz de responder às perguntas dos usuários de forma eficaz. Além disso, o sistema foi capaz de realizar validação assistida pelo usuário para melhorar a classificação das respostas para interações futuras. Esses resultados indicam que o framework proposto tem potencial para melhorar a</p>	<p>O estudo sobre o framework baseado em web semântica para sistemas de perguntas e respostas pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações e o suporte ao usuário. Ao integrar a tecnologia de processamento de linguagem natural e web semântica, os repositórios institucionais podem oferecer respostas mais precisas e relevantes às consultas dos usuários, melhorando a experiência de busca e recuperação de informações. Além disso, a associação de contexto às perguntas dos usuários e a validação assistida pelo usuário podem aprimorar a classificação e relevância das respostas fornecidas, contribuindo para uma melhor interação e utilização dos repositórios institucionais.</p>

		informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA		qualidade e a eficiência dos sistemas de perguntas e respostas baseados em web semântica.	
2019	Text Filtering through Multi-Pattern Matching: A Case Study of Wu-Manber-Uy on the Language of Uyghur	O problema de pesquisa abordado no estudo é a adaptação do algoritmo Wu-Manber para lidar com os desafios linguísticos específicos da língua Uyghur, como a fraqueza de vogais e o desalinhamento semântico. A pesquisa visa desenvolver o algoritmo Wu-Manber-Uy, uma variante dedicada a filtrar textos em Uyghur de forma eficiente, superando as dificuldades encontradas ao aplicar o algoritmo original a essa língua.  <b>Categoria:</b> Recursos de informação  <b>Subcategoria:</b> Documento	O objetivo do estudo é melhorar a eficácia da filtragem de texto em Uyghur, desenvolvendo e aplicando o algoritmo Wu-Manber-Uy. Esse algoritmo adaptado visa lidar com as particularidades linguísticas do Uyghur, como a fraqueza de vogais e desalinhamentos semânticos, para aprimorar o processo de correspondência de padrões e filtragem de texto nesse idioma específico. Além disso, o estudo busca demonstrar a viabilidade e eficácia do Wu-Manber-Uy em comparação com o algoritmo original, destacando as melhorias alcançadas na filtragem de texto em Uyghur.	Os resultados da pesquisa demonstraram que o algoritmo adaptado Wu-Manber-Uy foi eficaz na filtragem de texto em Uyghur, superando as limitações do algoritmo original ao lidar com as complexidades linguísticas específicas desse idioma. Os experimentos realizados mostraram um desempenho consistente e de alta qualidade do Wu-Manber-Uy, destacando sua eficiência na correspondência de padrões e filtragem de texto em Uyghur. Além disso, o estudo ressaltou a importância de considerar abordagens como dicionários e o uso de Word2vec para expandir os padrões de filtragem, contribuindo para a melhoria contínua do sistema de filtragem de texto em Uyghur.	O algoritmo Wu-Manber-Uy, ao ser aplicado em repositórios institucionais com documentos em Uyghur, pode aprimorar a eficiência da busca e recuperação de informações. Essa aplicação envolve a filtragem de conteúdo, organização dos documentos, melhoria na busca, atualizações regulares do algoritmo e implementação de recursos avançados. Ao adaptar o algoritmo para identificar padrões específicos no texto em Uyghur, é possível criar um ambiente mais preciso e eficaz, contribuindo para uma melhor organização e acesso a informações específicas nos repositórios institucionais.
2019	The state-of-the-art in expert recommendation systems	O problema de pesquisa abordado no documento é a análise e avaliação dos sistemas de recomendação de especialistas, com foco nas técnicas, conjuntos de dados, métodos de avaliação e desafios enfrentados por esses sistemas. O estudo visa compreender as	O objetivo do estudo é fornecer uma visão abrangente e crítica do estado-da-arte dos sistemas de recomendação de especialistas, destacando suas vantagens e desvantagens, explorando métricas de avaliação, e delineando desafios existentes e direções	Os resultados da pesquisa incluem uma revisão abrangente dos sistemas de recomendação de especialistas, uma classificação dos métodos utilizados nesses sistemas, uma análise das vantagens e desvantagens de cada abordagem, uma descrição das	O estudo sobre sistemas de recomendação de especialistas propõe a aplicação dessas abordagens em repositórios institucionais, visando aprimorar a identificação e recomendação de especialistas em áreas específicas do conhecimento. Ao implementar as técnicas

		<p>características dos sistemas de recomendação de especialistas, identificar suas vantagens e desvantagens, e propor direções futuras de pesquisa para superar os desafios existentes na área.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>promissoras para pesquisas futuras na área. O estudo busca consolidar o conhecimento atual sobre os sistemas de recomendação de especialistas e oferecer insights para pesquisadores e profissionais interessados nesse campo.</p>	<p>métricas de avaliação empregadas e uma visão sobre os desafios atuais e futuras direções de pesquisa na área. O estudo fornece uma compreensão aprofundada do estado-da-arte dos sistemas de recomendação de especialistas e serve como um recurso valioso para pesquisadores e profissionais interessados nesse campo.</p>	<p>discutidas, os repositórios podem aperfeiçoar a descoberta de especialistas, facilitando a colaboração acadêmica e fomentando a troca de conhecimento entre pesquisadores. A utilização de métricas de avaliação adequadas é crucial para medir a eficácia desses sistemas, contribuindo para a melhoria contínua de suas funcionalidades.</p>
<p>2020</p>	<p>A building regulation question answering system: A deep learning methodology</p>	<p>O problema de pesquisa abordado neste estudo é a ineficiência e a dificuldade enfrentada por engenheiros e profissionais similares ao tentar acessar informações regulatórias específicas para garantir que um edifício esteja em conformidade com os padrões especificados. Apesar da disponibilidade de mecanismos de busca e bancos de dados digitais, muitas vezes é desafiador para os usuários recuperar informações relevantes de forma rápida e precisa, resultando em um processo demorado de navegação e filtragem de informações. Assim, o objetivo é desenvolver uma metodologia robusta que integre recuperação de informações com um modelo de Processamento de Linguagem Natural (PLN) baseado em deep learning para</p>	<p>O objetivo deste estudo é desenvolver uma metodologia que combine Processamento de Linguagem Natural (PLN) e deep learning para fornecer respostas precisas e rápidas às perguntas dos usuários a partir de uma coleção de regulamentações de construção. A metodologia visa melhorar a eficiência e eficácia na recuperação de consultas relacionadas às regulamentações de construção, integrando a recuperação de informações com um modelo de PLN. Além disso, o estudo propõe a criação de um sistema de perguntas e respostas (chatbot) para auxiliar na recuperação de informações regulatórias de forma mais eficiente.</p>	<p>Os resultados da pesquisa demonstraram que a aplicação do modelo BERT para responder a perguntas sobre regulamentações de construção obteve um desempenho significativamente bom, com uma precisão de EM = 87,4 e F1 = 93,2. Esses resultados superaram o desempenho humano na mesma tarefa. A utilização do modelo BERT mostrou-se eficaz sem a necessidade de incorporar fontes de conhecimento específico do domínio ou realizar engenharia de recursos humanos pesada. Além disso, o estudo destacou que o modelo BERT foi mais preciso ao responder a textos curtos em comparação com textos mais longos, onde a probabilidade de perda de informações durante o processo de extração de respostas aumenta.</p>	<p>O estudo propõe a implementação de uma metodologia que utiliza uma combinação de Processamento de Linguagem Natural (PLN) e deep learning para aprimorar a eficiência na recuperação de informações regulatórias de construção. Essa abordagem pode ser aplicada em repositórios institucionais, facilitando o acesso e a busca por informações específicas em documentos regulatórios. A integração de um modelo de PLN baseado em deep learning possibilita uma recuperação mais ágil de consultas relacionadas às regulamentações de construção, permitindo que os usuários encontrem respostas precisas de forma rápida e eficaz. O <b>deep learning</b>, uma subárea do machine learning, utiliza redes neurais artificiais para aprender</p>

		<p>melhorar a eficiência e eficácia na recuperação de consultas relacionadas às regulamentações de construção.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>			<p>representações hierárquicas de dados, sendo amplamente empregado em diversas aplicações devido à sua capacidade de lidar com grandes volumes de dados e aprender representações significativas automaticamente.</p>
<p>2020</p>	<p>A Deep Learning Architecture for Psychometric Natural Language Processing</p>	<p>O problema de pesquisa é a extração de dimensões psicométricas ricas a partir de textos gerados pelos usuários. A proposta do PyNDA, uma arquitetura de aprendizado profundo, é superar as limitações dos métodos tradicionais e das arquiteturas de aprendizado profundo existentes na extração dessas dimensões psicométricas, tornando-se uma ferramenta valiosa para análises psicométricas em textos do mundo real.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo do estudo é desenvolver e apresentar uma arquitetura de aprendizado profundo chamada PyNDA, projetada especificamente para extrair dimensões psicométricas complexas e ricas, como confiança, ansiedade e alfabetização, a partir de textos gerados pelos usuários. A intenção é superar as limitações dos métodos tradicionais e das arquiteturas de aprendizado profundo existentes, demonstrando a eficácia do PyNDA na extração precisa e eficiente dessas dimensões psicométricas para diversas aplicações do mundo real.</p>	<p>A pesquisa revelou que a arquitetura de aprendizado profundo PyNDA superou significativamente os classificadores tradicionais baseados em características e as arquiteturas de aprendizado profundo existentes na extração de dimensões psicométricas complexas de textos gerados pelos usuários. Cada componente do PyNDA contribuiu significativamente para seu desempenho, conforme evidenciado pela análise de ablação. Os experimentos em conjuntos de dados do mundo real, abrangendo 11 dimensões psicométricas, validaram a eficácia da arquitetura proposta para análises psicométricas detalhadas e precisas. Esses resultados têm implicações importantes para sistemas de extração e recuperação de</p>	<p>O estudo sobre a arquitetura de aprendizado profundo PyNDA para processamento de linguagem natural psicométrica pode ser aplicado em repositórios institucionais para analisar e extrair informações valiosas dos textos disponíveis. Ao utilizar o PyNDA, os repositórios podem identificar e medir dimensões psicométricas complexas, como confiança, ansiedade e alfabetização, presentes nos documentos. Isso possibilita uma análise mais aprofundada e precisa do conteúdo, permitindo uma melhor compreensão dos usuários e suas interações com os materiais do repositório. Essa abordagem inovadora pode enriquecer a recuperação de informações e a modelagem de comportamento nos repositórios institucionais, contribuindo para uma experiência mais</p>

				informações centrados no usuário, especialmente aqueles que buscam medir e incorporar dimensões psicométricas em suas análises.	personalizada e eficaz para os usuários.
2020	A History and Theory of Textual Event Detection and Recognition	<p>O problema de pesquisa abordado neste estudo é a detecção e reconhecimento de eventos em textos. Os autores exploram como identificar eventos significativos em grandes volumes de dados textuais, bem como reconhecer e extrair informações relevantes desses eventos. Eles buscam aprimorar as técnicas existentes de processamento de texto para melhorar a detecção e reconhecimento de eventos, contribuindo assim para avanços na área de análise de texto e recuperação de informações.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é fornecer uma visão abrangente sobre a detecção e reconhecimento de eventos textuais. Os pesquisadores buscam explorar as metodologias para construir eventos para exploração de informações, dividindo-os em três tipos: eventos documentais, eventos de quadro e eventos gráficos. Eles se propõem a analisar e discutir as diferentes abordagens e técnicas utilizadas para identificar eventos em textos, visando melhorar a eficácia da recuperação de informações e do reconhecimento de eventos em grandes conjuntos de dados textuais.</p>	<p>Neste estudo, os pesquisadores desenvolveram metodologias inovadoras para a detecção e reconhecimento de eventos em textos, utilizando redes neurais para acelerar o processo e modelos de aprendizado profundo para detectar subeventos. Métricas de avaliação, como a matriz de confusão e a curva de troca de erros de decisão, foram empregadas para avaliar a eficácia dos sistemas de detecção de eventos. A importância da organização hierárquica de eventos foi ressaltada em contraste com a estrutura plana de partição de corpus. Esses resultados são relevantes para avançar a área de detecção de eventos em textos, oferecendo insights cruciais para lidar com grandes volumes de dados textuais em diversas aplicações, como aquisição de informações, recomendação colaborativa, gerenciamento de conhecimento e tomada de decisões.</p>	<p>O estudo sobre detecção e reconhecimento de eventos textuais pode ser aplicado em repositórios institucionais para melhorar a organização e recuperação de informações. Ao utilizar as metodologias e técnicas propostas no estudo, os repositórios podem aprimorar a detecção de eventos relevantes em documentos, facilitando a navegação e busca por conteúdo específico. Isso pode beneficiar pesquisadores, estudantes e profissionais que buscam acessar e utilizar informações de forma mais eficiente e eficaz. Além disso, a organização hierárquica de eventos pode proporcionar uma visão mais estruturada e abrangente dos dados armazenados nos repositórios, facilitando a identificação de padrões e tendências ao longo do tempo.</p>
2020	A New Hybrid Improved Method for Measuring Concept	<p>O problema de pesquisa abordado neste estudo é a necessidade de melhorar a</p>	<p>O objetivo do estudo é propor um novo método híbrido aprimorado para medir a similaridade</p>	<p>A pesquisa propôs um novo método híbrido para melhorar a medição da similaridade</p>	<p>O estudo pode ser aplicado em repositórios institucionais para melhorar a precisão na</p>

	<p>Semantic Similarity in WordNet</p>	<p>medição da similaridade semântica entre conceitos em WordNet, especialmente em relação à distinção de similaridades entre nós de um nó para diferentes nós do mesmo nível.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>semântica entre conceitos, superando as limitações do método de Wu e Palmer.</p>	<p>semântica entre conceitos em WordNet, visando resolver as limitações identificadas no método de Wu e Palmer. Os resultados indicaram a eficácia do método proposto em distinguir com precisão as similaridades entre nós, tanto de um nó para diferentes nós do mesmo nível quanto entre outros contextos semânticos.</p>	<p>recuperação de informações e no gerenciamento do conhecimento, permitindo uma análise mais aprofundada da similaridade semântica entre conceitos armazenados nos repositórios. Isso pode facilitar a organização e a recuperação de informações relevantes de forma mais eficiente e precisa.</p>
<p>2020</p>	<p>A Persian Medical Question Answering System</p>	<p>O problema de pesquisa abordado no artigo é o desenvolvimento de um sistema de Perguntas e Respostas (QA) médico em persa. O objetivo é criar um sistema capaz de responder a perguntas relacionadas a doenças e medicamentos na língua persa, utilizando técnicas avançadas de processamento de linguagem natural e recuperação de documentos. O sistema visa melhorar a acessibilidade e a precisão na busca de informações médicas em persa, atendendo à demanda por respostas rápidas e precisas em linguagem natural na área da saúde.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade</p>	<p>O objetivo do estudo é desenvolver um sistema de Perguntas e Respostas (QA) médico em persa que possa responder a perguntas relacionadas a doenças e medicamentos. O sistema visa fornecer respostas precisas e rápidas em linguagem natural na área da saúde, atendendo às necessidades de médicos, pacientes e prestadores de serviços de saúde que buscam informações médicas em persa. Além disso, o estudo tem como objetivo coletar e estruturar um conjunto de dados de doenças e medicamentos em persa, bem como desenvolver ferramentas de processamento de linguagem específicas para o sistema de QA médico em persa.</p>	<p>O resultado da pesquisa foi a implementação de um sistema de Perguntas e Respostas (QA) médico em persa, que foi avaliado com sucesso usando métricas variadas em um conjunto de testes projetado. O sistema demonstrou uma precisão de 83,6% ao responder a 500 perguntas de amostra sobre doenças e medicamentos. O sistema foi estruturado em três módulos principais: processamento de perguntas, recuperação de documentos e extração de respostas. Durante o estudo, um conjunto de dados contendo documentos médicos relacionados a doenças e medicamentos em persa foi preparado, totalizando 1767 documentos semi-estruturados. O sistema mostrou potencial para responder a uma variedade de perguntas médicas em persa, contribuindo para a</p>	<p>O estudo sobre o sistema de Perguntas e Respostas (QA) médico em persa pode ser aplicado em repositórios institucionais de saúde para melhorar a acessibilidade e precisão na busca de informações médicas em persa. Ao integrar o sistema de QA em repositórios institucionais, os usuários poderão fazer perguntas em linguagem natural sobre doenças e medicamentos, obtendo respostas rápidas e precisas a partir dos documentos disponíveis. Isso facilitaria a recuperação de informações médicas específicas em persa, beneficiando médicos, pacientes e prestadores de serviços de saúde que utilizam os repositórios institucionais como fonte de conhecimento médico.</p>

		percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA		acessibilidade e precisão na busca de informações médicas na língua persa.	
2020	A Pseudo-relevance feedback framework combining relevance matching and semantic matching for information retrieval	<p>O problema de pesquisa abordado neste trabalho é a lacuna semântica entre a intenção da consulta e a representação da consulta, bem como entre a representação da consulta e a representação do documento em recuperação de informações. A dificuldade de um computador entender com precisão a intenção real da consulta do usuário pode levar a desvios semânticos entre os documentos e a consulta. O objetivo é desenvolver um framework de feedback de pseudo-relevância que combine correspondência de relevância e correspondência semântica para abordar esses problemas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>O principal objetivo deste trabalho é fornecer um framework de feedback de pseudo-relevância que combine correspondência de relevância e correspondência semântica para melhorar a qualidade dos documentos de feedback. Os objetivos específicos incluem determinar a importância da correspondência de relevância e correspondência semântica para a recuperação de informações, bem como reduzir a lacuna semântica entre a intenção da consulta e sua representação, e entre a representação da consulta e a representação do documento na recuperação de informações.</p>	<p>Os resultados da pesquisa mostraram que o modelo proposto superou os modelos de linha de base robustos em termos de precisão média (MAP) e precisão P na posição 10 (P@10) em quatro conjuntos de dados da Text Retrieval Conference (TREC). Além disso, os resultados destacaram que o uso do método combinado de correspondência de relevância e correspondência semântica é mais eficaz do que o uso de correspondência de relevância ou correspondência semântica isoladamente para melhorar a qualidade dos documentos de feedback.</p>	<p>O estudo propõe um framework de feedback de pseudo-relevância que combina correspondência de relevância e correspondência semântica para melhorar a recuperação de informações. Esse framework pode ser aplicado em repositórios institucionais para aprimorar a precisão e relevância dos resultados de pesquisa. Ao integrar a correspondência de relevância e semântica, o modelo pode reduzir a lacuna semântica entre consultas e documentos, resultando em uma recuperação mais eficaz e relevante de informações nos repositórios institucionais.</p>

<p>2020</p>	<p>A Semantic Based Approach for Topic Evaluation in Information Filtering</p>	<p>O problema de pesquisa abordado neste estudo é a avaliação da qualidade de modelos de tópicos em sistemas de filtragem de informações. O foco está em garantir que as palavras tópicas sejam semanticamente significativas, evitando palavras tópicas que não contribuem para a representação precisa dos tópicos. A pesquisa visa melhorar a precisão dos modelos de tópicos gerados a partir de documentos do usuário para aprimorar os sistemas de filtragem de informações.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>O objetivo deste estudo é propor um novo modelo de avaliação de tópicos que utiliza técnicas semânticas para avaliar a qualidade dos tópicos em um modelo de tópicos. Especificamente, o estudo visa avaliar a qualidade dos tópicos por meio da correspondência das palavras tópicas com conceitos em uma ontologia, utilizando técnicas de incorporação de palavras para lidar com palavras tópicas ambíguas ou sem sentido. O modelo proposto é destinado a melhorar a eficácia dos sistemas de filtragem de informações que utilizam modelos de tópicos para representar o interesse do usuário em informações.</p>	<p>Os resultados da pesquisa indicaram que o modelo proposto superou os modelos de linha de base do estado-da-arte em termos de representações de interesse do usuário baseadas em termos, frases e tópicos, bem como em modelos de avaliação de tópicos. A avaliação comparativa demonstrou que o novo modelo proposto teve um desempenho superior aos modelos existentes, destacando sua eficácia na avaliação e melhoria da qualidade dos tópicos em modelos de tópicos utilizados em sistemas de filtragem de informações.</p>	<p>O estudo propõe um modelo de avaliação de tópicos que utiliza técnicas semânticas para melhorar a qualidade dos modelos de tópicos em sistemas de filtragem de informações. Essa abordagem pode ser aplicada em repositórios institucionais para aprimorar a representação do interesse do usuário em informações, garantindo que os tópicos gerados sejam semanticamente significativos. Ao utilizar a correspondência de palavras tópicas com conceitos em uma ontologia e técnicas de incorporação de palavras, é possível melhorar a precisão da representação do interesse do usuário nos repositórios institucionais, resultando em uma filtragem mais eficaz e relevante de documentos para os usuários.</p>
<p>2020</p>	<p>A syntactic path-based hybrid neural network for negation scope detection</p>	<p>O problema de pesquisa é a detecção do escopo de negação em textos, uma tarefa crucial em processamento de linguagem natural. O objetivo é desenvolver um modelo híbrido de rede neural que possa identificar com precisão as regiões de um texto que são afetadas por negações,</p>	<p>O objetivo do estudo é propor uma abordagem inovadora para a detecção do escopo de negação em textos utilizando um modelo de rede neural híbrida baseado em caminhos sintáticos. Especificamente, o objetivo é desenvolver um sistema que possa identificar</p>	<p>O estudo obteve resultados promissores na detecção do escopo de negação, alcançando um F-score de 90.82% (78.31% PCS) na sub-corpus de resumos do Bioscope, superando abordagens que dependem de características manualmente codificadas. Esses resultados</p>	<p>O estudo propõe um modelo de rede neural híbrida para detecção de escopo de negação em textos, que pode ser aplicado em repositórios institucionais para melhorar a análise e compreensão de conteúdos textuais. Ao utilizar informações sintáticas e contextuais para</p>

		<p>contribuindo para a compreensão e análise de sentenças negativas em diferentes contextos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>com eficácia se um token em uma sentença está dentro ou fora do escopo de negação, sem depender de intervenção humana. A combinação de uma rede Bi-LSTM bidirecional e uma rede CNN permite capturar informações relevantes para determinar a presença de negação, superando abordagens dependentes de características específicas.</p>	<p>demonstram a eficácia do modelo proposto, que combina redes neurais e informações sintáticas, sem a necessidade de engenharia de características manual, destacando-se como uma alternativa poderosa para lidar com a detecção de escopo de negação em textos de diferentes domínios.</p>	<p>identificar regiões afetadas por negações, o modelo pode auxiliar na extração de informações relevantes e na categorização de conteúdo, contribuindo para uma indexação mais precisa e uma recuperação de informações mais eficiente em repositórios institucionais.</p>
2020	<p>An encoder–decoder approach to mine conditions for engineering textual data</p>	<p>O problema de pesquisa abordado neste artigo é a mineração de condições a partir de dados textuais escritos em linguagem natural. O objetivo é desenvolver um método que possa identificar e extrair condições que influenciam fatos e opiniões presentes nos textos, sem a necessidade de padrões definidos pelo usuário ou recursos específicos. A proposta visa preencher uma lacuna científica ao demonstrar a viabilidade de extrair condições de textos em linguagem natural de forma eficaz e precisa, utilizando abordagens de redes neurais profundas.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo deste estudo é propor um novo método de mineração de condições que supere as limitações dos métodos existentes, utilizando uma abordagem de codificador-decodificador baseada em redes neurais profundas. O método proposto busca extrair condições que influenciam fatos e opiniões presentes em dados textuais em linguagem natural. O objetivo é demonstrar que o novo método pode alcançar um desempenho significativamente melhor em termos de pontuação F1 em comparação com os métodos existentes. Além disso, o estudo visa realizar uma análise abrangente de um grande conjunto de dados multilíngue e multi-tópico para avaliar a eficácia do método proposto.</p>	<p>Os resultados da pesquisa indicam que o novo método proposto para mineração de condições utilizando uma abordagem de codificador-decodificador baseada em redes neurais profundas superou os métodos existentes em termos de pontuação F1. A análise experimental realizada em um grande conjunto de dados com 4,7 milhões de sentenças em 16 tópicos comuns em quatro idiomas (inglês, espanhol, francês e italiano) demonstrou que o método proposto apresentou uma melhoria significativa na pontuação F1 em comparação com os métodos existentes. Além disso, a análise abrangente do conjunto de dados revelou que as condições extraídas apresentam uma distribuição de conectivos de cauda longa e são</p>	<p>O estudo propõe um método inovador de mineração de condições em dados textuais utilizando redes neurais profundas, que pode ser aplicado em repositórios institucionais para extrair e analisar condições que influenciam fatos e opiniões presentes nos documentos. Essa abordagem avançada pode melhorar a extração de conhecimento e insights a partir dos textos armazenados nos repositórios, contribuindo para uma análise mais precisa e eficaz dos dados disponíveis.</p>

				semanticamente bastante diferentes.	
2020	An end-to-end pseudo relevance feedback framework for neural document retrieval	<p>O problema de pesquisa abordado neste estudo é a melhoria do desempenho dos modelos tradicionais de recuperação de informações (IR) por meio do uso de feedback de relevância pseudo (PRF) em conjunto com arquiteturas neurais. O objetivo é reduzir o impacto das inconsistências de vocabulário entre consulta e documento, visando aprimorar a eficácia da recuperação de documentos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>O objetivo deste estudo é propor um framework neural de feedback de relevância pseudo (NPRF) que possibilite a integração do PRF com modelos neurais de IR existentes. O NPRF visa expandir a consulta inicial considerando os documentos mais bem classificados, permitindo uma interpretação mais abrangente da consulta e melhorando o desempenho da recuperação de documentos neurais.</p>	<p>Os resultados da pesquisa indicam que o framework NPRF proposto é capaz de melhorar significativamente o desempenho de três modelos neurais de IR de ponta em dois conjuntos de dados de referência amplamente utilizados. Além disso, o NPRF demonstrou melhorar o desempenho dos modelos neurais em diferentes tipos de consultas, como consultas curtas e consultas verbais em linguagem natural. A análise também revelou a robustez do framework NPRF em várias configurações de parâmetros, e as curvas de aprendizado mostraram que o NPRF melhora a representação da necessidade de informação do usuário ao considerar múltiplos documentos de PRF em vez de uma única consulta.</p>	<p>O estudo sobre o framework NPRF para recuperação de documentos neurais pode ser aplicado em repositórios institucionais para melhorar a eficiência da recuperação de informações. Ao integrar o feedback de relevância pseudo com modelos neurais de IR, é possível aprimorar a precisão e relevância dos resultados de busca, facilitando a localização de documentos relevantes nos repositórios institucionais. Essa abordagem pode otimizar a experiência do usuário ao buscar informações específicas dentro desses repositórios.</p>
2020	An ensemble clustering approach for topic discovery using implicit text segmentation	<p>O problema de pesquisa abordado neste estudo é a necessidade de melhorar a descoberta de tópicos em coleções de texto multi-tópico, superando as limitações e resultados abaixo da média encontrados em estudos anteriores. Os autores destacam</p>	<p>O objetivo deste estudo é propor e validar um novo modelo de clustering denominado LDA-onto, que combina segmentação de texto e clustering para melhorar a descoberta de tópicos em coleções de texto multi-tópico. Os autores buscam demonstrar a eficácia do LDA-</p>	<p>os resultados da pesquisa demonstram que o modelo de clustering proposto, baseado no LDA-onto, é uma abordagem eficaz e inovadora para a descoberta de tópicos em coleções de texto multi-tópico, oferecendo melhorias significativas em relação às</p>	<p>O estudo propõe um modelo de clustering inovador, o LDA-onto, que combina segmentação de texto e clustering para melhorar a descoberta de tópicos em coleções de texto multi-tópico. Essa abordagem pode ser aplicada em repositórios institucionais para melhorar a</p>

		<p>a importância de gerar clusters de tópicos de alta qualidade a partir de documentos que contenham múltiplos tópicos, enfatizando a relevância temática e coerência dos clusters resultantes. Eles também apontam as deficiências de abordagens de clustering tradicionais, que tratam cada cluster como um tópico único, e a falta de exploração de palavras discriminadas nos documentos originais. Assim, o problema de pesquisa central é como desenvolver uma abordagem de clustering que segmente efetivamente documentos em diferentes partes e gere clusters de alta qualidade, superando as limitações das abordagens tradicionais.</p> <p><b>Categoria:</b> Recursos de informação  <b>Subcategoria:</b> Representação do documento</p>	<p>onto em gerar clusters de tópicos de alta qualidade, superando as abordagens tradicionais de clustering. Eles também visam validar a capacidade do LDA-onto em melhorar a segmentação de texto por meio de similaridade ontológica e semântica, resultando em clusters mais coesos e relevantes em termos de tópicos.</p>	<p>técnicas tradicionais de clustering e segmentação de texto.</p>	<p>organização e recuperação de informações. Ao utilizar o LDA-onto, os repositórios podem gerar clusters de tópicos mais coesos e relevantes, facilitando a navegação e busca por conteúdo. Isso pode aprimorar a experiência do usuário, tornando mais fácil encontrar e explorar informações em repositórios institucionais.</p>
<p>2020</p>	<p>An intelligent use of stemmer and morphology analysis for Arabic information retrieval</p>	<p>O problema de pesquisa abordado neste estudo é a melhoria da extração de raízes/prefixos em árabe para construir sistemas eficazes de recuperação de informações em árabe. Os autores propõem o método AMIR (Arabic Morphology Information Retrieval) como uma abordagem inovadora para gerar prefixos para sistemas de busca de texto em árabe. Eles buscam resolver</p>	<p>O objetivo deste estudo é melhorar a precisão e eficácia da extração de raízes e prefixos em árabe para a construção de sistemas de recuperação de informações em árabe. Os autores propõem o método AMIR como uma abordagem inteligente para gerar e extrair prefixos em árabe, aplicando um conjunto de regras que consideram a relação entre as letras árabes para encontrar a</p>	<p>O resultado da pesquisa demonstrou que o método AMIR (Arabic Morphology Information Retrieval) superou outros sistemas de extração de raízes em árabe, como LUCENE e FARASA, em termos de precisão média. O AMIR alcançou uma média de precisão de 0,34%, enquanto o LUCENE, FARASA e o sistema sem extrator de raízes obtiveram 0,27%, 0,28% e 0,21%,</p>	<p>O estudo sobre o uso inteligente de stemmer e análise de morfologia para recuperação de informações em árabe pode ser aplicado em repositórios institucionais para aprimorar a precisão e eficácia da busca de informações em textos em árabe. Ao implementar o método AMIR (Arabic Morphology Information Retrieval), os repositórios institucionais podem melhorar a extração de raízes e</p>

		<p>desafios como a ambiguidade de palavras, variações ortográficas e a rica morfologia da língua árabe, visando otimizar as pesquisas dos usuários nesse idioma.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>raiz/prefixo das palavras usadas como termos de indexação na busca de texto em árabe. A intenção é demonstrar a utilidade das regras propostas para diferentes sistemas de recuperação de informações em árabe.</p>	<p>respectivamente. Esses resultados indicam que o AMIR é capaz de melhorar a extração de raízes em árabe e aumentar o desempenho de recuperação de informações nesse idioma, sendo mais eficaz em comparação com outros métodos existentes.</p>	<p>prefixos em árabe, facilitando a indexação e recuperação de documentos nesse idioma. Isso resultaria em uma busca mais precisa e eficiente para os usuários que buscam informações em árabe nos repositórios institucionais.</p>
2020	<p>ArA*summarizer: An Arabic text summarization system based on subtopic segmentation and using an A* algorithm for reduction</p>	<p>O problema de pesquisa abordado no artigo é a falta de trabalhos dedicados à sumarização automática de textos em língua árabe que atendam às expectativas da comunidade. A escassez de esforços e pesquisas nessa área levou os autores a proporem uma abordagem inovadora que visa produzir resumos automáticos coerentes e não redundantes, utilizando técnicas de segmentação de tópicos com detecção de coesão lexical.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo do estudo é desenvolver o ArA*summarizer, um sistema automático de sumarização de documentos em língua árabe. Este sistema baseia-se em uma abordagem híbrida não supervisionada que combina técnicas estatísticas, baseadas em cluster e baseadas em grafo. O principal propósito é dividir o texto em subtópicos e selecionar as sentenças mais relevantes nos subtópicos mais pertinentes, utilizando um algoritmo A* executado em um grafo que representa as diferentes relações léxico-semânticas entre as sentenças. O sistema foi avaliado em relação a outros trabalhos existentes, demonstrando um bom desempenho.</p>	<p>Os resultados da pesquisa mostraram que o ArA*summarizer obteve um bom desempenho em comparação com outros sistemas existentes de sumarização automática de textos em língua árabe. A avaliação foi realizada utilizando métricas de avaliação de recall-oriented understudy for gisting, engenharia de sumarização automática, modelos de gráficos combinados e avaliação de gráficos n-grama. Esses resultados positivos destacam a eficácia do ArA*summarizer na produção de resumos automáticos de alta qualidade.</p>	<p>O estudo sobre o ArA*summarizer, um sistema de sumarização automática de textos em língua árabe, pode ser aplicado em repositórios institucionais para facilitar a organização e recuperação de informações. Ao utilizar técnicas avançadas de segmentação de tópicos e seleção de sentenças relevantes, o ArA*summarizer pode ajudar a criar resumos concisos e informativos de documentos, tornando mais fácil para os usuários identificarem o conteúdo essencial. Isso pode melhorar a eficiência na busca e navegação em repositórios institucionais, permitindo uma melhor utilização e aproveitamento das informações disponíveis.</p>
2020	<p>Automatic CV processing for scientific research</p>	<p>O problema de pesquisa abordado no documento é a necessidade de automatizar o processamento de currículos</p>	<p>O objetivo do estudo é automatizar o processamento de currículos para pesquisadores de diversas disciplinas em</p>	<p>Os resultados da pesquisa indicam que a combinação de processamento de linguagem natural com o classificador</p>	<p>O estudo sobre automatização do processamento de currículos para pesquisadores utilizando técnicas de processamento de</p>

	<p>using data mining algorithm</p>	<p>para pesquisadores de diferentes disciplinas pertencentes a vários laboratórios de pesquisa. A proposta visa melhorar o sistema de informação, especialmente no contexto da pesquisa científica, utilizando processamento de linguagem natural em conjunto com o classificador de mineração de dados Decision Tree para prever o campo de trabalho de cada pesquisador.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>laboratórios de pesquisa, utilizando técnicas de processamento de linguagem natural e o classificador de mineração de dados Decision Tree. O foco é prever o campo de trabalho de cada pesquisador, visando melhorar o sistema de informação e auxiliar na tomada de decisões no contexto da pesquisa científica.</p>	<p>Decision Tree proporcionou resultados promissores na automação do processamento de currículos para pesquisadores. Comparando o desempenho do Decision Tree com outros classificadores, como Naive Bayes e One Rule, verificou-se que o Decision Tree foi mais eficaz em termos de precisão, recall, F-measure, estatística Kappa, erro quadrático médio, erro absoluto relativo e erro quadrático relativo. Esses resultados sugerem que a abordagem proposta pode ser uma ferramenta valiosa para melhorar a gestão da informação e a tomada de decisões no âmbito da pesquisa científica.</p>	<p>linguagem natural e classificador Decision Tree pode ser aplicado em repositórios institucionais para facilitar a organização e recuperação de informações sobre os pesquisadores e suas áreas de atuação. Ao automatizar a classificação dos currículos com base em campos de trabalho previstos, os repositórios institucionais podem oferecer uma maneira mais eficiente de gerenciar e acessar informações relevantes sobre os pesquisadores e suas contribuições científicas. Isso pode melhorar a eficácia na busca por especialistas, na identificação de colaborações e no monitoramento do progresso da pesquisa dentro da instituição.</p>
<p>2020</p>	<p>Automatic keyphrase extraction: a survey and trends</p>	<p>O problema de pesquisa abordado no documento é a extração automática de frases-chave de documentos ou dados da web. Devido ao crescimento exponencial de dados textuais e fontes da web, a necessidade de identificar informações relevantes dentro deles levou ao desenvolvimento de técnicas de Extração Automática de Frases-Chave (AKPE). Essas técnicas são amplamente adotadas em várias aplicações de Recuperação de Informação (IR), Processamento de</p>	<p>O objetivo do documento é fornecer uma revisão abrangente das pesquisas atuais sobre a tarefa de Extração Automática de Frases-Chave (AKPE). Isso inclui destacar o processo comum, ilustrar as várias abordagens e técnicas utilizadas, examinar as complexidades abordadas e os principais desafios enfrentados pelas técnicas de AKPE. Além disso, o documento visa realizar um estudo comparativo das técnicas de melhor desempenho, discutir por que</p>	<p>O resultado da pesquisa apresentada no documento inclui uma análise abrangente do processo de Extração Automática de Frases-Chave. Durante o estudo, foram identificadas 34 técnicas existentes que demonstraram eficiência em diferentes conjuntos de dados, incluindo o conjunto de dados mais comumente utilizado, o Semeval-2010. Foram destacados vários desafios significativos em todas as etapas do processo de AKPE, como a</p>	<p>O estudo sobre Extração Automática de Frases-Chave pode ser aplicado em repositórios institucionais para melhorar a indexação e recuperação de informações. Ao implementar técnicas de AKPE, os repositórios podem automaticamente extrair frases-chave relevantes de documentos depositados, facilitando a busca e navegação pelos conteúdos. Isso pode aumentar a visibilidade e acessibilidade do material armazenado, melhorando a</p>

		<p>Linguagem Natural (NLP) e Mineração de Texto (TM), visando resolver desafios relacionados à extração de informações valiosas.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>algumas técnicas se saem melhor do que outras e propor recomendações para melhorar cada estágio da AKPE.</p>	<p>redundância causada por frases-chave semanticamente equivalentes, a falta de frases-chave latentes, o aumento do tamanho ou número de documentos e a insuficiência dos métodos de avaliação comumente utilizados. Recomendações foram propostas para abordar esses desafios e aprimorar a tarefa de AKPE em trabalhos futuros.</p>	<p>experiência dos usuários ao encontrar informações específicas dentro do repositório.</p>
2020	<p>Chatbot Design Method Using Hybrid Word Vector Expression Model Based on Real Telemarketing Data</p>	<p>O problema de pesquisa abordado no artigo é a necessidade de aprimorar os métodos de design de chatbot, especialmente em relação à representação de palavras e à compreensão de declarações contínuas, utilizando dados reais de telemarketing. O estudo propõe um modelo híbrido de expressão de vetores de palavras que visa melhorar a precisão das respostas do chatbot e a capacidade de comunicação com os humanos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo do estudo é propor um método aprimorado de design de chatbot utilizando um modelo híbrido de expressão de vetores de palavras baseado em dados reais de telemarketing. Especificamente, o objetivo é melhorar a precisão das respostas do chatbot e a capacidade de comunicação com os humanos, superando as limitações dos métodos convencionais que utilizam apenas o modelo bag-of-words.</p>	<p>O resultado da pesquisa demonstrou que a abordagem de design de chatbot utilizando o modelo híbrido de expressão de vetores de palavras baseado em dados reais de telemarketing foi capaz de melhorar a precisão das respostas do chatbot e a eficácia da comunicação com os humanos. A utilização do modelo híbrido, que combina o bag-of-words e o skip-gram, permitiu uma representação mais rica das palavras e uma melhor compreensão de declarações contínuas, superando as limitações dos métodos convencionais. Os resultados experimentais indicaram uma melhoria significativa na precisão das respostas do chatbot, tornando o sistema mais eficaz na interação com os usuários.</p>	<p>O estudo sobre o design de chatbot utilizando um modelo híbrido de expressão de vetores de palavras baseado em dados reais de telemarketing pode ser aplicado em repositórios institucionais para melhorar a interação com os usuários. Ao implementar esse modelo, os repositórios podem aprimorar a precisão das respostas fornecidas aos usuários, oferecer uma comunicação mais eficaz e personalizada, e facilitar o acesso e a busca por informações dentro do repositório. Isso pode resultar em uma experiência mais satisfatória para os usuários e em uma melhor utilização dos recursos disponíveis no repositório institucional.</p>
2020	<p>Conceptualising a Library Chatbot using</p>	<p>O problema de pesquisa abordado no documento é a</p>	<p>O objetivo do estudo é conceituar um novo chatbot de</p>	<p>O uso de software de inteligência artificial, como o Rasa, pode</p>	<p>O estudo sobre a conceptualização de um chatbot</p>

	<p>Open Source Conversational Artificial Intelligence</p>	<p>necessidade de explorar e implementar plataformas de software conversacional de inteligência artificial de código aberto, como o Rasa, nas bibliotecas. O autor destaca a falta de escalabilidade, complexidade na gestão de diálogos contextuais e interações de API compostas em chatbots existentes, bem como a necessidade de documentação profissional e facilidade de aprendizado. Além disso, a falta de independência de plataforma em muitas ferramentas de desenvolvimento de chatbots é identificada como um desafio a ser superado.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>biblioteca utilizando a plataforma de software conversacional de código aberto e baseada em inteligência artificial chamada Rasa. O autor propõe a adoção deste chatbot por bibliotecas, visando melhorar os serviços oferecidos aos usuários. O estudo também busca apresentar a essência da tecnologia de chatbots e sua aplicação atual em bibliotecas, além de ilustrar os fundamentos técnicos do Rasa Stack que podem ser aproveitados para desenvolver um chatbot de biblioteca e refletir sobre possíveis pesquisas futuras nessa área.</p>	<p>transformar as bibliotecas, melhorando interações com usuários. Embora necessite de mais pesquisa para aprimoramento, essa tecnologia pode impactar positivamente os serviços de informação e referência. A implementação requer treinamento técnico e recursos financeiros significativos. Os chatbots podem ser aplicados em diversas áreas, como entrevistas virtuais, suporte 24 horas, informação comunitária e sugestões de recursos. O estudo também destaca áreas de pesquisa futura, como lidar com negações, extração de entidades multilíngues e integração com sistemas de gerenciamento de conhecimento semântico.</p>	<p>de biblioteca usando inteligência artificial conversacional de código aberto, como o Rasa, pode ser aplicado em repositórios institucionais para melhorar a interação com os usuários. Ao implementar um chatbot baseado em tecnologias avançadas de processamento de linguagem natural e gerenciamento de conhecimento semântico, os repositórios institucionais podem oferecer suporte mais eficiente aos usuários na busca e recuperação de informações. O chatbot pode auxiliar na navegação do repositório, responder a perguntas dos usuários, fornecer recomendações personalizadas de conteúdo e facilitar o acesso a recursos relevantes. Essa abordagem inovadora pode aprimorar a experiência do usuário nos repositórios institucionais, tornando a descoberta e o uso de informações mais intuitivos e eficazes.</p>
<p>2020</p>	<p>A literature embedding model for cardiovascular disease prediction using risk factors, symptoms, and genotype information</p>	<p>O problema de pesquisa abordado neste estudo é a predição de doenças cardiovasculares, com foco na identificação de novos riscos, sintomas associados, e genótipos relevantes. A pesquisa visa avaliar a relação entre fatores genéticos,</p>	<p>O objetivo do estudo é desenvolver um modelo de incorporação de literatura para identificar fatores de risco, sintomas e genes associados a doenças cardiovasculares, visando aprimorar a predição dessas doenças, melhorar a tomada de decisão dos</p>	<p>O resultado da pesquisa foi a criação de um modelo de incorporação de literatura que identifica com precisão fatores de risco, sintomas, mecanismos e genes associados a doenças cardiovasculares a partir de consultas de palavras-chave. O modelo demonstrou uma média</p>	<p>O modelo de incorporação de literatura pode ajudar o Repositório Institucional melhorando a recuperação de informações, identificando relações entre fatores de risco e sintomas, predizendo doenças cardiovasculares, auxiliando na tomada de decisão, e</p>

		<p>sintomas e riscos para melhorar a precisão na predição de doenças cardiovasculares. Além disso, o estudo utiliza um modelo de incorporação de literatura para analisar e extrair informações relevantes de grandes volumes de dados biomédicos, como abstracts de artigos científicos, a fim de aprimorar a compreensão e a predição de doenças cardiovasculares.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>profissionais de saúde, aumentar a eficiência na análise de dados biomédicos e contribuir para a saúde pública através de medidas preventivas e terapêuticas mais eficazes.</p>	<p>de acurácia superior a 96% na identificação de fatores de risco e sintomas associados a uma palavra-chave de consulta. Além disso, o modelo apresentou um desempenho superior em tarefas de seleção de características e redução de dimensionalidade para a predição de doenças cardiovasculares em dados de coorte, com um tempo de execução mais rápido em comparação com outros métodos populares. Esses resultados indicam que o modelo desenvolvido é eficaz na identificação de informações relevantes para a predição de doenças cardiovasculares e pode contribuir significativamente para a tomada de decisões informadas e aprimoradas nessa área da saúde.</p>	<p>acelerando a análise de dados biomédicos, contribuindo para a pesquisa e prática em saúde cardiovascular.</p>
2020	Cross-modal retrieval with dual multi-angle self-attention	<p>O problema de pesquisa abordado neste artigo é a melhoria do desempenho da recuperação cruzada de informações entre modalidades diferentes, como texto e imagem, por meio da utilização de atenção dupla de múltiplos ângulos (DMASA). A pesquisa visa aprimorar a precisão e eficiência da recuperação de informações em tarefas de correspondência visual-textual, que são desafiadoras devido às</p>	<p>O objetivo deste estudo é propor um novo framework de recuperação cruzada de informações chamado Dual Multi-Angle Self-Attention (DMASA) para melhorar a precisão e eficácia da recuperação de informações em tarefas de correspondência visual-textual. O DMASA é projetado para extrair características detalhadas de diferentes ângulos em imagens e textos, permitindo a</p>	<p>O framework DMASA proposto na pesquisa apresentou sucesso ao capturar detalhes em imagens e textos, permitindo a identificação de regiões destacadas em imagens e palavras semânticas em descrições textuais de diferentes ângulos. A estratégia de treinamento em múltiplos estágios mostrou-se eficaz, resultando em um desempenho superior na recuperação cruzada de informações em</p>	<p>O estudo sobre o framework DMASA para recuperação cruzada de informações pode ser aplicado em repositórios institucionais para melhorar a precisão e eficácia da busca e recuperação de informações multimodais, como imagens e textos. Ao utilizar a atenção dupla de múltiplos ângulos, o framework pode ajudar a localizar características detalhadas em diferentes modalidades, facilitando a</p>

		<p>representações heterogêneas e à diferença entre modalidades cruzadas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>localização de regiões salientes em imagens visuais e palavras semânticas em descrições textuais. O objetivo principal é alcançar um desempenho de recuperação cruzada de informações de última geração em conjuntos de dados de referência, como Flickr8k, Flickr30k e MSCOCO.</p>	<p>comparação com abordagens tradicionais de treinamento em um único estágio. Experimentos e visualizações destacaram a eficácia dos componentes do DMASA, alcançando resultados de última geração nas tarefas de recuperação cruzada nos conjuntos de dados Flickr8k, Flickr30k e MSCOCO.</p>	<p>correspondência entre conteúdos visuais e textuais. Isso pode aprimorar a experiência de busca e recuperação de informações em repositórios institucionais, tornando mais eficiente e precisa a localização de conteúdos relevantes para os usuários.</p>
<p>2020</p>	<p>Developing Amaia: A Conversational Agent for Helping Portuguese Entrepreneurs—An Extensive Exploration of Question-Matching Approaches for Portuguese</p>	<p>O problema de pesquisa abordado no documento sobre o desenvolvimento de Amaia é a criação de um agente de conversação para empreendedores portugueses. O objetivo é desenvolver um sistema que possa corresponder automaticamente as perguntas dos usuários com Perguntas Frequentes (FAQs) em português, a fim de fornecer respostas precisas e úteis de forma eficiente. A pesquisa se concentra em explorar abordagens de correspondência de perguntas para melhorar a interação entre os usuários e o sistema de conversação, visando aprimorar a experiência do usuário e fornece suporte eficaz aos empreendedores em Portugal.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de</p>	<p>O objetivo do trabalho é desenvolver o agente de conversação Amaia para auxiliar empreendedores portugueses, fornecendo respostas automáticas em linguagem natural para perguntas relacionadas ao empreendedorismo e atividades econômicas em Portugal. Além disso, o objetivo é criar um agente capaz de manter conversas mais leves e informais sobre diversos tópicos, simulando uma interação mais próxima com os usuários. Amaia combina abordagens de Recuperação de Informação (IR) e modelos de Similaridade Textual Semântica (STS) para corresponder perguntas dos usuários com FAQs e responder a perguntas fora do domínio com base em um corpus de chitchat.</p>	<p>Os resultados da pesquisa incluem o desenvolvimento do agente de conversação Amaia, que combina abordagens de Recuperação de Informação (IR) e modelos de Similaridade Textual Semântica (STS) para corresponder perguntas dos usuários com Perguntas Frequentes (FAQs) em português. Foram realizadas extensas comparações de diferentes abordagens para corresponder solicitações de usuários com perguntas existentes, incluindo abordagens baseadas em IR, modelos STS supervisionados e não supervisionados. Amaia foi avaliada em um novo conjunto de dados, o corpus AIA-BDE, e demonstrou capacidade de lidar com interações fora do domínio com base em um corpus de chitchat. A pesquisa também abordou a redução da complexidade do modelo STS e</p>	<p>O estudo sobre o desenvolvimento de Amaia, um agente de conversação para auxiliar empreendedores portugueses, pode ser aplicado em repositórios institucionais para melhorar a interação com os usuários. A integração de abordagens de Recuperação de Informação (IR) e modelos de Similaridade Textual Semântica (STS) pode aprimorar a correspondência de perguntas dos usuários com informações nos repositórios, facilitando a recuperação de dados relevantes. Além disso, a capacidade de lidar com interações fora do domínio, baseada em um corpus de chitchat, pode tornar a interação com os repositórios mais amigável e eficaz, proporcionando respostas mais personalizadas e naturais aos usuários.</p>

		informação, necessidade percebida de informação, solicitação e query)		a melhoria da escalabilidade, mantendo o desempenho.	
		<b>Subcategoria:</b> Interação e IA			
2020	Discrimination of Verbal/Visuospatial Memory Retrieval Processes by Measuring Prefrontal Lobe Blood Volume With Functional Near-Infrared Spectrometry	<p>O problema de pesquisa abordado neste estudo é a discriminação dos processos de recuperação de memória verbal e visuoespacial, utilizando a medição do volume sanguíneo no lobo frontal por meio da espectrometria funcional de infravermelho próximo (fNIRS).</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo foi desenvolver um método para discriminar entre os processos de recuperação de memória verbal e visuoespacial por meio da observação do volume sanguíneo no lobo frontal, utilizando fNIRS. O estudo buscou utilizar a ativação do executivo central de memória de curto prazo para diferenciar entre a recuperação de memória verbal e visuoespacial.</p>	<p>O resultado da pesquisa indicou que o método proposto foi capaz de discriminar com sucesso entre os processos de recuperação de memória verbal e visuoespacial por meio da observação do volume sanguíneo no lobo frontal. A precisão, os valores preditivos positivos e negativos do kNN e SVM foram todos iguais a 1, o que sugere que o método foi eficaz na distinção entre os dois tipos de recuperação de memória.</p>	<p>O estudo utilizando fNIRS para discriminar processos de recuperação de memória verbal e visuoespacial oferece aplicações potenciais em pesquisas de neurociência cognitiva nos repositórios institucionais. Essa abordagem inovadora proporciona insights sobre os mecanismos cerebrais envolvidos na memória, destacando-se pela diferenciação eficaz entre distintos processos de recuperação de memória. Os métodos de discriminação, como kNN e SVM, sugerem utilidade em futuras investigações sobre atividade cerebral em contextos de memória e cognição.</p>
2020	Document-Level Text Classification Using Single-Layer Multisize Filters Convolutional Neural Network	<p>O problema de pesquisa abordado neste estudo é a classificação de documentos de texto em nível de documento usando uma Rede Neural Convolutiva de Filtros de Tamanho Único de Camada Única (SMFCNN). O foco está na língua urdu e na necessidade de desenvolver um modelo eficaz para classificar documentos de texto nessa língua. O estudo investiga a</p>	<p>O objetivo deste estudo é projetar e disponibilizar um grande conjunto de dados de documentos de texto em urdu, organizado em seis classes, para pesquisa futura. Além disso, o estudo visa comparar o desempenho de um modelo de Rede Neural Convolutiva de Filtros de Tamanho Único de Camada Única (SMFCNN) com dezesseis modelos de aprendizado de máquina em três</p>	<p>Os resultados da pesquisa indicam que o SMFCNN superou os classificadores de referência de aprendizado de máquina e alcançou pontuações de precisão de 95,4%, 91,8% e 93,3% em conjuntos de dados de tamanho médio, grande e pequeno, respectivamente. O estudo demonstrou que a remoção de stopwords e palavras raras teve um impacto significativo no desempenho do</p>	<p>O estudo sobre Document-Level Text Classification usando SMFCNN pode ser aplicado em repositórios institucionais para aprimorar a organização e recuperação de documentos não estruturados. Ao implementar o modelo SMFCNN treinado com conjuntos de dados em urdu, os repositórios podem melhorar a classificação automática de documentos, facilitando a busca e recuperação de informações</p>

		<p>influência do tamanho e do número de filtros nas redes neurais convolucionais, bem como a divisão adequada de conjuntos de dados em treinamento e teste para melhorar o desempenho do classificador.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>conjuntos de dados desequilibrados de tamanhos variados. O objetivo é analisar os efeitos de diferentes métodos de pré-processamento no desempenho do SMFCNN e avaliar a influência de parâmetros como stopwords, palavras raras e a divisão adequada de conjuntos de dados no desempenho do classificador.</p>	<p>SMFCNN, especialmente em conjuntos de dados de grande porte. Além disso, a divisão adequada dos conjuntos de dados de treinamento e teste resultou em um aumento significativo no desempenho do classificador. Os resultados destacam a eficácia do SMFCNN na classificação de documentos de texto em urdu e abrem caminho para pesquisas futuras nessa área.</p>	<p>relevantes. A remoção de stopwords e palavras raras, juntamente com a divisão adequada dos conjuntos de dados, pode otimizar o desempenho do classificador, resultando em uma melhor experiência de usuário ao acessar e explorar os documentos armazenados nos repositórios institucionais.</p>
2020	Efficient Regular Language Search for Secure Cloud Storage	<p>O problema de pesquisa abordado no documento é a necessidade de projetar um novo esquema de busca de linguagem regular seguro para armazenamento em nuvem que supere as limitações de esquemas existentes, como a vulnerabilidade a ataques de adivinhação de palavras-chave offline e a dependência de geração de trapdoor pelo centro de geração de chaves (KGC). Além disso, o documento destaca a falta de flexibilidade nos padrões de busca de criptografia pesquisável disponíveis e a necessidade de um sistema que permita aos usuários definir um autômato finito determinístico (DFA) para recuperar dados de forma eficiente e segura. Portanto, o problema de pesquisa central é a concepção de um sistema de armazenamento de dados</p>	<p>O objetivo do estudo apresentado no documento é propor um esquema eficiente de criptografia pesquisável de linguagem regular para armazenamento seguro em nuvem. Esse esquema visa garantir a proteção da confidencialidade dos dados e a preservação da privacidade durante a recuperação de dados, ao mesmo tempo em que é resistente a ataques de adivinhação de palavras-chave offline. Além disso, o objetivo é desenvolver um sistema que suporte a busca de linguagem regular e seja capaz de estender-se para múltiplos usuários, permitindo que cada usuário gere um token DFA usando sua própria chave privada sem a necessidade de interagir com o centro de geração de chaves. O esquema proposto busca oferecer uma</p>	<p>O resultado da pesquisa apresentada no documento é a introdução de um esquema de criptografia pesquisável de universo grande para proteger a segurança do sistema de armazenamento em nuvem. Esse esquema realiza a criptografia de linguagem regular e a função de busca DFA. O provedor de serviços em nuvem pode testar se a linguagem regular criptografada no texto cifrado é aceitável pelo DFA incorporado no token de busca enviado. Durante o procedimento de teste, nenhum texto simples da linguagem regular ou do DFA será vazado para o servidor em nuvem. Além disso, foi proposta uma construção concreta com algoritmos de criptografia e geração de token leves. O esquema proposto é preservador de privacidade e</p>	<p>O estudo sobre o esquema de criptografia pesquisável de linguagem regular para armazenamento em nuvem pode ser aplicado em repositórios institucionais para garantir a segurança e privacidade dos dados armazenados. Ao implementar esse esquema, os repositórios institucionais podem permitir que os usuários realizem buscas eficientes e seguras em seus dados, protegendo a confidencialidade das informações e preservando a privacidade durante a recuperação de dados. Isso é especialmente relevante em ambientes onde a confiança no servidor de armazenamento em nuvem não é absoluta. A capacidade de suportar múltiplos usuários e permitir que cada usuário gere seu próprio token DFA sem depender de</p>

		<p>seguro que suporte a busca de linguagem regular, garantindo privacidade e segurança no modelo padrão com base na suposição de dificuldade de decisão do problema de Diffie-Hellman bilinear.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>solução eficiente, segura e flexível para a busca de dados em armazenamento em nuvem, superando as limitações dos esquemas existentes.</p>	<p>indistinguível contra ataques de adivinhação de palavras-chave offline, conforme demonstrado no modelo padrão. A comparação e os resultados experimentais confirmam a baixa sobrecarga de transmissão e computação do esquema.</p>	<p>interações com um centro de geração de chaves centralizado torna o esquema adequado para repositórios institucionais que lidam com dados sensíveis e confidenciais.</p>
2020	Empirical evaluation and study of text stemming algorithms	<p>O problema de pesquisa abordado no estudo é a avaliação empírica e estudo dos algoritmos de stemming de texto. Os autores investigam as dificuldades e desafios enfrentados ao avaliar a eficácia dos algoritmos de stemming em línguas que utilizam o script árabe, com foco especial no Urdu. Eles exploram as limitações dos métodos de avaliação atuais e propõem critérios para uma medição mais precisa do desempenho dos stemmers.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo do estudo é revisar e analisar os métodos de avaliação de stemming de texto, a fim de desenvolver critérios para uma medição mais precisa do desempenho dos stemmers. Os autores buscam discutir os diferentes aspectos da medição de desempenho dos stemmers, incluindo suas principais características, méritos e limitações, utilizando o Urdu como exemplo de uma língua com recursos limitados. Através de experimentos, eles procuram destacar as deficiências dos métodos de avaliação existentes e fornecer insights para uma avaliação mais eficaz do desempenho dos stemmers.</p>	<p>Os resultados da pesquisa indicam que os métodos de avaliação atuais para stemming de texto podem medir apenas uma média de confluência de palavras, independentemente da correção do radical. Além disso, alguns dos métodos de avaliação favorecem apenas um tipo específico de língua. Os autores concluem que nenhum dos métodos de avaliação existentes consegue medir perfeitamente o desempenho dos stemmers para todos os tipos de línguas. Eles sugerem que esse estudo pode auxiliar os pesquisadores na avaliação de seus stemmers utilizando métodos mais apropriados e eficazes.</p>	<p>O estudo sobre a avaliação empírica de algoritmos de stemming de texto pode ser aplicado em repositórios institucionais para aprimorar a eficácia da indexação e recuperação de informações. Ao utilizar métodos de avaliação mais precisos e abrangentes, os repositórios podem melhorar a qualidade da análise de texto, facilitando a busca e recuperação de informações relevantes para os usuários. Isso pode resultar em uma melhor organização e acessibilidade dos conteúdos armazenados nos repositórios institucionais, beneficiando pesquisadores, estudantes e demais usuários em suas atividades acadêmicas e de pesquisa.</p>
2020	Enriching Domain Concepts with Qualitative	<p>O problema de pesquisa abordado no artigo é a extração de valores de atributos</p>	<p>O objetivo do estudo apresentado no artigo é desenvolver um framework</p>	<p>Os resultados da pesquisa indicam que o sistema proposto, que integra o Modelo de Espaço</p>	<p>O estudo propõe um framework analítico para a extração e classificação de valores de</p>

	<p>Attributes: A Text Mining based Approach</p>	<p>qualitativos de texto não estruturado e a atribuição de atributos a esses valores por meio da comparação do significado ou contexto derivado do Google de atributos, bem como de valores de qualidade (adjetivos).</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>analítico para adquirir valores de atributos qualitativos a partir de texto não estruturado e atribuir atributos a esses valores por meio da comparação do significado ou contexto derivado do Google de atributos, bem como de valores de qualidade (adjetivos).</p>	<p>Vetorial (VSM) com um classificador probabilístico Multinomial Naive Bayes (MNB), teve um desempenho superior em comparação com frameworks baseados em padrões existentes no estado da arte. A avaliação de desempenho foi realizada em dois conjuntos de dados: o Conjunto de Dados de Desenvolvimento HeiPLAS e um conjunto de dados de texto no Domínio de Plantas Medicinais. O sistema demonstrou melhor desempenho com a abordagem probabilística em comparação com frameworks baseados em padrões existentes.</p>	<p>atributos qualitativos a partir de texto não estruturado. Essa abordagem pode ser aplicada em repositórios institucionais para enriquecer a descrição de conceitos de domínio com atributos qualitativos, melhorando a recuperação e organização da informação. Ao integrar o Modelo de Espaço Vetorial com um classificador probabilístico Multinomial Naive Bayes, o sistema pode automatizar o processo de atribuição de atributos a valores extraídos do texto, facilitando a indexação e busca de informações nos repositórios institucionais.</p>
2020	<p>Extracting Arabic Composite Names Using Genitive Principles of Arabic Grammar</p>	<p>O problema de pesquisa abordado neste trabalho é a extração de nomes compostos em árabe a partir de documentos não estruturados, utilizando os princípios da gramática genitiva árabe. A pesquisa visa desenvolver uma abordagem eficaz para reconhecer entidades compostas em árabe, levando em consideração a complexidade da morfologia da língua árabe e a falta de ferramentas avançadas de Processamento de Linguagem Natural (PLN) em árabe. O foco está na identificação e extração precisa de nomes compostos em árabe, considerando a classificação de substantivos</p>	<p>O objetivo deste trabalho é desenvolver uma abordagem baseada em regras para extrair nomes compostos em árabe a partir de textos não estruturados, utilizando os princípios da gramática genitiva árabe. O foco principal é aprimorar o reconhecimento de entidades nomeadas em árabe, especialmente no contexto de nomes compostos, que são comuns na língua árabe. O objetivo é alcançar alta precisão e recall na extração de nomes compostos, demonstrando a eficácia da abordagem proposta. Além disso, o estudo visa contribuir para o avanço da tecnologia de Processamento de</p>	<p>Os resultados promissores indicam a eficácia da abordagem na extração precisa de nomes compostos em árabe, preparando dados semanticamente marcados para análise inteligente de documentos. Além disso, o estudo ressaltou a necessidade de lidar com desafios sintáticos e erros gramaticais árabes. Por fim, apontou para futuras pesquisas explorando técnicas avançadas de Aprendizado de Máquina para o Reconhecimento de Entidades Nomeadas (NER) em textos árabes genéricos, visando superar desafios na recuperação</p>	<p>O estudo sobre a extração de nomes compostos em árabe usando princípios da gramática genitiva pode ser aplicado em repositórios institucionais para melhorar a extração de informações de documentos em árabe não estruturados. A abordagem baseada em regras e padrões linguísticos pode ser utilizada para identificar e extrair com precisão nomes compostos em árabe, contribuindo para a organização e recuperação eficiente de dados em repositórios institucionais que contenham documentos nesse idioma. Essa metodologia inovadora pode aprimorar a indexação e a busca de</p>

		<p>definidos e indefinidos, bem como a aplicação de regras de reconhecimento de padrões genitivos e artigos definidos árabes para a extração de nomes compostos.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>Linguagem Natural (PLN) em árabe, fornecendo uma metodologia inovadora para lidar com a complexidade da língua árabe e melhorar a extração de informações de textos não estruturados nesse idioma.</p>	<p>de informações em textos não estruturados.</p>	<p>informações em textos em árabe, facilitando a análise e o acesso a conteúdos relevantes para pesquisadores, estudantes e profissionais que utilizam repositórios institucionais como fonte de informação.</p>
<p>2020</p>	<p>Extractive Text Summarization Models for Urdu Language</p>	<p>O problema de pesquisa abordado no documento é a falta de recursos e conjuntos de dados publicamente disponíveis para a geração automática de resumos extrativos em urdu. Isso dificulta o trabalho dos pesquisadores em tarefas mais complexas de processamento de texto nesse idioma de recursos limitados. A pesquisa visa resolver esse problema ao desenvolver um framework público para gerar resumos textuais de textos em urdu, aplicando duas abordagens famosas de ponderação de frases.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O principal objetivo da pesquisa é desenvolver um framework público para gerar um resumo textual do idioma urdu, aplicando duas abordagens famosas de ponderação de frases. O primeiro objetivo é calcular os pesos das palavras com base no contexto do texto fornecido, enquanto o segundo objetivo é calcular os pesos a partir de um conjunto de dados independente. Além disso, o estudo visa gerar um conjunto de dados em urdu publicamente disponível para sumarização extrativa e avaliar os resultados com base em resumos gerados por especialistas humanos usando as métricas ROUGE-1 e ROUGE-2.</p>	<p>Os resultados da pesquisa indicam que as abordagens baseadas em pesos locais superam as abordagens baseadas em pesos globais na geração de resumos em urdu. Enquanto as estratégias de ponderação baseadas em TF-IDF e embedding de palavras são consideradas técnicas padrão para calcular pesos em diversas aplicações, como classificação de texto e recuperação de informações, o método de ponderação de frases e de frequência de termos ponderados, utilizados nas abordagens locais, calculam os pesos das frases de maneira diferente. Os experimentos revelaram que todas as abordagens tiveram desempenho limitado em relação aos resumos abstrativos devido à presença de sinônimos e novo vocabulário, sugerindo a necessidade de explorar</p>	<p>O estudo sobre modelos de sumarização de texto extrativo para o idioma urdu pode ser aplicado em repositórios institucionais para facilitar a geração automática de resumos de documentos em urdu. Ao implementar as abordagens de ponderação de frases locais e globais, os repositórios podem oferecer aos usuários resumos precisos e concisos dos documentos em urdu disponíveis. Isso pode melhorar a acessibilidade e a compreensão do conteúdo para os usuários que buscam informações específicas nos repositórios institucionais.</p>

				sinônimos para melhorar a precisão e expandir a abordagem para resumos abstrativos no futuro.	
2020	Ambiguidade e o contexto na representação de informações em domínios de especialidade	<p>O contexto da pesquisa abordada no documento está relacionado à Arquitetura e Organização do Conhecimento, especialmente no âmbito da Ciência da Informação. A autora discute a importância do contexto na minimização de problemas de ambiguidade na recuperação da informação, destacando a relevância dos Sistemas de Organização do Conhecimento como linguagem de comunicação composta por conceitos que estruturam o conhecimento em domínios específicos. Além disso, a pesquisa explora como os sistemas inteligentes podem utilizar o contexto como um mecanismo mediador para aprimorar a comunicação entre humanos e computadores.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>O objetivo da pesquisa é analisar a ambiguidade na representação de informações em domínios de especialidade, destacando a importância do contexto como um componente essencial para minimizar esses problemas. A autora busca discutir como o contexto pode ser considerado na construção de abstrações e modelos da realidade, especialmente no desenvolvimento de Sistemas de Organização do Conhecimento para uso computacional. Além disso, a pesquisa visa explorar como os sistemas inteligentes podem aplicar o contexto como um mecanismo mediador para melhorar a comunicação entre humanos e computadores.</p>	<p>Os resultados da pesquisa indicam que a consideração do contexto na representação de informações em domínios de especialidade é fundamental para minimizar problemas de ambiguidade. A autora destaca a importância dos Sistemas de Organização do Conhecimento como uma linguagem de comunicação composta por conceitos que estruturam o conhecimento em áreas específicas. Além disso, a pesquisa demonstra como os sistemas inteligentes podem utilizar o contexto como um mecanismo mediador para aprimorar a comunicação entre humanos e computadores. Em suma, a pesquisa ressalta a relevância do contexto na construção de modelos de realidade e na melhoria da interação humano-computador em ambientes de informação especializada.</p>	<p>O estudo sobre a importância do contexto na representação de informações em domínios de especialidade pode ser aplicado em repositórios institucionais para melhorar a recuperação da informação. Ao considerar o contexto na construção de metadados e na organização do conhecimento dentro desses repositórios, é possível minimizar problemas de ambiguidade e facilitar a busca por informações relevantes. Além disso, a utilização do contexto como um mecanismo mediador pode aprimorar a comunicação entre usuários e o sistema de informação, tornando a experiência de busca mais eficiente e satisfatória.</p>

2020	<p style="text-align: center;">Head Concepts Selection for Verbose Medical Queries Expansion</p>	<p>O problema de pesquisa abordado neste trabalho é a necessidade de aprimorar os sistemas de recuperação de informações médicas (MIR) para lidar com consultas médicas longas e detalhadas, que contêm vários conceitos específicos do domínio médico. Essas consultas exigem a utilização de recursos linguísticos e semânticos para sua interpretação e expansão, a fim de melhor corresponder a documentos médicos relevantes. O desafio reside na incompletude e na limitada cobertura dos recursos semânticos médicos existentes, o que motiva a busca por abordagens que combinem múltiplos recursos semânticos médicos e técnicas de expansão de consultas para melhorar a correspondência entre consultas e documentos médicos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>O objetivo deste trabalho é desenvolver um framework de busca de informações médicas que destaque e expanda os conceitos médicos principais em consultas médicas detalhadas (verbose), construa índices invertidos semanticamente aprimorados e contribua para uma técnica de ponderação heurística no processo de correspondência entre consulta e documento. O objetivo final é demonstrar a eficácia dessa abordagem combinando vários recursos semânticos extrínsecos na melhoria da precisão da busca de informações médicas, conforme evidenciado pelos experimentos realizados no conjunto de dados CLEF e-Health 2014.</p>	<p>Os resultados da pesquisa indicaram que a abordagem proposta, que combina vários recursos semânticos extrínsecos na expansão de consultas médicas verbose, foi mais eficaz do que abordagens relacionadas em termos de medida de precisão. Através de experimentos realizados no conjunto de dados CLEF e-Health 2014, foi demonstrado que a utilização de múltiplos recursos semânticos médicos resultou em uma melhoria na correspondência entre consultas e documentos médicos, contribuindo para uma melhor interpretação da intenção da consulta e, conseqüentemente, para a recuperação de resultados mais relevantes. Os resultados também destacaram a importância de considerar as relações semânticas entre os conceitos médicos identificados nas consultas dos usuários para aprimorar a compreensão da intenção da consulta e a qualidade dos resultados de recuperação.</p>	<p>O estudo sobre a seleção de conceitos principais em consultas médicas verbose e sua expansão pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações médicas. Ao combinar múltiplos recursos semânticos médicos na interpretação de consultas detalhadas, é possível melhorar a correspondência entre as consultas dos usuários e os documentos médicos armazenados nos repositórios. Isso resulta em uma recuperação mais precisa e relevante de informações médicas, contribuindo para uma melhor organização e acesso aos recursos médicos disponíveis nos repositórios institucionais.</p>
------	--	---	---	--	--

2020	HerCulB: content-based information extraction and retrieval for cultural heritage of the Balkans	<p>O problema de pesquisa abordado no artigo é a necessidade de desenvolver uma metodologia para a anotação automática de uma coleção de patrimônio cultural intangível dos Bálcãs. Os autores destacam os desafios enfrentados pelos pesquisadores que trabalham com o patrimônio cultural, como a falta de padronização da linguagem, a integração de vários tipos de mídia para extração de informações, a fraqueza no suporte de ferramentas e recursos adaptados a esse domínio, e a existência de múltiplos padrões em uso. Assim, o objetivo principal é fornecer um sistema de anotação semântica automática que seja compatível com a ontologia CIDOC CRM e outras vocabulários, ontologias e tesouros do domínio cultural, visando facilitar a recuperação de informações e preservação do patrimônio cultural dos Bálcãs.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo do trabalho apresentado no artigo é desenvolver um sistema para a anotação semântica automática de uma coleção de patrimônio cultural intangível dos Bálcãs. Esse sistema, chamado HerCulB, visa extrair metadados automaticamente a partir do domínio cultural, realizar a anotação semi-automática de documentos e permitir a recuperação de informações de um banco de dados multimídia. A metodologia proposta é aplicada principalmente à língua sérvia, mas não se limita a ela, podendo ser adaptada e aplicada a outras línguas. O objetivo final é facilitar o acesso e a preservação do patrimônio cultural dos Bálcãs, encapsulando conhecimento específico do domínio e da língua nos sistemas de anotação automática.</p>	<p>O resultado da pesquisa apresentada no artigo foi a criação de uma metodologia para a anotação automática de uma coleção de patrimônio cultural intangível dos Bálcãs. A aplicação dessa metodologia ao patrimônio cultural dos Bálcãs resultou em resultados positivos, com uma medida F de 0,87 para a anotação de entidades nomeadas e 0,90 para a anotação de tópicos. Esses resultados indicam que a metodologia desenvolvida teve um desempenho satisfatório na extração automática de metadados e na anotação de documentos no domínio do patrimônio cultural. Além disso, a metodologia permitiu encapsular conhecimento específico do domínio e da língua em coleções de transdutores de estado finito, abrindo caminho para futuras melhorias e desenvolvimentos na preservação e acesso ao patrimônio cultural dos Bálcãs.</p>	<p>O estudo sobre o HerCulB, que foca na extração e recuperação de informações para o patrimônio cultural dos Bálcãs, pode ser aplicado em repositórios institucionais de diversas maneiras. A metodologia desenvolvida para a anotação automática de coleções de patrimônio cultural intangível pode ser utilizada para enriquecer metadados, facilitar a recuperação de informações e melhorar a organização de conteúdos culturais em repositórios. Isso possibilita uma busca mais eficiente e precisa dentro desses repositórios, tornando o acesso ao patrimônio cultural mais fácil e enriquecedor para os usuários. Além disso, a aplicação da metodologia em repositórios institucionais pode contribuir para a preservação e divulgação do patrimônio cultural de forma mais abrangente e significativa.</p>
2020	Highly Language-Independent Word Lemmatization Using a	<p>O problema de pesquisa abordado no documento é a lematização de palavras de forma altamente independente</p>	<p>O objetivo do estudo é desenvolver um lematizador de palavras altamente independente de idioma,</p>	<p>O resultado da pesquisa foi a criação de um lematizador de palavras altamente independente de idioma,</p>	<p>O estudo sobre lematização de palavras altamente independente de idioma, utilizando um classificador de</p>

	<p>Machine-Learning Classifier</p>	<p>de idioma, utilizando um classificador de aprendizado de máquina. A pesquisa visa desenvolver um lematizador que possa ser aplicado a uma variedade de idiomas, independentemente de suas características linguísticas específicas. O objetivo é criar um método eficaz e versátil que possa determinar a forma base de uma palavra (lema) em diferentes idiomas, facilitando assim tarefas de processamento de linguagem natural em um contexto multilíngue.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>baseado em um classificador de aprendizado de máquina do tipo Random Forest. O foco é criar um sistema que possa identificar a forma base (lema) de uma palavra em diferentes idiomas, sem depender de regras específicas de cada idioma. O objetivo é fornecer uma solução eficaz e simples para a lematização de palavras em diversos idiomas, facilitando assim tarefas de processamento de linguagem natural em um contexto multilíngue.</p>	<p>baseado em um classificador de Random Forest. O lematizador desenvolvido mostrou bom desempenho em uma variedade de idiomas de diferentes grupos linguísticos e foi comparado com o lematizador UDPipe em 22 dos 25 idiomas testados. Os resultados indicaram que o lematizador baseado em árvores de decisão e Random Forest alcançou resultados significativamente melhores do que o UDPipe quando fornecido com uma quantidade adequada de dados de treinamento. A abordagem proposta mostrou-se eficaz e facilmente extensível para outros idiomas, representando uma contribuição significativa para a lematização de palavras em um contexto multilíngue.</p>	<p>Random Forest, pode ser aplicado em repositórios institucionais para melhorar a precisão e eficiência da recuperação de informações. Ao implementar esse lematizador nos repositórios, é possível padronizar as formas das palavras em diferentes idiomas, facilitando a busca e análise de conteúdo multilíngue. Isso pode aprimorar a indexação, a categorização e a recuperação de documentos, tornando o acesso à informação mais eficiente e preciso para os usuários dos repositórios institucionais.</p>
<p>2020</p>	<p>Implementation and Evaluation of a Retrieval-based Chinese Humor Chatbot</p>	<p>O problema de pesquisa abordado neste estudo é a necessidade de desenvolver um sistema de diálogo de humor em chinês que seja capaz de fornecer piadas relevantes e engraçadas para os usuários, visando melhorar a interação social e promover um ambiente descontraído em conversas informais. O foco está em automatizar a riqueza e a qualidade do corpus de piadas, bem como em fornecer um serviço de recomendação eficaz</p>	<p>O objetivo deste estudo é construir um corpus de humor, desenvolver tecnologias relacionadas e implementar um sistema de chatbot de humor chinês baseado em recuperação, conhecido como "icebreaker chatbot", que permite aos usuários encontrar piadas relevantes para uso em situações informais, com o intuito de relaxar o ambiente. Além disso, o estudo visa avaliar a eficácia do chatbot em quebrar o gelo por meio das piadas</p>	<p>Os resultados da pesquisa indicam que, após a implementação de melhorias no chatbot de humor chinês, houve uma redução significativa na taxa de falha na busca de piadas, passando de 25,4% para 16,7%. Além disso, a eficácia do chatbot em quebrar o gelo durante as interações aumentou de 27,9% para 39,9%. Esses resultados demonstram a importância de compilar um corpus de piadas de alta qualidade e de desenvolver um</p>	<p>O estudo sobre o chatbot de humor chinês baseado em recuperação pode ser aplicado em repositórios institucionais para melhorar a interação com os usuários, proporcionando um ambiente mais descontraído e facilitando a busca por informações. A implementação de um sistema de chatbot de humor pode ajudar a quebrar o gelo em interações online, tornando a experiência do usuário mais agradável e engajadora. Além disso, a</p>

		<p>para melhorar a experiência do usuário ao interagir com o chatbot de humor chinês.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>recuperadas. A pesquisa busca melhorar a qualidade do serviço oferecido pelo chatbot, reduzindo a taxa de falha na busca de piadas e aumentando a taxa de sucesso na quebra do gelo durante as interações.</p>	<p>sistema de diálogo de humor eficaz, destacando a necessidade contínua de esforços de pesquisa e desenvolvimento para aprimorar tais serviços.</p>	<p>utilização de tecnologias de processamento de linguagem natural e recuperação de informações pode aprimorar a eficiência na busca e recuperação de conteúdo nos repositórios, contribuindo para uma experiência mais interativa e personalizada para os usuários.</p>
<p>2020</p>	<p>Information Retrieval Based on Knowledge-Enhanced Word Embedding Through Dialog: A Case Study</p>	<p>O problema de pesquisa abordado no estudo é fornecer uma rota sistemática de recuperação de informações de um banco de dados baseado em conhecimento (ou conhecimento de domínio) por meio de um sistema de diálogo de interação em linguagem natural. O foco está em melhorar a compreensão semântica e a eficácia da recuperação de informações por meio da incorporação de conhecimento de domínio nos embeddings de palavras utilizados no processo de diálogo.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo do estudo é integrar abordagens baseadas em conhecimento e orientadas por dados em um sistema de diálogo, onde os embeddings de palavras são aprimorados com o conhecimento de domínio para se adequarem especificamente à linguagem natural utilizada na aplicação. Especificamente, o estudo visa criar uma rota sistemática para a recuperação de informações em um ambiente universitário, onde o conhecimento de domínio é central e é incorporado nos embeddings de palavras para melhorar a compreensão semântica e a interação em linguagem natural no sistema de diálogo.</p>	<p>Os resultados da pesquisa demonstram que os embeddings de palavras aprimorados com conhecimento de domínio são mais precisos e expressivos na transmissão de informações semânticas, especialmente quando treinados por modelos específicos, como o modelo LIC-skip-gram. Esses embeddings melhoram a compreensão semântica no sistema de diálogo, facilitando o processo de interação e tornando a recuperação de informações mais eficiente. Além disso, a integração de conhecimento de domínio em diferentes níveis de granularidade nos embeddings de palavras contribui para uma melhor compreensão semântica e torna o processo de diálogo mais fluente.</p>	<p>O estudo demonstra que a incorporação de conhecimento de domínio nos embeddings de palavras pode melhorar significativamente a compreensão semântica e a eficácia da recuperação de informações em sistemas de diálogo. Essa abordagem pode ser aplicada em repositórios institucionais para aprimorar a busca e recuperação de informações, tornando a interação com os usuários mais eficiente e precisa. Ao integrar o conhecimento de domínio nos embeddings de palavras utilizados nos repositórios, é possível melhorar a compreensão semântica das consultas dos usuários e fornecer resultados mais relevantes e precisos.</p>

<p>2020</p>	<p>Math-word embedding in math search and semantic extraction</p>	<p>O problema de pesquisa abordado neste estudo é a melhoria da recuperação de informações e extração de conhecimento semântico em documentos matemáticos, utilizando técnicas de incorporação de palavras (word embedding). A pesquisa visa superar os desafios associados à natureza imprecisa da linguagem matemática e à necessidade de compreender as relações entre conceitos matemáticos e seus descritores em um contexto mais amplo.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo deste estudo é explorar o uso e a eficácia das técnicas de incorporação de palavras (word embedding) em documentos matemáticos para tarefas de recuperação de informações matemáticas e extração de conhecimento semântico. Os pesquisadores testaram vários cenários, como similaridade de termos matemáticos, analogias matemáticas, modelagem de conceitos numéricos, busca matemática usando expansões de consulta e extração semântica de expressões matemáticas. O estudo busca avaliar a promessa das técnicas de incorporação de palavras para tarefas de similaridade, analogia e busca matemática, ao mesmo tempo em que identifica a necessidade de abordagens mais robustas nesse contexto.</p>	<p>Os resultados da pesquisa indicaram que as técnicas de incorporação de palavras (word embedding) têm um grande potencial para tarefas de similaridade, analogia e busca matemática em documentos matemáticos. No entanto, os pesquisadores observaram a necessidade de abordagens mais robustas de incorporação de palavras para lidar com os desafios específicos da linguagem matemática e da recuperação de informações nesse domínio. Embora tenham obtido resultados promissores em tarefas como similaridade de termos, analogias matemáticas e modelagem de conceitos, os pesquisadores identificaram problemas fundamentais que impedem os algoritmos de aprendizado de máquina de compreender efetivamente a matemática e extrair conhecimento semântico.</p>	<p>O estudo sobre a incorporação de palavras em documentos matemáticos pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações matemáticas e a extração de conhecimento semântico. Ao utilizar técnicas avançadas de incorporação de palavras, os repositórios podem aprimorar a busca por termos matemáticos, identificar analogias entre conceitos, modelar conceitos numéricos e expandir consultas de forma mais eficaz. Isso pode facilitar a navegação e a compreensão de documentos matemáticos, tornando o acesso a informações matemáticas mais eficiente e preciso para pesquisadores e estudantes.</p>
<p>2020</p>	<p>MSC+: Language Pattern Learning for Word Sense Induction and Disambiguation</p>	<p>O problema de pesquisa abordado neste trabalho é a dificuldade enfrentada pelos inventários e sistemas de semântica lexical em lidar com problemas de deriva semântica em conteúdos gerados pelo usuário e comunicação online. Essa deriva semântica ocorre devido à criação contínua de</p>	<p>O objetivo deste estudo é propor uma abordagem para induzir e desambiguar sentidos de palavras em textos curtos, como tweets, por meio do uso de padrões lexico-semânticos difusos, denominados padrões MSC+. A intenção é melhorar a capacidade de identificar corretamente o significado das</p>	<p>Os resultados da pesquisa demonstraram que a utilização de padrões MSC+ para induzir semântica em palavras-alvo em textos curtos, como tweets, levou a melhorias significativas na capacidade de desambiguação de sentidos de palavras. Os experimentos realizados mostraram que os</p>	<p>O estudo sobre Language Pattern Learning for Word Sense Induction and Disambiguation propõe uma abordagem inovadora para induzir e desambiguar sentidos de palavras em textos curtos, como tweets, utilizando padrões lexico-semânticos difusos chamados padrões MSC+. Essa</p>

		<p>novos significados para palavras existentes, resultando em desafios para identificar corretamente o significado das palavras em contextos específicos e descobrir novos sentidos das palavras. A proposta do estudo é desenvolver uma abordagem para induzir e desambiguar sentidos de palavras-alvo em coleções de textos curtos, como tweets, utilizando padrões lexico-semânticos difusos chamados de padrões MSC+.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>palavras em contextos específicos e lidar com os desafios de deriva semântica em conteúdos gerados pelo usuário e comunicação online. Os experimentos realizados visam demonstrar a eficácia desses padrões na indução de semântica em palavras-alvo e na melhoria dos métodos de desambiguação de sentidos de palavras.</p>	<p>padrões MSC+ foram capazes de identificar instâncias em tweets, mesmo quando palavras diferentes eram usadas para transmitir o sentido do componente morfo-semântico correspondente. Essa abordagem resultou em melhorias no estado da arte em relação à indução e desambiguação de sentidos de palavras, especialmente em contextos de comunicação online e conteúdos gerados pelo usuário.</p>	<p>metodologia pode ser aplicada em repositórios institucionais para melhorar a capacidade de identificar corretamente o significado das palavras em contextos específicos, especialmente em conteúdos gerados pelo usuário. Ao utilizar esses padrões, é possível aprimorar a qualidade das anotações semânticas e a precisão da desambiguação de sentidos de palavras, contribuindo para uma melhor organização e recuperação de informações em repositórios institucionais.</p>
<p>2020</p>	<p>Neural relational inference for disaster multimedia retrieval</p>	<p>O problema de pesquisa abordado neste estudo é como identificar adequadamente eventos de interesse, como desastres naturais, com base em dados multimodais disponíveis. Especificamente, o foco está em desenvolver soluções eficazes que possam inferir eventos com base em visualizações complementares fornecidas por diferentes modalidades, como</p>	<p>O objetivo deste estudo é propor um modelo de inferência neural relacional para a recuperação de multimídia relacionada a desastres naturais. O modelo foi desenvolvido para lidar com dados não estruturados, como texto sem uma ordem específica de palavras, comumente encontrado em dados de redes sociais. Além disso, o método pode ser combinado com uma CNN para lidar com dados</p>	<p>Os resultados da pesquisa indicam que o modelo de inferência neural relacional proposto obteve um desempenho competitivo em cenários unimodais, tanto para dados visuais quanto textuais, superando significativamente a linha de base no cenário multimodal. Além disso, o estudo demonstrou a eficácia do modelo em diferentes aplicações, incluindo a</p>	<p>O estudo sobre inferência neural relacional para recuperação de multimídia relacionada a desastres naturais pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de conteúdo multimodal relevante. Ao utilizar um modelo que combina dados visuais e textuais de forma eficaz, é possível melhorar a precisão na identificação de eventos de</p>

		<p>propriedades visuais e descrições textuais. Este problema ainda é uma questão em aberto em diversas aplicações, incluindo aquelas relacionadas à gestão adequada de coleções multimídia em redes sociais.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>multimodais, nos quais dados visuais e textuais estão disponíveis. O estudo visa demonstrar a eficácia do modelo proposto em diferentes cenários, incluindo avaliações de modalidades visual, textual e multimodal, com resultados competitivos em cada uma dessas modalidades.</p>	<p>recuperação de multimídia relacionada a desastres naturais, conforme introduzido pela Tarefa de Satélite Multimídia. Os experimentos realizados mostraram que o modelo proposto é capaz de lidar com dados não estruturados e pode inferir estruturas relacionais a partir de conjuntos de objetos, sendo uma abordagem promissora para a análise multimodal de objetos multimídia</p>	<p>interesse, como desastres naturais, a partir de coleções de multimídia. Isso pode facilitar a organização e recuperação de informações críticas em repositórios institucionais, tornando a busca por conteúdo multimodal mais eficiente e precisa.</p>
2020	<p>One deep music representation to rule them all? A comparative analysis of different representation learning strategies</p>	<p>O problema de pesquisa abordado no documento é a geração de representações profundas eficazes para dados musicais no campo da Recuperação de Informação Musical (Music Information Retrieval). O estudo investiga diferentes estratégias de aprendizado de representação para várias tarefas de aprendizado em música, com foco na transferência de aprendizado profundo e na avaliação da eficácia dessas representações profundas para novas tarefas de aprendizado e conjuntos de dados-alvo na área musical.</p> <p><b>Categoria:</b> Recursos de informação</p>	<p>O objetivo do estudo é investigar os fatores mais importantes para gerar representações profundas para dados e tarefas de aprendizado na área da música, por meio de uma análise comparativa de diferentes estratégias de aprendizado de representação. O estudo visa fornecer insights sobre a eficácia de várias estratégias de aprendizado de representação profunda no contexto de aprendizado multitarefa, oferecer uma compreensão aprofundada de como avaliar as propriedades desejadas de um procedimento de aprendizado de representação profunda e propor e disponibilizar várias redes de representação musical pré-</p>	<p>Os resultados da pesquisa fornecem insights sobre a eficácia de diferentes estratégias de aprendizado de representação profunda no contexto de aprendizado multitarefa na área da música. Além disso, o estudo oferece uma compreensão aprofundada de como avaliar as propriedades desejadas de um procedimento de aprendizado de representação profunda e propõe várias redes de representação musical pré-treinadas, baseadas em diferentes estratégias de aprendizado para múltiplas fontes de aprendizado semântico. A pesquisa contribui para a compreensão de como</p>	<p>O estudo sobre aprendizado de representação profunda para dados musicais na Recuperação de Informação Musical pode ser aplicado em repositórios institucionais para melhorar a organização e recuperação de informações musicais. Ao utilizar as estratégias de aprendizado de representação propostas no estudo, os repositórios institucionais podem aprimorar a categorização, indexação e busca de conteúdo musical, tornando mais eficiente a recuperação de informações relevantes para os usuários. Essas técnicas podem ajudar a criar representações profundas mais informativas e transferíveis para uma ampla gama de tarefas</p>

		<b>Subcategoria:</b> Representação do documento	treinadas, baseadas em diferentes estratégias de aprendizado para múltiplas fontes de aprendizado semântico.	gerar representações profundas informativas para dados musicais e tarefas de aprendizado na área da Recuperação de Informação Musical.	e conjuntos de dados musicais, melhorando assim a experiência de busca e descoberta de música nos repositórios institucionais.
2020	Quantitative similarity assessment of construction projects using WBS-based metrics	<p>O problema de pesquisa abordado neste estudo é a avaliação quantitativa da similaridade entre projetos de construção utilizando métricas baseadas na Estrutura Analítica do Trabalho (WBS). O objetivo é desenvolver um método que possa identificar e comparar projetos semelhantes com base em suas estruturas de WBS, visando melhorar a gestão do conhecimento e a tomada de decisões em projetos de construção.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>O objetivo deste estudo é propor um método inovador para medir a similaridade entre projetos de construção com base na comparação semântica de suas Estruturas Analíticas do Trabalho (WBS). O método desenvolvido utiliza técnicas de processamento de linguagem natural para calcular a similaridade entre as WBS dos projetos, empregando três métricas distintas: similaridade de nó, similaridade estrutural e similaridade total. O principal objetivo é avaliar a eficácia desse método na identificação de projetos similares e na recuperação de informações relevantes de projetos passados, contribuindo assim para a melhoria da gestão do conhecimento e da tomada de decisões em projetos de construção.</p>	<p>Os resultados da pesquisa mostraram que o método proposto para avaliação da similaridade entre projetos de construção com base nas métricas de WBS teve resultados promissores em conformidade com as propriedades de similaridade, como simetria e reflexividade, com pequenos erros. A métrica de similaridade estrutural demonstrou o melhor desempenho na recuperação de projetos similares, com limiares na faixa de 0,7 a 0,75. Além disso, o estudo identificou oportunidades para futuras pesquisas, como a inclusão de atributos quantitativos importantes, como quantidade de trabalho das tarefas e suas durações, para aprimorar a avaliação de similaridade dos projetos de construção.</p>	<p>O estudo sobre a avaliação quantitativa da similaridade entre projetos de construção usando métricas baseadas em WBS pode ser aplicado em repositórios institucionais para facilitar a recuperação de informações relevantes de projetos passados. Ao utilizar técnicas de processamento de linguagem natural e métricas de similaridade, é possível identificar projetos similares com base em suas estruturas de WBS, auxiliando na gestão do conhecimento e na tomada de decisões em novos projetos. Essa abordagem pode melhorar a eficiência na busca e recuperação de informações em repositórios institucionais, contribuindo para a reutilização de conhecimento e experiências em projetos futuros.</p>
2020	ReBoost: a retrieval-boosted sequence-to-sequence model for neural response generation	<p>O problema de pesquisa abordado no artigo é como melhorar a geração de respostas em conversas entre humanos e computadores, utilizando informações recuperadas como um recurso adicional para</p>	<p>O objetivo do estudo é propor e avaliar o modelo ReBoost, que incorpora respostas recuperadas em um modelo de sequência para sequência de forma hierárquica, visando melhorar a geração de respostas</p>	<p>O estudo concluiu que o modelo ReBoost superou os modelos existentes na geração de respostas, demonstrando maior qualidade e relevância. A análise comparativa, tanto automática quanto humana, destacou o</p>	<p>O modelo ReBoost, que utiliza respostas recuperadas para melhorar a geração de respostas em interações entre humanos e computadores, pode ser aplicado em repositórios institucionais. Essa aplicação</p>

		<p>aumentar a qualidade e a relevância das respostas geradas pelo modelo de sequência para sequência.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>mais informativas e significativas em conversas entre humanos e computadores. O estudo busca demonstrar que a utilização de respostas recuperadas pode aprimorar o processo de geração de respostas em comparação com outros modelos existentes.</p>	<p>desempenho superior do ReBoost em termos de qualidade das respostas geradas. A inclusão de respostas recuperadas como informação adicional mostrou-se eficaz, contribuindo significativamente para a melhoria geral do modelo.</p>	<p>promete aprimorar a experiência do usuário ao proporcionar respostas mais precisas e úteis às consultas, resultando em uma busca de informações mais eficiente e na oferta de serviços mais eficazes nos repositórios institucionais.</p>
2020	Retrieval-Polished Response Generation for Chatbot	<p>O problema de pesquisa abordado neste trabalho é a falta de diversidade e relevância contextual nas interações de chatbot. Os autores buscam melhorar a qualidade das respostas geradas por chatbots, tornando-as mais diversas, fluentes e contextualmente relevantes para proporcionar uma comunicação mais eficaz entre humanos e robôs.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste trabalho é propor um modelo de geração de respostas para chatbots, denominado modelo de "retrieval-polished" (RP), que visa aprimorar a diversidade, fluência e relevância contextual das respostas geradas. O modelo RP busca polir uma resposta inicial com base em um protótipo recuperado, combinando elementos de geração e recuperação para melhorar a qualidade das interações entre humanos e chatbots.</p>	<p>O estudo conclui que o modelo "Retrieval-Polished" (RP) para geração de respostas em chatbots superou modelos concorrentes em fluidez, relevância contextual e diversidade. Experimentos mostraram melhorias significativas, destacando a fluidez notável do modelo. Além disso, o RP alcançou a melhor pontuação de perplexidade (PPL), indicando respostas mais fluentes em comparação com outros métodos avaliados.</p>	<p>O estudo sobre o modelo de "retrieval-polished" (RP) para geração de respostas em chatbots pode ser aplicado em repositórios institucionais para aprimorar a interação e comunicação com os usuários. Ao utilizar o modelo RP, os repositórios institucionais podem oferecer respostas mais diversificadas, fluentes e contextualmente relevantes aos usuários que buscam informações ou suporte. Isso pode melhorar a experiência do usuário, aumentar a eficiência na busca por informações e fortalecer a comunicação entre a instituição e seus stakeholders.</p>
2020	Semantic Comprehension of	<p>O problema de pesquisa abordado neste trabalho é a</p>	<p>O objetivo deste trabalho é propor um método de</p>	<p>O resultado da pesquisa demonstrou que o método</p>	<p>O estudo sobre a compreensão semântica de perguntas em</p>

	<p>Questions in Q&amp;A System for Chinese Language Based on Semantic Element Combination</p>	<p>compreensão semântica de perguntas em sistemas de pergunta e resposta (Q&amp;A) para a língua chinesa. O foco está em desenvolver um método eficaz para entender a semântica das perguntas, o que é fundamental para a recuperação de informações e extração de respostas precisas em sistemas Q&amp;A.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>compreensão de perguntas com base na combinação de elementos semânticos para a língua chinesa. O método visa melhorar a compreensão semântica das perguntas em sistemas Q&amp;A, permitindo a transformação de perguntas em expressões lógicas normalizadas de forma eficaz. A precisão e eficácia do método proposto são avaliadas com o intuito de demonstrar sua capacidade de compreender adequadamente a semântica das perguntas em chinês.</p>	<p>proposto de compreensão de perguntas com base na combinação de elementos semânticos para a língua chinesa alcançou uma precisão de 97,8%. Isso indica que o método foi capaz de entender efetivamente a semântica das perguntas em chinês, convertendo-as em expressões semânticas normalizadas de forma precisa e eficiente. A pesquisa evidenciou a viabilidade e eficácia do método proposto para melhorar a compreensão semântica de perguntas em sistemas Q&amp;A para a língua chinesa.</p>	<p>sistemas Q&amp;A para a língua chinesa, utilizando a combinação de elementos semânticos, pode ser aplicado em repositórios institucionais para otimizar a recuperação de informações e a extração de respostas precisas. Essa abordagem melhora a compreensão de perguntas em chinês, transformando-as eficientemente em expressões lógicas normalizadas, contribuindo para uma busca mais eficaz e obtenção de respostas precisas nos repositórios.</p>
2020	<p>SemBioNLQA: A semantic biomedical question answering system for retrieving exact and ideal answers to natural language questions</p>	<p>O problema de pesquisa abordado no artigo é a necessidade de recuperar respostas de alta qualidade para perguntas em linguagem natural na área biomédica, diante do aumento constante do volume de literatura científica nesse domínio. A dificuldade de absorver todas as informações relevantes, mesmo para especialistas, destaca a importância de tecnologias mais sofisticadas para fornecer respostas relevantes de forma eficiente. Assim, o objetivo é desenvolver um sistema de perguntas e respostas</p>	<p>O objetivo do estudo é desenvolver e avaliar o SemBioNLQA, um sistema de perguntas e respostas biomédicas semânticas, que seja capaz de lidar com diferentes tipos de perguntas, como perguntas de sim/não, factoides, listas e resumos, na área biomédica. O sistema visa fornecer respostas precisas e ideais para perguntas em linguagem natural, contribuindo para a melhoria da precisão e eficácia na recuperação de informações biomédicas.</p>	<p>O resultado da pesquisa demonstrou que o SemBioNLQA, um sistema de perguntas e respostas biomédicas semânticas, é capaz de lidar com uma variedade de tipos de perguntas na área biomédica, fornecendo respostas exatas e ideais. O sistema mostrou-se eficaz na recuperação de informações relevantes e na geração de respostas precisas para perguntas de sim/não, factoides e listas, além de fornecer respostas ideais para perguntas de resumo. Comparado com outros sistemas de perguntas e</p>	<p>O estudo sobre o SemBioNLQA, um sistema de perguntas e respostas biomédicas semânticas, pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações biomédicas. Ao integrar esse sistema em repositórios, os usuários poderão fazer perguntas em linguagem natural e obter respostas precisas e ideais, facilitando a busca por informações relevantes e acelerando a tomada de decisões baseadas em evidências na área biomédica. Essa aplicação pode aprimorar a</p>

		<p>biomédicas semânticas (SemBioNLQA) que seja capaz de recuperar respostas exatas e ideais para perguntas em linguagem natural na área biomédica.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>		<p>respostas biomédicas, o SemBioNLQA obteve bons desempenhos em avaliações experimentais, mostrando-se como uma alternativa prática e competitiva para ajudar os usuários a encontrar respostas exatas e ideais para suas perguntas biomédicas.</p>	<p>eficiência e a qualidade da busca por informações em repositórios institucionais, beneficiando pesquisadores, profissionais de saúde e demais usuários que buscam acesso rápido e preciso a conhecimentos biomédicos.</p>
<p>2020</p>	<p>SIFRank: A New Baseline for Unsupervised Keyphrase Extraction Based on Pre-Trained Language Model</p>	<p>O problema de pesquisa abordado no documento é a extração de palavras-chave de forma não supervisionada a partir de documentos, com o objetivo de identificar um conjunto de palavras ou frases que possam resumir os principais tópicos discutidos no documento. A motivação para esse problema é a aceleração da recuperação de informações, permitindo que as pessoas obtenham rapidamente e com precisão as informações essenciais de um texto longo.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo do estudo é propor o SIFRank, um novo método de extração de palavras-chave baseado em modelos de linguagem pré-treinados, que combina o modelo de incorporação de sentenças SIF e o modelo de linguagem pré-treinado ELMo. O SIFRank visa melhorar o desempenho da extração de palavras-chave em documentos curtos, superando abordagens tradicionais e estabelecendo um novo referencial para a extração de palavras-chave não supervisionada.</p>	<p>Os resultados da pesquisa mostram que o SIFRank supera outros modelos de referência em três conjuntos de dados amplamente utilizados: Inspec, DUC2001 e SemEval2017. O modelo SIFRank alcança um nível de estado-da-arte em termos de precisão, recall e valor F1 na extração de palavras-chave. Além disso, o estudo demonstra que o SIFRank é especialmente eficaz na extração de palavras-chave de documentos curtos, enquanto o SIFRank+ (uma versão aprimorada para documentos longos) melhora significativamente o desempenho em textos extensos.</p>	<p>O estudo sobre o SIFRank, um método de extração de palavras-chave baseado em modelos de linguagem pré-treinados, pode ser aplicado em repositórios institucionais para melhorar a indexação e recuperação de informações. Ao utilizar o SIFRank, os repositórios podem automaticamente extrair palavras-chave relevantes de documentos depositados, facilitando a busca e a organização de conteúdo. Isso pode aumentar a visibilidade e acessibilidade dos materiais armazenados nos repositórios, tornando mais fácil para os usuários encontrar informações específicas de seu interesse.</p>

<p>2020</p>	<p>Transfer Learning for Arabic Named Entity Recognition With Deep Neural Networks</p>	<p>O problema de pesquisa abordado neste estudo é a necessidade de desenvolver métodos eficazes de recuperação e extração de informações para lidar com a vasta quantidade de dados não estruturados disponíveis diariamente. Em particular, o foco está na tarefa desafiadora de Reconhecimento de Entidades Nomeadas (NER) para a língua árabe, devido às suas características especiais e natureza complexa. O objetivo principal é desenvolver um novo modelo de Aprendizado Profundo para o Reconhecimento de Entidades Nomeadas em Árabe Padrão, visando fornecer resultados mais refinados para uso nos campos de Processamento de Linguagem Natural.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é apresentar uma abordagem inovadora de Aprendizado Profundo para o Reconhecimento de Entidades Nomeadas em Árabe Padrão, que demonstre um desempenho superior em comparação com trabalhos anteriores. O principal propósito de construir um novo modelo é fornecer resultados mais refinados para uso nos campos de Processamento de Linguagem Natural. A metodologia proposta utiliza transfer learning com redes neurais profundas para construir um modelo Pooled-GRU combinado com o Codificador de Sentenças Universal Multilíngue. O modelo proposto obteve cerca de 17% de melhoria em comparação com trabalhos anteriores, destacando a busca por avanços na área de Reconhecimento de Entidades Nomeadas em árabe.</p>	<p>Os resultados da pesquisa demonstraram que o novo modelo proposto, Pooled-GRU, utilizando transfer learning com redes neurais profundas, superou significativamente o modelo anterior, Bi-LSTM-CRF, em termos de desempenho para o Reconhecimento de Entidades Nomeadas em árabe. O modelo Pooled-GRU obteve uma melhoria de aproximadamente 17% em relação ao modelo anterior, com uma precisão comprovada de 91.20% em comparação com 75.73% no trabalho anterior. Esses resultados promissores indicam que a abordagem de transfer learning com o modelo Pooled-GRU pode ser altamente eficaz na classificação de entidades nomeadas em árabe, destacando a importância e o potencial do uso de redes neurais profundas e transfer learning nesse contexto específico.</p>	<p>O estudo sobre Transfer Learning for Arabic Named Entity Recognition with Deep Neural Networks apresenta um modelo inovador, o Pooled-GRU, que utiliza transfer learning e redes neurais profundas para melhorar o Reconhecimento de Entidades Nomeadas em árabe. Esse modelo pode ser aplicado em repositórios institucionais para aprimorar a extração de informações de textos em árabe, facilitando a organização e recuperação de dados não estruturados. A precisão aprimorada do modelo Pooled-GRU pode contribuir significativamente para a classificação e identificação de entidades nomeadas em documentos em árabe, melhorando assim a eficiência e a precisão da indexação e busca de informações em repositórios institucionais.</p>
<p>2020</p>	<p>Unambiguous Scene Text Segmentation With Referring Expression Comprehension</p>	<p>O problema de pesquisa abordado neste trabalho é a segmentação de texto em cenas de forma não ambígua, utilizando compreensão de</p>	<p>O objetivo deste trabalho é extrair textos designados e prever máscaras de segmentação de texto em cenas a partir de descrições em</p>	<p>O resultado da pesquisa foi a proposição de um novo framework que efetivamente segmenta instâncias de texto em nível de pixel a partir de</p>	<p>O estudo sobre segmentação de texto em cenas com compreensão de expressões referenciais apresenta uma aplicação significativa em</p>

		<p>expressões referenciais. O objetivo é desenvolver um método que possa segmentar instâncias de texto em nível de pixel a partir de descrições em linguagem natural, permitindo uma segmentação precisa de texto em meio a cenários complexos.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>linguagem natural, possibilitando a segmentação precisa de instâncias de texto em meio a contextos complexos. A proposta visa combinar informações visuais e linguísticas para realizar uma segmentação eficaz de instâncias de texto, superando abordagens convencionais de localização de texto e recuperação de objetos por meio de linguagem natural.</p>	<p>descrições de contexto, por meio da previsão de máscaras de segmentação de texto em cenas de forma não ambígua. Além disso, foi explorada e modelada a relação entre instâncias de texto e seus contextos, utilizando triplets {texto-predicado-contexto}, onde o predicado pode ser espacial, preposicional, comparativo, atributos visuais ou suas combinações. O modelo proposto demonstrou um desempenho significativamente superior em comparação com métodos de referência em um novo conjunto de dados de benchmark, o COCO-CharRef.</p>	<p>repositórios institucionais, contribuindo para a melhoria da extração e organização de informações textuais em imagens. Ao utilizar descrições em linguagem natural para segmentar texto em nível de pixel, esse método oferece vantagens notáveis na indexação e recuperação de documentos visuais. Essa abordagem inovadora facilita a busca e análise de conteúdo textual em repositórios institucionais, proporcionando uma eficiente integração entre linguagem natural e processamento de imagens.</p>
<p>2020</p>	<p>Unified Medical Language System resources improve sieve-based generation and Bidirectional Encoder Representations from Transformers (BERT)-based ranking for concept normalization</p>	<p>O problema de pesquisa abordado neste estudo é a normalização de conceitos, que consiste na tarefa de vincular frases em texto a conceitos em uma ontologia. A normalização de conceitos é fundamental para diversas tarefas subsequentes, como extração de relações e recuperação de informações na área médica.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo deste estudo é apresentar um sistema de normalização de conceitos baseado em um processo de geração e classificação, com o intuito de vincular frases em texto a conceitos em uma ontologia. Esse sistema foi desenvolvido com base na participação dos autores na Tarefa Compartilhada de Desafios Clínicos de Processamento de Linguagem Natural (NLP) Nacional de 2019, na Trilha 3 de Normalização de Conceitos.</p>	<p>Os resultados da pesquisa indicam que o sistema de normalização de conceitos baseado em geração e classificação proposto pelos autores obteve um desempenho significativo. O sistema superou o estado da arte anterior e alcançou a terceira posição em uma competição, demonstrando melhorias na precisão e na capacidade de lidar com conceitos não vistos durante o treinamento. A precisão do modelo foi aprimorada por meio da incorporação de recursos do Unified Medical Language System (UMLS) e da utilização de um regularizador de tipos semânticos do UMLS.</p>	<p>O estudo propõe um sistema de normalização de conceitos que pode ser aplicado em repositórios institucionais para melhorar a vinculação de informações textuais a conceitos em ontologias médicas. Ao utilizar recursos do Unified Medical Language System (UMLS) e técnicas de geração e classificação baseadas em redes neurais, o sistema pode aprimorar a recuperação e organização de informações em repositórios, facilitando a busca e análise de dados clínicos e médicos de forma mais precisa e eficiente.</p>

<p>2020</p>	<p>Word Sense Disambiguation: A comprehensive knowledge exploitation framework</p>	<p>O problema de pesquisa abordado neste trabalho é o da Desambiguação de Sentido de Palavras (Word Sense Disambiguation - WSD). O objetivo é desenvolver um framework abrangente baseado em conhecimento que possa melhorar a capacidade de desambiguação de palavras em textos, levando a resultados mais precisos e eficazes.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>O objetivo deste trabalho é propor um método baseado em conhecimento para a Desambiguação de Sentido de Palavras (WSD) que utilize o espaço semântico e o caminho semântico ocultos por trás de uma determinada sentença. O foco é melhorar a eficácia da desambiguação de palavras, utilizando o Knowledge Base WordNet e técnicas como Análise Semântica Latente (LSA) e PageRank para modelar o espaço semântico e o caminho semântico, respectivamente. O objetivo final é alcançar um desempenho de ponta em vários conjuntos de dados de WSD e contribuir para avanços significativos nessa área de pesquisa.</p>	<p>O framework proposto para a Desambiguação de Sentido de Palavras (WSD) destacou-se ao superar outros sistemas baseados em conhecimento em três conjuntos de dados padrão de WSD. Com um desempenho excepcional na desambiguação de substantivos e verbos, componentes-chave em documentos, a abordagem mostrou potencial significativo. Especialmente na desambiguação de substantivos, alcançou resultados comparáveis ao melhor sistema supervisionado. A pesquisa enfatizou a importância da exploração de relações de sentido em WordNet e do uso de documentos específicos do domínio para aprimorar a desambiguação semântica de palavras.</p>	<p>O estudo sobre Desambiguação de Sentido de Palavras (WSD) utilizando um framework baseado em conhecimento pode ser aplicado em repositórios institucionais para melhorar a precisão na recuperação de informações. Ao incorporar o conhecimento semântico e relacional das palavras em documentos, o sistema pode ajudar a identificar e recuperar informações relevantes com maior eficácia. Isso pode resultar em uma melhor organização e recuperação de documentos em repositórios institucionais, tornando a busca por informações mais eficiente e precisa.</p>
<p>2021</p>	<p>A decade of in-text citation analysis based on natural language processing and machine learning techniques: an overview of empirical studies</p>	<p>O problema de pesquisa abordado no documento é a análise de citações em texto, com foco no uso de técnicas de processamento de linguagem natural e aprendizado de máquina. O estudo visa explorar como essas técnicas podem ser aplicadas para identificar e classificar citações, realizar análises de sentimento em citações, e desenvolver sistemas de recomendação e sumarização baseados em citações. Além disso, o</p>	<p>O objetivo do estudo é revisar narrativamente as pesquisas que utilizaram técnicas de processamento de linguagem natural e aprendizado de máquina para analisar citações em texto. O foco principal está nas publicações que exploraram o uso dessas técnicas para análise de citações em contexto, classificação de citações, análise de sentimento em citações, sumarização baseada em citações e sistemas de recomendação baseados em</p>	<p>O resultado da pesquisa destaca a evolução significativa na análise de citações em texto por meio do uso de técnicas de processamento de linguagem natural e aprendizado de máquina. Os estudos revisados demonstram a aplicação bem-sucedida dessas técnicas para identificar e classificar citações, realizar análises de sentimento em citações, desenvolver sistemas de recomendação e sumarização baseados em citações. Além disso, o estudo</p>	<p>O estudo sobre análise de citações em texto com técnicas de processamento de linguagem natural e aprendizado de máquina pode ser aplicado em repositórios institucionais para aprimorar a identificação e classificação de citações, realizar análises de sentimento em citações, e desenvolver sistemas de recomendação e sumarização baseados em citações. Isso pode melhorar a organização e recuperação de informações em repositórios</p>

		<p>documento discute os desafios associados à análise de citações em texto, como a identificação de contextos de citação relevantes, a generalização dos resultados devido à falta de variedade nos conjuntos de dados utilizados e a limitada disponibilidade de conjuntos de dados completos devido a restrições de direitos autorais.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>citações. O estudo busca fornecer uma visão geral das pesquisas empíricas realizadas nessa área ao longo da última década, destacando os avanços e desafios enfrentados pelos pesquisadores.</p>	<p>ressalta a importância do acesso a bases de dados completas de publicações e a necessidade de abordagens mais abertas e colaborativas na pesquisa nessa área. Os pesquisadores também identificaram desafios, como a dificuldade na extração de contextos de citação devido à variabilidade nos estilos de citação e a falta de disponibilidade de conjuntos de dados e códigos de estudos anteriores para comparação e replicação de resultados.</p>	<p>acadêmicos, facilitando a pesquisa e a descoberta de conteúdo relevante para os usuários. Além disso, a aplicação dessas técnicas pode contribuir para a análise e avaliação da produção científica, auxiliando na identificação de tendências e na avaliação do impacto de publicações.</p>
<p>2021</p>	<p>A New Approach for Video Concept Detection Based on User Comments</p>	<p>O problema de pesquisa abordado neste estudo é a necessidade de aprimorar a detecção de conceitos em vídeos, especialmente em plataformas de mídia social, utilizando comentários de usuários como uma fonte adicional de informações. A pesquisa visa explorar como a análise de comentários de usuários pode complementar a análise de conteúdo visual e melhorar a recuperação de vídeos com base em conceitos específicos.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo deste estudo é investigar como os comentários de usuários em vídeos compartilhados em mídias sociais podem ser utilizados para extrair conceitos semânticos presentes nos vídeos. O estudo não assume que os comentários são uma solução perfeita para a extração de conceitos semânticos, mas busca explorar como esses comentários podem apoiar os esforços de pesquisa que trabalham com o conteúdo visual, seguindo as novas tendências de trabalho do YouTube. O objetivo é mostrar como os comentários podem ser uma fonte valiosa de informações para a detecção de conceitos em vídeos.</p>	<p>Os resultados da pesquisa indicam que os comentários dos usuários são uma fonte valiosa de informações para compreender o conteúdo visual de vídeos. Com taxas de recall superiores a 60% e taxas de precisão acima de 80%, confirmadas por avaliadores humanos, a abordagem proposta mostrou eficácia. A quantidade e qualidade dos comentários influenciam a eficácia da extração de conceitos, sendo observado que a qualidade varia entre categorias de vídeos, sendo melhor em áreas como "telefone e tecnologia", "veículos automotores" e "jogos". Estes resultados enfatizam a necessidade de considerar ambos os aspectos ao analisar</p>	<p>O estudo sobre a extração de conceitos de vídeos com base em comentários de usuários pode ser aplicado em repositórios institucionais para melhorar a recuperação de vídeos com base em conceitos específicos. Ao analisar os comentários dos usuários associados aos vídeos, os repositórios podem aprimorar a indexação e a busca de conteúdo audiovisual, tornando mais fácil para os usuários encontrar vídeos relevantes com base em suas necessidades e interesses. Isso pode melhorar a experiência do usuário e a eficiência na utilização de repositórios institucionais de vídeos.</p>

				comentários para a detecção de conceitos em vídeos.	
2021	A New Enhanced Arabic Light Stemmer for IR in Medical Documents	<p>O problema de pesquisa abordado neste estudo é a melhoria da recuperação de informações em documentos médicos por meio de um novo algoritmo aprimorado de stemmer leve para a língua árabe. O foco está em desenvolver uma abordagem mais eficaz para extrair raízes e hastes de palavras em documentos médicos em árabe, visando aprimorar a precisão e eficiência do processo de recuperação de informações nesse contexto específico.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo deste estudo é apresentar um novo algoritmo aprimorado de stemmer leve para a língua árabe, com foco na recuperação de informações em documentos médicos. O algoritmo proposto visa extrair raízes e hastes de palavras de forma mais precisa e eficiente, utilizando um modelo de Grafo Acíclico Direcionado (DAG) para segmentação aprimorada. O objetivo final é melhorar a eficácia da recuperação de informações em documentos médicos em árabe por meio da aplicação desse novo algoritmo de stemmer.</p>	<p>O resultado da pesquisa foi a introdução de um novo algoritmo aprimorado de stemmer leve para a língua árabe, especificamente voltado para a recuperação de informações em documentos médicos. O algoritmo proposto utiliza um modelo de Grafo Acíclico Direcionado (DAG) para segmentação aprimorada, permitindo a extração mais precisa de raízes e hastes de palavras em árabe. A avaliação comparativa do algoritmo proposto com outros cinco stemmers árabes conhecidos demonstrou que o novo algoritmo superou os demais em termos de eficácia na extração de raízes e hastes, especialmente em conjuntos de dados específicos, como o NAFIS e o Quran Index Corpus. Os resultados indicaram que o algoritmo proposto obteve uma precisão superior na extração de hastes em comparação com os outros stemmers avaliados.</p>	<p>O estudo sobre o novo algoritmo aprimorado de stemmer leve para a língua árabe, focado em documentos médicos, pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações em documentos em árabe. Ao implementar esse algoritmo nos repositórios, é possível aprimorar a precisão e eficiência da busca por conteúdo em língua árabe, especialmente em contextos médicos. Isso pode facilitar a organização, busca e recuperação de documentos relevantes, contribuindo para uma melhor gestão da informação em repositórios institucionais.</p>
2021	A simple and fast method for Named Entity context extraction from patents	<p>O problema de pesquisa abordado no artigo é a extração de informações técnicas de patentes de forma eficiente e precisa. Especificamente, o desafio está em identificar entidades específicas em</p>	<p>O objetivo do estudo é desenvolver um método simples e rápido para extração de contexto de Entidades Nomeadas (Named Entities) a partir de patentes, sem depender de conjuntos de dados</p>	<p>O estudo apresenta um método eficaz para extrair contexto de Entidades Nomeadas (NER) de patentes, utilizando uma abordagem baseada em conhecimento que possibilita a criação automática de conjuntos</p>	<p>O estudo propõe um método de extração de contexto de Entidades Nomeadas a partir de patentes, sem depender de dados anotados manualmente. Esse método pode ser aplicado em repositórios institucionais</p>

		<p>documentos técnicos sem depender de conjuntos de dados anotados manualmente, devido à escassez e ao alto custo de entidades específicas de domínio, à falta de disponibilidade de conjuntos de dados anotados de documentos técnicos e à limitação de tempo de especialistas para realizar a marcação manual. A pesquisa propõe um método baseado em abordagens de Processamento de Linguagem Natural para criar um sistema de Reconhecimento de Entidades Nomeadas (NER) específico de domínio, sem a necessidade de dados anotados manualmente.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>anotados manualmente. Especificamente, o estudo visa criar um sistema de NER específico de domínio para identificar entidades específicas em documentos técnicos, utilizando abordagens de Processamento de Linguagem Natural e técnicas de Aprendizado de Máquina. O foco está em coletar o contexto em que os usuários da invenção aparecem dentro das patentes, avaliar a eficácia do método proposto e discutir sua generalização para outras entidades e documentos.</p>	<p>de treinamento. Três métodos de classificação supervisionada foram testados (árvores de classificação, máquinas de vetores de suporte e redes neurais) para a tarefa específica de NER. Ao focalizar nos contextos em que as entidades aparecem, o método permite uma avaliação mais profunda das extrações. Os resultados, aplicados em patentes de diversas categorias da Classificação Internacional de Patentes (IPC), indicaram uma maior generalização do classificador, especialmente em categorias com diferentes focos tecnológicos. Apesar dos resultados promissores, os autores reconhecem a necessidade de validação humana adicional para garantir a qualidade do método, sugerindo também a combinação com abordagens anteriores para uma extração mais precisa e abrangente de informações de invenções em patentes.</p>	<p>para identificar automaticamente entidades específicas em documentos técnicos, como patentes, melhorando a eficiência na recuperação de informações relevantes. Ao utilizar abordagens de Processamento de Linguagem Natural e Aprendizado de Máquina, é possível automatizar a identificação de entidades específicas, como usuários de invenções, em grandes volumes de documentos, facilitando a organização e recuperação de informações técnicas importantes para instituições de pesquisa e desenvolvimento.</p>
2021	<p>An ensemble model for classifying idioms and literal texts using BERT and RoBERTa</p>	<p>O problema de pesquisa abordado neste estudo é a classificação de expressões idiomáticas e textos literais. O objetivo é desenvolver um modelo de ensemble que possa distinguir com precisão entre expressões idiomáticas e textos literais, utilizando as capacidades de modelos pré-</p>	<p>O objetivo deste estudo é propor um modelo de ensemble para classificar expressões idiomáticas e textos literais utilizando os modelos BERT e RoBERTa, que são modelos de aprendizado profundo pré-treinados. O objetivo é melhorar a precisão na distinção entre expressões idiomáticas e textos</p>	<p>O resultado da pesquisa foi a criação de um modelo de ensemble para classificar expressões idiomáticas e textos literais, utilizando os modelos BERT e RoBERTa, que foram finetuned com o conjunto de dados TroFi. O modelo proposto foi testado em um conjunto de dados contendo 1470</p>	<p>O estudo propõe um modelo de ensemble para classificar expressões idiomáticas e textos literais, utilizando os modelos BERT e RoBERTa. Essa abordagem pode ser aplicada em repositórios institucionais para melhorar a classificação e organização de documentos que contenham expressões</p>

		<p>treinados como BERT e RoBERTa. A pesquisa visa lidar com a complexidade e ambiguidade das expressões idiomáticas na linguagem natural, o que pode ser desafiador para sistemas de processamento de linguagem natural.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>literais, contribuindo para avanços na área de processamento de linguagem natural. Além disso, o estudo visa comparar o desempenho do modelo proposto com os modelos tradicionais e demonstrar a eficácia da abordagem de ensemble para essa tarefa específica de classificação.</p>	<p>expressões idiomáticas e literais, e foi avaliado por especialistas da área. Os resultados mostraram que o modelo de ensemble superou os modelos de referência em termos de métricas como F-score e precisão, com um aumento de 2% na precisão. Isso indica que a abordagem de ensemble utilizando BERT e RoBERTa é eficaz na classificação de expressões idiomáticas e textos literais.</p>	<p>idiomáticas. Ao implementar esse modelo de ensemble, os repositórios institucionais podem aprimorar a precisão na identificação e categorização de textos com expressões idiomáticas, facilitando a recuperação de informações relevantes e a análise de conteúdo de forma mais eficiente.</p>
2021	Arabic Question Answering Systems: Gap Analysis	<p>O problema de pesquisa abordado no documento é a análise das lacunas existentes nos sistemas de perguntas e respostas em árabe. Os autores buscam investigar o estado atual da pesquisa em sistemas de QA em árabe, identificar as experiências realizadas até o momento, avaliar seu grau de sucesso, destacar as lacunas existentes e propor maneiras de melhorar os sistemas de QA em árabe.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade</p>	<p>O objetivo do estudo é analisar os sistemas de perguntas e respostas criados para texto em árabe, por meio da revisão, categorização e análise das lacunas existentes, fornecendo orientações para aqueles que desejam trabalhar nesse campo. Além disso, os autores buscam expandir a pesquisa em diferentes componentes dos sistemas de QA, como análise de perguntas, recuperação de informações e extração de respostas. O estudo também visa fornecer conselhos sobre como os sistemas de QA em árabe podem ser aprimorados.</p>	<p>Os resultados da pesquisa indicam que os recursos disponíveis para sistemas de perguntas e respostas em árabe são limitados e têm tamanhos pequenos. Foram propostos seis conjuntos de dados de QA em árabe para serem usados como referência, e um total de 25 sistemas de QA em árabe foram propostos. Além disso, o estudo fornece uma análise detalhada dos sistemas de QA em árabe, destacando suas contribuições e limitações. Os autores concluem que há uma necessidade de mais pesquisa e desenvolvimento nessa área devido às limitações dos recursos disponíveis.</p>	<p>O estudo sobre sistemas de perguntas e respostas em árabe pode ser aplicado em repositórios institucionais para orientar pesquisadores e desenvolvedores interessados em trabalhar nessa área. Os insights fornecidos no estudo, como a necessidade de grandes conjuntos de dados, a importância de benchmark datasets e o uso de ontologias e dados específicos de domínio, podem orientar a criação e avaliação de sistemas de QA em árabe. Além disso, a análise das lacunas existentes e as recomendações para futuras pesquisas podem direcionar o desenvolvimento de novas</p>

		percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA			soluções e aprimoramentos nos repositórios institucionais que lidam com sistemas de QA em árabe.
2021	Arabic Biomedical Community Question Answering Based on Contextualized Embeddings	<p>O estudo foca na aplicação de modelos de linguagem contextualizada em um contexto de perguntas e respostas na comunidade biomédica árabe. Os autores destacam a importância e os desafios de implementar modelos de aprendizado profundo em idiomas com recursos limitados. Introduzindo uma abordagem inovadora, combinam dois modelos neurais com embeddings contextualizados pré-treinados para aprender representações de sentenças e calcular pontuações para classificação. Os resultados experimentais indicam um desempenho superior aos modelos de referência, com melhorias significativas na média de precisão média (mAP) para as edições de 2016 e 2017.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo da pesquisa é abordar o problema de responder a perguntas da comunidade aplicadas a conjuntos de dados da área médica em árabe, utilizando embeddings contextualizados de última geração. Os autores buscam melhorar a eficácia dos sistemas de perguntas e respostas da comunidade ao calcular a similaridade entre a pergunta original e o par pergunta/resposta, a fim de classificar os pares como relevantes ou irrelevantes. Eles visam demonstrar que o uso de embeddings contextualizados como representações de sentenças leva a resultados superiores, pois esses embeddings são sensíveis ao contexto e capazes de medir a similaridade em um nível semântico, contornando erros sintáticos ou uso de paráfrases.</p>	<p>O modelo proposto pelos autores obteve sucesso em melhorar a eficiência dos sistemas de perguntas e respostas da comunidade na área médica em árabe. Os resultados experimentais mostraram que o modelo superou significativamente os modelos de referência existentes, melhorando a média de precisão média (mAP) em quase 2% para a edição de 2016 e 1% para a edição de 2017. A abordagem de utilizar embeddings contextualizados como representações de sentenças demonstrou ser eficaz, pois esses embeddings são capazes de capturar o contexto e a semântica das perguntas e respostas, resultando em uma classificação mais precisa e relevante dos pares de perguntas e respostas.</p>	<p>O estudo sobre o uso de embeddings contextualizados em sistemas de perguntas e respostas da comunidade na área médica em árabe pode ser aplicado em repositórios institucionais para melhorar a eficiência na busca e recuperação de informações. Ao utilizar modelos de representação de linguagem contextualizada, é possível aprimorar a similaridade entre consultas de usuários e conteúdo nos repositórios, resultando em classificações mais precisas e relevantes de perguntas e respostas. Isso pode facilitar a descoberta de informações relevantes, acelerar a recuperação de dados e melhorar a experiência do usuário ao interagir com os repositórios institucionais.</p>

2021	Content analysis-based documentation and exploration of research articles	<p>O problema de pesquisa abordado no documento é a dificuldade enfrentada pelos usuários, desde usuários gerais até pesquisadores, em satisfazer suas necessidades de informação devido à vasta quantidade de informações disponíveis na World Wide Web. O principal desafio é categorizar os documentos de forma sistemática e considerar dados mais valiosos, como informações semânticas, para obter resultados de busca precisos e eficientes.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo do estudo é desenvolver um sistema de busca baseado em conceitos que aproveita recursos de conhecimento externo como base de conhecimento para obter resultados de busca significativos, precisos e eficientes. O sistema utiliza a Análise Formal de Conceitos (FCA) em conjunto com informações semânticas para apoiar a gestão de documentos na recuperação de informações. Além disso, o estudo propõe um algoritmo de navegação para recuperar a hierarquia de conceitos com base nas consultas dos usuários.</p>	<p>O resultado da pesquisa foi a criação de um sistema de busca conceitual que utiliza informações semânticas de recursos de conhecimento externo, como WordNet e Wikipedia, para gerar uma hierarquia de conceitos de documentos. O sistema baseado em Análise Formal de Conceitos (FCA) foi capaz de lidar com problemas de discrepância semântica nas necessidades dos usuários. O algoritmo de navegação proposto permitiu a integração e exploração aprimoradas de artigos científicos, contribuindo para o avanço da comunidade de pesquisa científica e de engenharia.</p>	<p>O estudo proposto, que utiliza Análise Formal de Conceitos (FCA) com informações semânticas de recursos externos, pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações. Ao criar uma hierarquia de conceitos para os documentos armazenados, o sistema pode facilitar a navegação e a busca de informações relevantes dentro do repositório. Isso pode resultar em uma recuperação mais eficiente e precisa de documentos, beneficiando pesquisadores, estudantes e demais usuários que acessam o repositório institucional.</p>
2021	Context-aware network with foreground recalibration for grounding natural language in video	<p>O problema de pesquisa abordado neste artigo é a correspondência entre linguagem natural e conteúdo de vídeo, especificamente a tarefa de "grounding", que envolve localizar e associar segmentos de vídeo relevantes a consultas em linguagem natural. O desafio principal é a precisão na localização desses segmentos em vídeos não editados, levando em consideração a flexibilidade e complexidade das consultas</p>	<p>O objetivo deste estudo é desenvolver um método eficaz e preciso para a correspondência entre linguagem natural e conteúdo de vídeo, especificamente no contexto de "grounding", onde o desafio é localizar e associar corretamente segmentos de vídeo a consultas em linguagem natural. O objetivo principal é melhorar a precisão e eficiência desse processo, levando em consideração a influência do</p>	<p>Os resultados da pesquisa demonstram que o método proposto, que inclui a utilização da rede context-aware com recalibração de primeiro plano, alcançou um desempenho superior em comparação com métodos existentes para a tarefa de grounding natural language in video. O estudo avaliou o método em três conjuntos de dados de referência (TACoS, Charades-STA e ActivityNet Captions) e obteve resultados de</p>	<p>O estudo sobre a rede context-aware com recalibração de primeiro plano para grounding natural language in video pode ser aplicado em repositórios institucionais para melhorar a correspondência entre consultas em linguagem natural e conteúdo de vídeo. Ao implementar esse método, os repositórios podem aprimorar a recuperação de informações em vídeos não editados, tornando mais eficiente e preciso o</p>

		<p>em linguagem natural e a distribuição desigual de momentos anotados nos vídeos. A proposta do artigo é apresentar um método que considera múltiplas escalas temporais, o papel orientador da semântica da consulta e a ênfase adequada nas regiões de momentos positivos para melhorar a correspondência entre linguagem natural e conteúdo de vídeo.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>ruído de fundo nos vídeos não editados, as complexas relações temporais introduzidas pelas consultas e a distribuição desigual de momentos anotados nos vídeos. O método proposto visa integrar diferentes características temporais, orientação semântica da consulta e recalibração do primeiro plano para aprimorar a correspondência entre linguagem natural e conteúdo de vídeo, alcançando um desempenho de ponta em termos de recuperação de momentos em vídeos.</p>	<p>ponta sem a necessidade de pós-processamento. Especificamente, o método alcançou taxas de acerto de 32,17%, 45,11% e 43,76% sob a métrica Rank@1, IoU@0.5 nos conjuntos de dados TACoS, Charades-STA e ActivityNet Captions, respectivamente. Além disso, foram realizados estudos de ablação para demonstrar a eficácia dos componentes individuais do método proposto, destacando a importância da integração de diferentes características temporais, orientação semântica da consulta e recalibração do primeiro plano para melhorar a recuperação de momentos em vídeos.</p>	<p>processo de associação de segmentos de vídeo relevantes a consultas específicas. Isso pode beneficiar instituições que lidam com grandes volumes de dados de vídeo e desejam melhorar a acessibilidade e a relevância das informações disponíveis em seus repositórios.</p>
2021	<p>Deep Graph Matching and Searching for Semantic Code Retrieval</p>	<p>O problema de pesquisa abordado neste artigo é a recuperação semântica de código, ou seja, a tarefa de encontrar trechos de código em repositórios de código-fonte que correspondam de forma significativa a consultas descritas em linguagem natural. O objetivo é desenvolver um modelo de correspondência e busca de código baseado em redes neurais de grafos profundos para melhorar a precisão e eficácia da recuperação de código com base em informações semânticas mais ricas.</p>	<p>O objetivo deste trabalho é propor um modelo de correspondência e busca de código semântico denominado Deep Graph Matching and Searching (DGMS) baseado em redes neurais de grafos para a tarefa de recuperação de código. O modelo DGMS visa representar consultas em linguagem natural e trechos de código em linguagem de programação como dados estruturados de grafos unificados e, em seguida, utilizar operações de correspondência semântica baseadas em atenção cruzada para calcular a</p>	<p>O resultado da pesquisa demonstrou que o modelo Deep Graph Matching and Searching (DGMS) superou significativamente os modelos de referência do estado da arte em termos de métricas de avaliação em dois conjuntos de dados de recuperação de código público: FB-Java e CSN-Python, que representam as linguagens de programação Java e Python, respectivamente. Os experimentos mostraram que o DGMS obteve um desempenho superior em todas as métricas de avaliação, destacando sua eficácia na recuperação</p>	<p>O estudo sobre Deep Graph Matching and Searching (DGMS) para recuperação semântica de código pode ser aplicado em repositórios institucionais de diversas maneiras. Ao utilizar o modelo DGMS, os repositórios institucionais podem melhorar a eficiência e precisão da recuperação de código, facilitando a busca por trechos de código relevantes para os desenvolvedores. Isso pode acelerar o desenvolvimento de software, permitindo que os programadores encontrem rapidamente soluções pré-</p>

		<p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>similaridade entre eles. O objetivo final é melhorar significativamente o desempenho da recuperação de código em comparação com modelos de referência do estado da arte, demonstrando a eficácia e superioridade do modelo proposto em relação a abordagens tradicionais de processamento de linguagem natural.</p>	<p>semântica de código. Além disso, estudos de ablação foram realizados para avaliar a contribuição de cada parte do modelo DGMS, reforçando a importância e eficácia do modelo proposto.</p>	<p>existentes para problemas específicos. Além disso, a abordagem baseada em grafos e redes neurais pode ajudar a preservar e aprender informações estruturais e semânticas importantes nos códigos armazenados nos repositórios, tornando a busca e a reutilização de código mais eficazes e produtivas para as equipes de desenvolvimento.</p>
2021	<p>Dialogue History Matters! Personalized Response Selection in Multi-turn Retrieval-based Chatbots</p>	<p>O problema de pesquisa abordado no documento é a seleção personalizada de respostas em chatbots baseados em recuperação de informações em diálogos de várias rodadas. O estudo busca melhorar a capacidade dos chatbots de selecionar respostas relevantes e personalizadas com base no histórico de diálogo do usuário, visando aprimorar a interação e a eficácia das conversas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo do estudo é desenvolver um modelo de rede de correspondência híbrida personalizada (PHMN) que leve em consideração o histórico de diálogo do usuário para melhorar a seleção de respostas em chatbots de várias rodadas. O foco é aprimorar a correspondência entre o contexto do diálogo e as respostas possíveis, incorporando informações personalizadas sobre o comportamento linguístico do usuário. O objetivo final é melhorar a personalização e a relevância das respostas fornecidas pelos chatbots durante interações de conversas de várias rodadas.</p>	<p>O resultado da pesquisa demonstrou que o modelo de rede de correspondência híbrida personalizada (PHMN) proposto obteve um desempenho superior em comparação com modelos de referência em dois conjuntos de dados desafiadores: o corpus de diálogo personalizado do Ubuntu (P-Ubuntu) e o conjunto de dados personalizado do Weibo (P-Weibo). O PHMN alcançou um desempenho de ponta ao incorporar informações personalizadas sobre o histórico de diálogo do usuário, como comportamentos linguísticos específicos e pesos de atenção personalizados para as respostas candidatas. Esses resultados destacam a eficácia do modelo na seleção de respostas personalizadas em chatbots de várias rodadas, melhorando a interação e a relevância das conversas.</p>	<p>O estudo sobre a rede de correspondência híbrida personalizada (PHMN) para seleção de respostas em chatbots de várias rodadas pode ser aplicado em repositórios institucionais para melhorar a interação com os usuários. Ao incorporar informações personalizadas do histórico de diálogo dos usuários, como comportamentos linguísticos e preferências, os chatbots podem fornecer respostas mais relevantes e personalizadas. Isso pode aprimorar a experiência do usuário ao buscar informações em repositórios institucionais, tornando as interações mais eficazes e satisfatórias.</p>

2021	Differences in information accessed in a pharmacologic knowledge base using a conversational agent vs traditional search methods	<p>O problema de pesquisa abordado neste estudo é a avaliação das diferenças na informação acessada em uma base de conhecimento farmacológico utilizando um agente de conversação versus métodos de busca tradicionais. Os pesquisadores investigaram como a introdução de um agente de conversação baseado em inteligência artificial impactou a frequência e o tipo de informações acessadas por profissionais de saúde em busca de informações sobre medicamentos. Eles compararam a utilização do agente de conversação com métodos tradicionais de busca, como busca por palavras-chave e navegação por títulos, para entender as diferenças nos tipos de informações acessadas e a frequência de acesso a essas informações.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo foi avaliar as diferenças na frequência e nos tipos de informações acessadas em uma base de conhecimento farmacológico por meio de um agente de conversação baseado em inteligência artificial, em comparação com métodos de busca tradicionais, como busca por palavras-chave e navegação por títulos. Os pesquisadores buscaram entender como a introdução do agente de conversação impactou a forma como os profissionais de saúde acessam informações sobre medicamentos e se houve diferenças significativas na preferência e eficácia de uso entre o agente de conversação e os métodos tradicionais de busca.</p>	<p>Os resultados da pesquisa indicaram que a adição do agente de conversação permitiu aos usuários acessar 22 tipos diferentes de informações contidas na seção de "respostas rápidas" da base de conhecimento farmacológico. Durante o período do estudo, esses tipos de informações foram acessados 117.550 vezes com o agente de conversação, em comparação com 33.649.651 vezes usando métodos de busca tradicionais. A distribuição dos tipos de informações acessadas diferiu de acordo com o método utilizado, com diferenças estatisticamente significativas. Alguns dos tipos de informações mais acessados tanto pelo agente de conversação quanto pelos métodos tradicionais incluíram dosagem de medicamentos, usos (FDA/não FDA), efeitos adversos e administração de medicamentos. Além disso, foi observado que a compatibilidade intravenosa foi acessada com mais frequência usando o agente de conversação, enquanto ajustes de dose foram mais frequentemente acessados por métodos tradicionais.</p>	<p>O estudo sobre as diferenças na informação acessada em uma base de conhecimento farmacológico usando um agente de conversação versus métodos tradicionais pode ser aplicado em repositórios institucionais de saúde para melhorar a eficiência e eficácia na busca de informações sobre medicamentos. A implementação de agentes de conversação baseados em inteligência artificial pode facilitar o acesso a uma variedade de informações farmacológicas de forma mais intuitiva e natural, atendendo às necessidades dos profissionais de saúde. Isso pode resultar em uma melhor utilização dos recursos de informação, maior satisfação do usuário e possivelmente em melhores resultados clínicos. Portanto, a integração de agentes de conversação em repositórios institucionais pode ser uma estratégia promissora para aprimorar a busca e o acesso a informações essenciais sobre medicamentos.</p>
2021	Dynamic Boundary Time Warping for Sub-	O problema de pesquisa abordado no documento é a identificação de trechos de texto	O objetivo do estudo é apresentar um novo método baseado em Dynamic Time	Os resultados da pesquisa mostram que o método Dynamic Boundary Time Warping	O estudo sobre Dynamic Boundary Time Warping (DBTW) pode ser aplicado em

	<p>sequence Matching with Few Examples</p>	<p>com correspondência semântica em um texto longo, com base em um conjunto de exemplos mais curtos. A pesquisa propõe um método inovador que não depende do cálculo da sequência média a partir dos exemplos de consulta, mas sim utiliza os exemplos de consulta como estão, empregando todos eles simultaneamente. O foco está em tarefas de recuperação de consulta por exemplo com poucos exemplos, especialmente no campo do Processamento de Linguagem Natural.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>Warping (DTW) chamado Dynamic Boundary Time Warping (DBTW) para encontrar um fragmento em uma sequência temporal longa que seja semelhante a um conjunto de sequências mais curtas. O método é projetado especificamente para tarefas de recuperação de consulta por exemplo com poucos exemplos, com foco em problemas de correspondência de sub-sequências em Processamento de Linguagem Natural. O estudo visa avaliar o desempenho do DBTW em duas diferentes tarefas de poucos exemplos em NLP e demonstrar que ele supera ou alcança resultados comparáveis em relação a abordagens anteriores quando há um baixo número de exemplos disponíveis.</p>	<p>(DBTW) supera ou alcança resultados comparáveis em relação a abordagens anteriores quando há um baixo número de exemplos disponíveis em tarefas de recuperação de consulta por exemplo em Processamento de Linguagem Natural. O DBTW foi avaliado em duas diferentes tarefas de poucos exemplos em NLP e demonstrou sua eficácia. Em comparação com calcular uma sequência de consenso e utilizá-la para realizar a busca de sub-sequências DTW, o DBTW pode encontrar uma solução aproximada para o problema de consulta por múltiplos exemplos de forma mais eficaz, sem perda desnecessária de informações. Além disso, a inclusão de ponderação de frequência inversa específica para problemas de NLP melhorou ainda mais a eficácia do método. Assim, o DBTW foi capaz de superar métodos anteriormente propostos para a Descoberta de Contratos de Poucos Exemplos com o mesmo Modelo de Linguagem aplicado.</p>	<p>repositórios institucionais para melhorar a recuperação de informações em textos longos, permitindo a identificação de trechos semânticos semelhantes a partir de exemplos mais curtos. Essa abordagem pode aprimorar a busca e recuperação de documentos relevantes em repositórios, especialmente em cenários de poucos exemplos disponíveis. Ao utilizar o DBTW, os repositórios institucionais podem melhorar a eficiência na identificação de informações específicas em documentos extensos, facilitando a pesquisa e análise de conteúdo de forma mais precisa e eficaz.</p>
2021	<p>Enhanced Natural Language Interface for Web-Based Information Retrieval</p>	<p>O problema de pesquisa abordado neste artigo é a limitação na recuperação de informações em sistemas baseados na web devido à falta de suporte para processar consultas de pesquisa expressas em linguagem natural</p>	<p>O objetivo deste trabalho é propor um modelo neural aprimorado, baseado em um framework existente chamado IRNet, para consultas em linguagem natural em bancos de dados. O modelo proposto utiliza uma representação de Gated</p>	<p>Os resultados da pesquisa demonstraram que a abordagem proposta, que inclui a representação de Gated Graph Neural Network (GGNN) e o uso de valores do banco de dados no modelo de previsão, apresentou melhorias significativas em</p>	<p>O estudo propõe um modelo aprimorado de interface de linguagem natural para recuperação de informações em bancos de dados baseados na web. Esse modelo, que incorpora Gated Graph Neural Network (GGNN) e valores do</p>

		<p>(NL). Atualmente, a maioria das interfaces da web para bancos de dados aceita apenas consultas de pesquisa inseridas em alguma forma de combinação lógica de palavras-chave ou strings de texto, o que restringe o escopo e a profundidade do que um usuário da web realmente deseja pesquisar. A falta de uma interface que compreenda consultas em linguagem natural impede que os usuários aproveitem ao máximo as ferramentas de busca disponíveis, mesmo que a recuperação de dados ou informações baseada em linguagem natural tenha avançado significativamente nos últimos anos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>Graph Neural Network (GGNN) para codificar as entidades e relações do banco de dados. Além disso, o sistema representa e utiliza os valores do banco de dados no modelo de previsão para identificar e corresponder nomes de tabelas e colunas, a fim de sintetizar automaticamente uma declaração SQL correta a partir de uma consulta expressa em uma frase em linguagem natural. Os experimentos realizados com um conjunto de dados público demonstram o potencial promissor dessa abordagem.</p>	<p>relação ao modelo original IRNet. Os experimentos realizados com o conjunto de dados público validaram empiricamente os méritos do modelo revisado, mostrando seu potencial para melhorar o desempenho da recuperação de dados e informações baseada na web. Os resultados indicam que a inclusão de GGNN e dos valores do banco de dados no modelo de previsão pode ajudar a superar as dificuldades na correspondência de nomes de colunas devido à falta de detalhes suficientes nas consultas em linguagem natural.</p>	<p>banco de dados na previsão, pode ser aplicado em repositórios institucionais para melhorar a eficiência e precisão da recuperação de informações. Ao utilizar a representação GGNN para codificar entidades e relações do banco de dados, juntamente com a correspondência de valores do banco de dados com nomes de tabelas e colunas, o sistema pode automatizar a geração de consultas SQL corretas a partir de consultas em linguagem natural, facilitando a busca e recuperação de informações em repositórios institucionais de forma mais eficaz.</p>
2021	Event classification from the Urdu language text on social media	<p>O problema de pesquisa abordado no documento é a classificação de eventos a partir de texto em língua urdu nas redes sociais. O estudo visa utilizar classificadores de aprendizado de máquina para</p>	<p>O objetivo do estudo é apresentar a tarefa de classificação de eventos para o texto em língua urdu existente nas redes sociais e nos canais de notícias, utilizando classificadores de aprendizado</p>	<p>A pesquisa sobre a estrutura de sentenças em língua urdu identificou que a combinação de recursos como título, comprimento e últimas quatro palavras é crucial para a classificação eficaz de eventos</p>	<p>O estudo sobre a classificação de eventos a partir de texto em língua urdu em redes sociais pode ser aplicado em repositórios institucionais para aprimorar a extração e categorização de informações</p>

		<p>extrair e classificar eventos a partir de dados multilíngues, lidando com os desafios da falta de recursos nessa área.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>de máquina. O foco é na extração e classificação de eventos a partir de dados multilíngues, visando superar os desafios decorrentes da escassez de recursos nessa área.</p>	<p>multiclasse. Diferentes combinações desses elementos, especialmente título em conjunto com últimas quatro palavras e comprimento, demonstraram resultados notáveis. A utilização de técnicas como Count_Vectorizer e TF-IDF revelou que o título é a característica mais influente para melhorar o desempenho dos modelos de classificação de eventos em nível de sentença. A combinação mencionada alcançou altas precisões de 98,00% e 99,00% para os classificadores Random Forest e k-NN, respectivamente.</p>	<p>multilíngues. A utilização de técnicas de aprendizado de máquina e recursos como título, comprimento e últimas quatro palavras das sentenças pode melhorar a organização e recuperação de dados em repositórios, especialmente em contextos multiculturais. Essa abordagem pode contribuir para uma melhor compreensão e classificação de eventos em diferentes idiomas, facilitando a pesquisa e a análise de conteúdo em ambientes institucionais.</p>
2021	<p>Extending latent semantic analysis to manage its syntactic blindness</p>	<p>O problema de pesquisa abordado no artigo é a limitação do Latent Semantic Analysis (LSA) em compreender a estrutura sintática das sentenças, levando a resultados imprecisos na avaliação da similaridade semântica entre textos. O estudo propõe uma abordagem para melhorar a capacidade do LSA em lidar com informações sintáticas, visando aprimorar a precisão da análise semântica em aplicações de Processamento de Linguagem Natural (PLN).</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade</p>	<p>O objetivo do estudo é propor uma extensão do Latent Semantic Analysis (LSA), denominada xLSA, que incorpora informações sintáticas para superar a limitação do LSA em compreender a estrutura das sentenças. O objetivo é melhorar a precisão na avaliação da similaridade semântica entre textos curtos, visando aprimorar a eficácia do LSA em tarefas de Processamento de Linguagem Natural (PLN).</p>	<p>O resultado da pesquisa demonstrou que a extensão proposta, xLSA, conseguiu mitigar o problema de cegueira sintática do LSA, proporcionando pontuações de similaridade semântica mais realistas. Os testes realizados com pares de frases que continham palavras semelhantes, mas com significados significativamente diferentes, mostraram que o xLSA foi capaz de diferenciar adequadamente entre essas frases, fornecendo pontuações de similaridade mais precisas em comparação com o LSA tradicional. Os resultados foram considerados promissores, indicando que o xLSA pode</p>	<p>O estudo sobre a extensão do Latent Semantic Analysis (LSA) para gerenciar sua cegueira sintática, através do xLSA, pode ser aplicado em repositórios institucionais para aprimorar a análise semântica de textos curtos. Ao incorporar informações sintáticas, o xLSA pode melhorar a precisão na identificação da similaridade entre documentos, facilitando a recuperação de informações relevantes e a organização eficiente de conteúdo em repositórios institucionais. Essa abordagem pode contribuir para uma melhor indexação, busca e recuperação de documentos, tornando a experiência de usuários mais eficiente e eficaz.</p>

		percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA		melhorar a precisão na avaliação da similaridade semântica entre textos.	
2021	Fine-grained Visual Textual Alignment for Cross-Modal Retrieval using Transformer Encoders	<p>O problema de pesquisa abordado neste trabalho é a melhoria da precisão e eficácia da recuperação cruzada de informações entre modalidades visual e textual, especificamente no contexto de recuperação de imagens e sentenças. O foco está em desenvolver uma abordagem que preserve a riqueza informativa de ambas as modalidades por meio de um alinhamento fino entre regiões da imagem e palavras do texto, visando obter resultados de recuperação mais precisos e relevantes.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste trabalho é propor uma abordagem inovadora chamada TERAN (Transformer Encoder Reasoning and Alignment Network) para a recuperação cruzada de informações entre modalidades visual e textual. O TERAN visa obter resultados de estado-da-arte na tarefa de recuperação de imagens em conjuntos de dados como MS-COCO e Flickr30k, superando abordagens existentes. Além disso, o objetivo é melhorar o desempenho na tarefa de recuperação de sentenças no conjunto de dados MS-COCO. A abordagem também busca manter separadas as pipelines de dados visuais e textuais, integrando as informações apenas durante a fase final de alinhamento, antes do cálculo da perda. O trabalho visa contribuir para métodos eficazes e eficientes de recuperação de informações em larga escala entre modalidades cruzadas.</p>	<p>Os resultados da pesquisa mostram que o TERAN alcançou resultados de estado-da-arte na tarefa de recuperação de imagens nos conjuntos de dados MS-COCO e Flickr30k. Além disso, o TERAN superou abordagens existentes na tarefa de recuperação de sentenças no conjunto de dados MS-COCO. A abordagem demonstrou a capacidade de produzir alinhamentos região-palavra finos e interpretáveis sem a necessidade de supervisão a nível de região-palavra. Os experimentos realizados mostraram que o TERAN foi capaz de generalizar melhor em comparação com a abordagem anterior TERN. Os resultados foram avaliados utilizando métricas como Recall@K e NDCG, demonstrando a eficácia do TERAN na recuperação cruzada de informações entre modalidades visual e textual.</p>	<p>O estudo sobre Fine-grained Visual Textual Alignment for Cross-Modal Retrieval using Transformer Encoders destaca o TERAN, uma abordagem inovadora para a recuperação cruzada de informações entre modalidades visual e textual. Este modelo pode ser implementado em repositórios institucionais, aprimorando a precisão e eficácia da recuperação de informações. Ao preservar a riqueza informativa de ambas as modalidades por meio de alinhamentos finos entre regiões da imagem e palavras do texto, o TERAN promove uma busca mais eficiente e relevante de imagens e textos. Sua aplicação contribui para a melhoria da experiência de busca e recuperação de conteúdo em repositórios institucionais, favorecendo a organização e acesso eficazes aos dados disponíveis.</p>
2021	GIS-KG: Building a Large-Scale Hierarchical Knowledge Graph for	<p>A pesquisa destaca a necessidade de aprimorar os sistemas de recuperação de informações na ciência da</p>	<p>O objetivo da pesquisa é melhorar os sistemas de recuperação de informações específicas do domínio para a</p>	<p>Os resultados da pesquisa incluem a construção de um grande grafo de conhecimento hierárquico para a ciência e</p>	<p>O estudo pode ser aplicado em repositórios institucionais para aprimorar a organização e recuperação de informações</p>

	<p>Geographic Information Science</p>	<p>informação geográfica e suas tecnologias. Apesar da existência de ferramentas robustas, como Google Scholar e Semantic Scholar, estas carecem de fornecer insights específicos sobre a incorporação de novo conhecimento em um framework existente. Diante desse desafio, propôs-se a construção do GIS-KG, um extenso grafo de conhecimento hierárquico. Este abrange diversas coleções de conhecimento em Sistemas de Informação Geográfica (SIG) em uma ontologia unificada. A abordagem utiliza inteligência artificial e grafos de conhecimento para representar conexões entre conceitos e propriedades, facilitando uma compreensão aprofundada e exploração do conhecimento geoespacial.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>ampla disciplina da ciência da informação geográfica e suas tecnologias. O foco está em organizar e disseminar o conhecimento de forma mais eficaz, incorporando novos conhecimentos em um framework existente e fornecendo insights focados sobre como o novo conhecimento é incorporado.</p>	<p>tecnologia da informação geográfica, o GIS-KG, que organiza mais de 500 mil publicações dentro de uma estrutura de tópicos integrados com base em métodos de aprendizado profundo. Além disso, a pesquisa demonstrou a utilidade do GIS-KG em tarefas de recuperação de informações em SIG, melhorando a qualidade da recuperação em métodos tradicionais e de aprendizado profundo.</p>	<p>geoespaciais. A construção de um grande grafo de conhecimento hierárquico, como o GIS-KG, pode facilitar a descoberta de conhecimento em publicações científicas relacionadas a SIG. Isso pode melhorar a precisão e relevância das buscas de informações, tornando mais eficiente o acesso a conteúdos relevantes para pesquisadores, estudantes e profissionais que utilizam repositórios institucionais para acessar informações geográficas.</p>
2021	<p>Grounded Vocabulary for Image Retrieval Using a Modified Multi-Generator Generative Adversarial Network</p>	<p>O problema de pesquisa abordado neste trabalho é a melhoria da recuperação de informações em imagens através do uso de um vocabulário fundamentado, combinando informações visuais</p>	<p>O objetivo deste trabalho é apresentar uma abordagem inovadora para a recuperação de imagens, utilizando processamento multi-modal para alinhar e comparar informações de linguagem natural e visual a</p>	<p>Os resultados da pesquisa demonstraram que a abordagem proposta, que combina um modelo modificado de Multi-Generator Generative Adversarial Network (MGAN) com embeddings BERT e mapas</p>	<p>O estudo pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações visuais com base em texto. Ao integrar informações visuais e linguísticas em uma</p>

		<p>e linguísticas de forma integrada para aprimorar a precisão e eficácia da recuperação de imagens com base em texto.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>partir de uma perspectiva equalizada. O modelo proposto visa criar uma biblioteca de vocabulários que são aplicados para a recuperação de imagens, gerando representações visuais que correspondem não apenas ao objeto do texto de entrada, mas também aos detalhes do objeto. O objetivo final é combinar informações de texto e imagem em uma plataforma única para recuperar imagens com alta precisão, mesmo com textos descritivos ou frases curtas como entrada.</p>	<p>de atenção, foi capaz de construir uma biblioteca de vocabulários para a recuperação de imagens. O modelo desenvolvido conseguiu gerar representações visuais que não apenas correspondem ao objeto mencionado no texto de entrada, mas também capturam os detalhes do objeto, resultando em uma recuperação de imagens mais precisa e eficaz. Além disso, a pesquisa destacou a capacidade do modelo de aprender e armazenar vocabulários individuais, permitindo a criação de bibliotecas de vocabulários fundamentados que podem ser aplicadas em diversas tarefas de visão computacional.</p>	<p>representação unificada, o modelo proposto pode melhorar a precisão e eficácia da recuperação de imagens em repositórios institucionais. Isso possibilita uma busca mais precisa e detalhada de imagens com base em descrições textuais, contribuindo para uma melhor organização e acesso aos recursos visuais nos repositórios.</p>
2021	Hierarchical Matching Network for Multi-Turn Response Selection in Retrieval-Based Chatbots	<p>O problema de pesquisa abordado neste trabalho é a seleção apropriada de respostas em chatbots baseados em recuperação, especialmente em cenários de diálogos multi- turnos. Os métodos existentes enfrentam desafios em capturar informações de correspondência local e global entre respostas e contextos de conversação em diferentes níveis, levando a uma seleção de respostas menos eficaz. Assim, o objetivo é desenvolver um modelo hierárquico de correspondência que possa melhorar significativamente o</p>	<p>O objetivo deste trabalho é propor e avaliar um modelo de rede de correspondência hierárquica (HMN) para a seleção de respostas em chatbots baseados em recuperação em diálogos multi- turnos. O HMN é projetado para combinar informações de correspondência em diferentes níveis - nível de palavra e nível de enunciado - a fim de capturar informações de correspondência importantes e complementares para melhorar a seleção de respostas. O objetivo final é demonstrar a eficácia do modelo proposto em comparação com</p>	<p>O estudo sobre o Hierarchical Matching Network (HMN) para seleção de respostas em chatbots pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações e interações com usuários. Ao integrar informações de correspondência em diferentes níveis (palavra e enunciado), o HMN pode aprimorar a precisão na seleção de respostas em diálogos multi- turnos, tornando a interação com os usuários mais eficaz e personalizada. Isso pode beneficiar repositórios institucionais ao facilitar a busca</p>	<p>O estudo sobre o Hierarchical Matching Network (HMN) para seleção de respostas em chatbots pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações e interações com usuários. Ao integrar informações de correspondência em diferentes níveis (palavra e enunciado), o HMN pode aprimorar a precisão na seleção de respostas em diálogos multi- turnos, tornando a interação com os usuários mais eficaz e personalizada. Isso pode beneficiar repositórios institucionais ao facilitar a busca</p>

		<p>desempenho na seleção de respostas em diálogos multi-turnos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>métodos de referência em dois conjuntos de dados públicos de grande escala.</p>	<p>e recuperação de informações relevantes de forma mais precisa e contextualizada, melhorando a experiência do usuário e a eficiência na obtenção de conhecimento.</p>	<p>e recuperação de informações relevantes de forma mais precisa e contextualizada, melhorando a experiência do usuário e a eficiência na obtenção de conhecimento.</p>
2021	Improving Bilingual Word Embeddings Mapping with Monolingual Context Information	<p>O problema de pesquisa abordado neste estudo é a melhoria do mapeamento de embeddings de palavras bilíngues utilizando informações contextuais monolíngues, especialmente em cenários de baixos recursos linguísticos onde recursos bilíngues supervisionados não estão disponíveis.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é propor um método não supervisionado para aprimorar os embeddings de palavras bilíngues, utilizando informações contextuais monolíngues otimizadas, sem a necessidade de corpora paralelos. O foco é melhorar o desempenho de mapeamentos de embeddings de palavras em idiomas de baixos recursos.</p>	<p>Os resultados da pesquisa mostraram que o método proposto superou significativamente outros sistemas de referência, incluindo resultados para quatro pares de idiomas de baixos recursos. A abordagem demonstrou melhorias substanciais nos embeddings de palavras bilíngues, sendo particularmente benéfica para pares de idiomas com poucos recursos disponíveis.</p>	<p>O estudo sobre a melhoria do mapeamento de embeddings de palavras bilíngues com informações contextuais monolíngues pode ser aplicado em repositórios institucionais para aprimorar a tradução automática, a recuperação de informações e a análise de texto em diferentes idiomas. Ao utilizar esse método, os repositórios podem melhorar a precisão e eficiência das ferramentas de processamento de linguagem natural, especialmente em cenários de baixos recursos linguísticos, onde a disponibilidade de recursos bilíngues é limitada.</p>
2021	Information Extraction From Free-Form CV	<p>O problema de pesquisa abordado neste trabalho é a</p>	<p>O objetivo do trabalho é desenvolver modelos de processamento de linguagem</p>	<p>O estudo resultou no desenvolvimento de modelos eficazes de processamento de</p>	<p>O estudo pode ser aplicado em repositórios institucionais para automatizar a extração de</p>

	Documents in Multiple Languages	<p>extração de informações úteis de documentos de CV em formato livre, em múltiplos idiomas. Os autores propõem modelos de processamento de linguagem natural que visam identificar as seções relevantes do documento de CV e extrair informações específicas em diferentes níveis hierárquicos. Para lidar com a diversidade linguística, os modelos são desenvolvidos e testados em cinco idiomas: inglês, sueco, norueguês, polonês e finlandês, pertencentes a diferentes famílias linguísticas. A abordagem adotada utiliza a arquitetura transformer e a implementação multilíngue do codificador na forma do modelo de linguagem BERT.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>natural para extrair informações de CVs em vários idiomas, identificando seções importantes e detalhes específicos, como nomes, datas e habilidades. Os modelos são avaliados em termos de precisão e interpretabilidade, visando facilitar a análise automatizada de CVs para recrutamento.</p>	<p>linguagem natural para extrair informações de CVs em vários idiomas, com alta precisão e desempenho. Os modelos foram capazes de identificar seções importantes dos CVs e detalhes específicos, como nomes, datas e habilidades. Além disso, a pesquisa investigou a interpretabilidade dos modelos, fornecendo insights sobre seu funcionamento.</p>	<p>informações de CVs depositados, facilitando a indexação e recuperação de dados relevantes, como nomes, habilidades e experiências profissionais. Isso pode melhorar a eficiência na gestão de currículos e na identificação de perfis adequados para oportunidades de pesquisa, colaboração e recrutamento dentro da instituição.</p>
2021	Information Retrieval-Based Bug Localization Approach with Adaptive Attribute Weighting	<p>O problema de pesquisa abordado neste estudo é a localização de bugs em software, que é uma etapa crucial no processo de garantia de qualidade de software. A bug localization (BL) é desafiadora e demorada, pois os desenvolvedores precisam entender o fluxo, a estrutura de codificação e a lógica do</p>	<p>O objetivo deste estudo é desenvolver uma ferramenta adaptativa de localização de bugs baseada em recuperação de informações (IRBL) que possa ser utilizada por diversas empresas. A ferramenta BugSTAiR utiliza o algoritmo de ponderação adaptativa de atributos (AAW) e é avaliada em quatro projetos de código aberto</p>	<p>Os resultados da pesquisa demonstram que a ferramenta BugSTAiR supera outras ferramentas de localização de bugs em termos de desempenho, com melhorias significativas nas métricas de avaliação com todos os conjuntos de dados testados. Em comparação com as ferramentas BLIA e BLUIR+, o</p>	<p>O estudo sobre a ferramenta BugSTAiR, uma abordagem de localização de bugs baseada em recuperação de informações com ponderação adaptativa de atributos, pode ser aplicado em repositórios institucionais de software para melhorar a eficiência e precisão na identificação e correção de bugs. Ao utilizar técnicas de IR e</p>

		<p>programa para identificar e corrigir os bugs de forma eficaz.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>conhecidos, com o intuito de melhorar a eficiência e precisão da localização de bugs em software.</p>	<p>BugSTAIR apresentou um aumento de 6,1%, 10% e 30% na métrica MAP. Além disso, o BugSTAIR é a primeira ferramenta genérica de localização de bugs que incorpora o processo AAW, contribuindo de forma significativa para o avanço do estado da arte em BL. A etapa de adaptação do BugSTAIR evita a necessidade de realizar inúmeras experimentações manuais para alcançar os pesos ótimos em todos os conjuntos de dados, tornando mais fácil a ampliação dos conjuntos de dados utilizados.</p>	<p>algoritmos adaptativos, o BugSTAIR pode ajudar os desenvolvedores a localizar rapidamente as partes do código associadas aos relatórios de bugs, reduzindo o tempo e custo de manutenção de software. A incorporação do processo AAW na ferramenta permite uma adaptação fácil a diferentes conjuntos de dados, tornando-a uma solução versátil e eficaz para empresas que lidam com desenvolvimento de software.</p>
2021	<p>Learning to Translate Kannada and English Queries for Mixed Script Information Retrieval</p>	<p>O problema de pesquisa abordado neste trabalho é a tradução de consultas em scripts mistos (como Kannada e Inglês) para facilitar a recuperação de informações em sistemas de recuperação de informações. Os sistemas atuais enfrentam desafios ao lidar com consultas que contêm termos em scripts mistos devido à falta de recursos, como dicionários transliterados e sistemas de tradução automática. A busca semântica para consultas em scripts mistos ainda é um problema não resolvido, especialmente quando aplicado à busca na web. Portanto, o estudo visa desenvolver métodos eficazes para traduzir</p>	<p>O objetivo da pesquisa foi desenvolver e avaliar um modelo de tradução de consultas entre Kannada e Inglês em scripts mistos para melhorar a precisão da recuperação de informações em sistemas de busca, abordando a crescente necessidade de lidar com consultas que combinam palavras de diferentes idiomas e scripts.</p>	<p>O estudo resultou no desenvolvimento de um modelo de Recuperação de Informações em Scripts Mistos (MSIR) que utiliza algoritmos avançados para traduzir consultas entre Kannada e Inglês. Foi criado um dicionário bilíngue e avaliadas três abordagens de tradução, com destaque para o modelo que combinou dicionário, embeddings de palavras e Google Translate. Os resultados mostraram melhorias significativas na precisão da tradução de consultas em scripts mistos, demonstrando a eficácia do modelo proposto para a recuperação de informações em sistemas de busca.</p>	<p>O estudo sobre tradução de consultas em scripts mistos pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações em diferentes idiomas. Ao implementar o modelo MSIR desenvolvido, os repositórios podem oferecer aos usuários a capacidade de pesquisar e acessar conteúdo em vários idiomas, facilitando a descoberta de informações relevantes. Isso pode aumentar a acessibilidade e a utilidade dos repositórios institucionais, tornando o conteúdo mais acessível e relevante para uma audiência global diversificada.</p>

		consultas em scripts mistos para melhorar a precisão e eficácia da recuperação de informações.  <b>Categoria:</b> Recursos de informação  <b>Subcategoria:</b> Representação do documento			
2021	Leaving No Stone Unturned: Flexible Retrieval of Idiomatic Expressions from a Large Text Corpus	O problema de pesquisa abordado neste estudo é a necessidade de encontrar expressões idiomáticas em um corpus de texto grande de forma eficiente, substituindo as expressões regulares por métodos mais eficazes com base na tecnologia de recuperação de informações. O objetivo é preencher a lacuna existente na pesquisa relacionada, uma vez que encontrar idiomas no texto é um requisito para qualquer raciocínio automatizado sobre o uso de idiomas. Nenhuma das abordagens existentes oferece uma solução abrangente para esse problema que seja generalizável e escalável, e este estudo visa abordar essa questão.	O principal objetivo deste estudo é criar um mecanismo de busca de idiomas que, ao receber um idioma como entrada, encontre todas as suas ocorrências em um determinado corpus de texto. O intuito é fornecer uma maior flexibilidade do que as técnicas simples de correspondência de strings comumente usadas em mecanismos de busca existentes, visando auxiliar os linguistas de corpus a investigar como os idiomas são usados no discurso. Além disso, o estudo busca preencher a lacuna na identificação sistemática de menções de idiomas em texto livre, permitindo a automação do reconhecimento de idiomas em texto livre e, assim, possibilitando o uso de idiomas como recursos em aplicações de processamento de linguagem natural (NLP).	Os resultados da pesquisa mostraram a eficácia de uma estratégia de busca sofisticada para recuperar expressões idiomáticas de um grande corpus de texto em comparação com abordagens tradicionais, como busca por palavras-chave e busca por frases. A estratégia proposta superou essas abordagens básicas, que são comumente suportadas por mecanismos de busca, demonstrando ser mais eficaz na recuperação de expressões idiomáticas devido à sua capacidade de lidar com a variação nas formas dos idiomas automaticamente.	O estudo sobre a flexível recuperação de expressões idiomáticas de um grande corpus de texto pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de informações específicas, como expressões idiomáticas, em documentos acadêmicos e de pesquisa. Ao integrar a estratégia de busca sofisticada proposta no estudo em repositórios institucionais, os usuários poderiam encontrar com mais facilidade e precisão expressões idiomáticas em textos acadêmicos, facilitando a pesquisa e análise de idiomas em diferentes contextos linguísticos. Isso poderia beneficiar pesquisadores, estudantes e profissionais que buscam explorar o uso de idiomas em textos acadêmicos e científicos.

		<p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>			
2021	Lyrics segmentation via bimodal text–audio representation	<p>O problema de pesquisa abordado neste estudo é a segmentação das letras de músicas em segmentos significativos, como versos, refrões, pontes, etc., utilizando uma abordagem bimodal que combina informações de texto e áudio. A pesquisa visa superar as dificuldades associadas à variação acústica na música, explorando a rica informação semântica contida nas letras das músicas para melhorar a detecção da estrutura das mesmas.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo deste estudo é desenvolver e avaliar um método de segmentação de letras de músicas que incorpore tanto informações textuais quanto de áudio, a fim de melhorar a detecção da estrutura das músicas. A pesquisa visa superar as limitações das abordagens anteriores que se baseavam apenas em características textuais, explorando a complementaridade entre as modalidades de texto e áudio para aprimorar os modelos de segmentação de letras de músicas.</p>	<p>Os resultados da pesquisa demonstraram que a abordagem proposta, que combina informações de texto e áudio na segmentação de letras de músicas, superou o estado da arte ao detectar com eficácia os limites dos segmentos musicais. Além disso, foi observado que a representação bimodal das letras de músicas, que incorpora características de áudio, obteve um desempenho significativamente melhor do que a abordagem unimodal. A pesquisa também identificou quais características de texto e áudio são mais relevantes para detectar os segmentos de letras de músicas e destacou a complementaridade entre as modalidades de texto e áudio.</p>	<p>O estudo sobre a segmentação de letras de músicas utilizando uma abordagem bimodal texto–áudio pode ser aplicado em repositórios institucionais de música para melhorar a organização e recuperação de informações. Ao incorporar informações de áudio juntamente com as letras das músicas, os repositórios podem oferecer uma indexação mais precisa e detalhada, permitindo aos usuários pesquisar e navegar por músicas com base em elementos textuais e acústicos. Isso pode aprimorar a experiência de descoberta de músicas e facilitar a identificação de músicas com estruturas específicas, como versos, refrões e pontes.</p>
2021	MABAN: Multi-Agent Boundary-Aware Network for Natural Language Moment Retrieval	<p>O problema de pesquisa abordado neste trabalho é a recuperação de momentos de linguagem natural (NLMR) em vídeos, que visa identificar as fronteiras temporais de momentos de vídeo que correspondem a uma descrição em linguagem natural fornecida. O desafio principal é lidar com a</p>	<p>O objetivo deste trabalho é propor um modelo chamado MABAN (Multi-Agent Boundary-Aware Network) para abordar os problemas de seleção limitada de momentos e compreensão insuficiente do contexto estrutural na recuperação de momentos de linguagem natural (NLMR) em vídeos. O MABAN</p>	<p>Os resultados da pesquisa demonstram que o modelo MABAN (Multi-Agent Boundary-Aware Network) proposto neste trabalho supera os métodos do estado da arte em recuperação de momentos de linguagem natural (NLMR) em vídeos. O MABAN aborda com sucesso os desafios de seleção limitada de</p>	<p>O estudo sobre o Multi-Agent Boundary-Aware Network (MABAN) para recuperação de momentos de linguagem natural em vídeos pode ser aplicado em repositórios institucionais para melhorar a busca e recuperação de informações em vídeos com base em descrições em linguagem natural. Ao utilizar o</p>

		<p>seleção limitada de momentos e a compreensão insuficiente do contexto estrutural, buscando desenvolver um modelo que possa ajustar de forma flexível e orientada por objetivos as fronteiras temporais dos momentos candidatos.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>utiliza aprendizado por reforço multiagente para decompor a tarefa de NLMR em localizar os dois pontos de fronteira temporais para cada momento, visando melhorar a seleção de momentos de forma flexível e orientada por objetivos. Além disso, o modelo emprega uma interação cruzada modal em duas fases para explorar informações semânticas contextuais ricas e considera a regressão da distância temporal para aprimorar a compreensão do contexto estrutural. Os experimentos realizados demonstram a eficácia do MABAN em comparação com métodos do estado da arte em dois conjuntos de dados desafiadores.</p>	<p>momentos e compreensão insuficiente do contexto estrutural, proporcionando uma abordagem flexível e orientada por objetivos para a seleção de momentos. Através do uso de aprendizado por reforço multiagente, interação cruzada modal em duas fases e regressão da distância temporal, o MABAN consegue localizar com precisão as fronteiras temporais dos momentos de vídeo que correspondem às descrições em linguagem natural. Os experimentos realizados em dois conjuntos de dados desafiadores, ActivityNet Captions e Charades-STA, validam a eficácia e o desempenho superior do MABAN em comparação com outras abordagens existentes.</p>	<p>MABAN, os repositórios podem aprimorar a precisão e eficiência da recuperação de momentos específicos em vídeos, tornando a pesquisa e navegação em grandes conjuntos de dados de vídeo mais eficazes e orientadas por objetivos. Isso pode beneficiar instituições acadêmicas, empresas e organizações que lidam com grandes volumes de conteúdo de vídeo e buscam formas avançadas de indexação e recuperação de informações.</p>
2021	Meta-search based approach for Arabic information retrieval	<p>O problema de pesquisa abordado no artigo é a necessidade de aprimorar a recuperação de informações em árabe, explorando e combinando diferentes níveis morfológicos de representação de palavras. Os autores destacam que a maioria das pesquisas anteriores em recuperação automática de informações (IR) se concentrou principalmente em índices baseados em raízes ou hastes, enquanto os lemas e padrões são subutilizados. Eles argumentam que cada um dos</p>	<p>O objetivo é combinar diferentes níveis de representação morfológica de palavras em árabe para aprimorar os resultados de recuperação de informações. Os autores buscam comparar as ferramentas de processamento de linguagem natural (NLP) em árabe mais precisas em cada nível de representação, como raízes, hastes, lemas verbais e lemas. Eles propõem a combinação desses níveis de representação por meio de abordagens sofisticadas para</p>	<p>Os resultados da pesquisa indicam que a abordagem baseada em meta-pesquisa, especialmente aquela que utiliza classificação SVM (Support Vector Machine), resultou em melhorias significativas e consistentes nos resultados de recuperação de informações em árabe. A combinação de diferentes níveis de representação de palavras demonstrou aprimorar os resultados da IR de forma significativa em comparação com abordagens mais simples,</p>	<p>O estudo sobre o meta-search baseado em abordagem para recuperação de informações em árabe pode ser aplicado em repositórios institucionais para aprimorar a eficiência e eficácia da busca e recuperação de informações em documentos em árabe. Ao combinar diferentes níveis de representação morfológica de palavras, como raízes, hastes, lemas verbais e lemas, é possível melhorar a precisão dos resultados de busca e facilitar a recuperação de informações relevantes. A</p>

		<p>quatro níveis morfológicos encapsula uma parte do significado das palavras e propõem a combinação desses níveis por meio de abordagens mais sofisticadas para alcançar a combinação ideal que aprimora a IR.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>otimizar a IR e melhorar a eficácia dos sistemas de recuperação de informações em árabe.</p>	<p>como a combinação linear. A abordagem SVM mostrou melhorias consistentes em todos os conjuntos de dados avaliados, destacando sua eficácia na melhoria do desempenho da recuperação de informações em árabe.</p>	<p>utilização de abordagens sofisticadas, como a classificação SVM, pode otimizar a busca em repositórios institucionais, tornando-a mais eficaz e precisa para os usuários que buscam informações em árabe.</p>
2021	Multi-level similarity learning for image-text retrieval	<p>O problema de pesquisa abordado neste estudo é a melhoria da recuperação de informações entre imagens e texto, especificamente no contexto de image-text retrieval. O desafio principal é superar a lacuna de heterogeneidade entre as modalidades visual e textual, que possuem distribuições inconsistentes, a fim de aprimorar a precisão e a eficácia da recuperação de informações entre essas duas modalidades distintas.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo deste estudo é propor um método inovador de aprendizado de representação em vários níveis para a tarefa de recuperação de imagem-texto. O método visa explorar informações em níveis semântico, estrutural e contextual de pares de imagem-texto para reduzir a discrepância entre diferentes modalidades. A meta é melhorar a qualidade das representações visuais e textuais, permitindo uma recuperação mais precisa e eficaz de informações entre imagens e texto.</p>	<p>Os resultados da pesquisa demonstraram a superioridade do método proposto em comparação com os métodos existentes para a tarefa de recuperação de imagem-texto. A avaliação extensiva realizada nos conjuntos de dados Flickr30k e MSCOCO mostrou que o método de aprendizado de similaridade em vários níveis conseguiu melhorar significativamente a qualidade das representações visuais e textuais, resultando em uma recuperação mais eficaz de informações entre imagens e texto. Os experimentos mostraram que o método proposto superou os métodos comparativos em várias métricas de avaliação, validando a eficácia e a relevância da abordagem de aprendizado de</p>	<p>O estudo sobre aprendizado de similaridade em vários níveis para recuperação de imagem-texto pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de informações multimodais, como imagens e texto. Ao utilizar a abordagem proposta, os repositórios podem melhorar a qualidade das representações visuais e textuais, permitindo uma recuperação mais precisa e eficaz de conteúdo. Isso pode beneficiar pesquisadores, acadêmicos e profissionais que buscam informações multimodais em repositórios institucionais, melhorando a experiência de busca e facilitando a localização de conteúdo relevante.</p>

				representação em vários níveis para a recuperação de imagem-texto.	
2021	Natural language inference for Malayalam language using language agnostic sentence representation	<p>O problema de pesquisa abordado no artigo é a inferência de linguagem natural (NLI) para a língua malaiala, que é um idioma de recursos limitados. Os pesquisadores exploram como representações de sentenças independentes de idioma podem ser aplicadas para melhorar a precisão da inferência de linguagem natural em malaiala. Eles buscam desenvolver e avaliar um conjunto de dados (MaNLI dataset) e comparar diferentes modelos para realizar essa tarefa de forma eficaz.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo do estudo é aplicar e avaliar diferentes modelos de representação de sentenças independentes de idioma, como Doc2Vec, fastText, BERT e LASER, para melhorar a precisão da inferência de linguagem natural (NLI) na língua malaiala. Os pesquisadores buscam investigar como esses modelos podem ser eficazes na tarefa de NLI em malaiala, tanto em classificação binária quanto em classificação multiclasse. Eles também procuram comparar o desempenho desses modelos e determinar qual deles é mais adequado para lidar com a inferência de linguagem natural nesse idioma de recursos limitados.</p>	<p>Os resultados da pesquisa indicam que o modelo LASER (Language Agnostic Sentence Representation) superou os outros modelos (Doc2Vec, fastText e BERT) em termos de desempenho na tarefa de inferência de linguagem natural (NLI) para a língua malaiala. O LASER obteve uma melhoria significativa na precisão, com uma margem de 12% para a classificação multiclasse e um aumento de 9% para a classificação binária em comparação com os outros modelos. Isso sugere que as representações de sentenças independentes de idioma, como as fornecidas pelo LASER, podem ser altamente eficazes na realização de tarefas de NLI em idiomas de recursos limitados, como a malaiala.</p>	<p>O estudo sobre inferência de linguagem natural para o idioma malaiala, utilizando representações de sentenças independentes de idioma como o LASER, pode ser aplicado em repositórios institucionais para melhorar a precisão e eficácia de sistemas de processamento de linguagem natural em idiomas de recursos limitados. Ao incorporar essas técnicas avançadas de NLI, os repositórios institucionais podem aprimorar a busca, recuperação e análise de informações em idiomas menos comuns, beneficiando pesquisadores, estudantes e profissionais que trabalham com esses idiomas específicos.</p>
2021	Polysemy Needs Attention: Short-Text Topic Discovery With Global and Multi-Sense Information	<p>O problema de pesquisa abordado é a necessidade de lidar com a polissemia em textos curtos. A polissemia, que se refere à presença de múltiplos significados para uma mesma palavra, pode dificultar a precisão na análise de textos curtos, como tweets e manchetes de notícias, devido à</p>	<p>O objetivo do estudo é propor um modelo de descoberta de tópicos em textos curtos que leve em consideração a polissemia, ou seja, a presença de múltiplos significados para uma mesma palavra. O modelo proposto visa melhorar a precisão na identificação de tópicos em textos curtos, como tweets e manchetes de notícias,</p>	<p>Os resultados da pesquisa indicam que o modelo proposto, denominado MuSeS (Multi-Sense Short-text topic discovery), demonstrou um desempenho superior em comparação com outros modelos existentes, como o GloSS, em termos de descoberta de tópicos em textos curtos. O MuSeS conseguiu</p>	<p>O estudo sobre descoberta de tópicos em textos curtos com informações globais e de múltiplos sentidos pode ser aplicado em repositórios institucionais para aprimorar a organização e recuperação de informações. Ao utilizar o modelo MuSeS, que considera a polissemia e a semelhança semântica entre palavras, os</p>

		<p>falta de contexto. Assim, o estudo busca desenvolver um modelo de tópicos que leve em consideração informações globais e de múltiplos sentidos para melhorar a descoberta de tópicos em textos curtos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>incorporando informações globais e de múltiplos sentidos. Ao fazer isso, o objetivo é aprimorar a qualidade da descoberta de tópicos e a coerência dos mesmos, contribuindo para avanços na análise de textos curtos e na recuperação de informações.</p>	<p>alcançar a melhor coerência de tópicos em diferentes conjuntos de dados, como o conjunto de dados de Perguntas Online e o conjunto de dados de Notícias Online. Além disso, o MuSeS mostrou eficácia na geração de tópicos de alta qualidade, considerando a semelhança semântica entre palavras e a polissemia, o que contribui para uma melhor compreensão e organização dos tópicos presentes nos textos curtos analisados. Os resultados experimentais também destacaram a importância da qualidade dos vetores de palavras de múltiplos sentidos na performance do MuSeS, evidenciando a relevância de considerar informações globais e de múltiplos sentidos na descoberta de tópicos em textos curtos.</p>	<p>repositórios podem melhorar a precisão na identificação de tópicos em documentos curtos, facilitando a navegação e busca por conteúdo relevante. Isso pode beneficiar pesquisadores, estudantes e demais usuários ao tornar a exploração de repositórios institucionais mais eficiente e eficaz, contribuindo para uma melhor compreensão e utilização dos recursos disponíveis.</p>
2021	<p>Question Similarity Measurement of Chinese Crop Diseases and Insect Pests Based on Mixed Information Extraction</p>	<p>O problema de pesquisa abordado neste trabalho é a medição da similaridade de perguntas relacionadas a doenças de culturas agrícolas e pragas de insetos na China. O objetivo é melhorar a precisão dos sistemas de perguntas e respostas de conhecimento agrícola, fornecendo uma abordagem inovadora para medir a similaridade entre perguntas relacionadas a esses tópicos específicos.</p>	<p>O objetivo principal deste trabalho é desenvolver um sistema de medição de similaridade de perguntas relacionadas a doenças de culturas agrícolas e pragas de insetos na China. Esse sistema visa auxiliar na construção de um sistema de perguntas e respostas de conhecimento agrícola mais preciso e eficiente, permitindo a correta identificação e diagnóstico de doenças e pragas, bem como a</p>	<p>Os resultados da pesquisa incluem o desenvolvimento de um método inovador para medir a similaridade de perguntas relacionadas a doenças de culturas agrícolas e pragas de insetos na China. Esse método é projetado para melhorar a precisão e eficiência dos sistemas de perguntas e respostas de conhecimento agrícola, permitindo uma melhor compreensão e classificação das perguntas feitas pelos</p>	<p>O estudo sobre a medição da similaridade de perguntas relacionadas a doenças de culturas agrícolas e pragas de insetos na China pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações e a classificação de perguntas relacionadas a esses tópicos específicos. Ao utilizar o método desenvolvido neste estudo, os repositórios institucionais podem melhorar a precisão na</p>

		<p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>formulação de estratégias de controle eficazes. A medição da similaridade de perguntas é considerada uma tarefa fundamental para a construção desse sistema de conhecimento agrícola.</p>	<p>usuários. Além disso, o estudo aborda desafios específicos, como a presença de terminologia específica e a falta de corpus no campo da medição de similaridade de perguntas relacionadas a doenças de culturas agrícolas e pragas de insetos na China.</p>	<p>identificação e categorização de perguntas sobre doenças de culturas agrícolas e pragas de insetos, facilitando o acesso e a busca por informações relevantes nessa área de conhecimento.</p>
2021	Semantic Similarity Analysis for Corpus Development and Paraphrase Detection in Arabic	<p>O problema de pesquisa abordado neste estudo é a falta de recursos publicamente disponíveis na língua árabe para detecção de paráfrases. A escassez de recursos de paráfrases em árabe dificulta a realização de tarefas de Processamento de Linguagem Natural (PLN) que requerem a compreensão e identificação de sentenças com significados semelhantes. A pesquisa visa preencher essa lacuna desenvolvendo um corpus de paráfrases em árabe e aplicando modelos de aprendizado profundo para melhorar a detecção de paráfrases nesse idioma.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo deste estudo é desenvolver um corpus de paráfrases em árabe com base nas especificidades sintáticas e semânticas das sentenças nesse idioma. Além disso, o estudo visa aplicar um modelo de Rede Neural Convolutiva Recorrente (RCNN) para detectar paráfrases em árabe. A pesquisa busca preencher a lacuna de recursos disponíveis para detecção de paráfrases em árabe e contribuir para a melhoria de tarefas de Processamento de Linguagem Natural, como detecção de plágio e resposta a perguntas, nesse idioma.</p>	<p>O resultado da pesquisa demonstrou que o modelo proposto, baseado em Global Word Embedding (GloVe) e Recurrent Convolutional Neural Network (RCNN), superou os métodos tradicionais no que diz respeito à precisão e recall na detecção de paráfrases em árabe. A abordagem utilizada para desenvolver um corpus de paráfrases preservando as estruturas sintáticas e semânticas das sentenças em árabe mostrou-se eficaz na captura de relações semânticas ocultas, mesmo quando as sentenças não continham palavras em comum ou a coocorrência de palavras era rara. Esses resultados destacam a relevância e eficácia do uso de modelos de aprendizado profundo para a análise de similaridade semântica e</p>	<p>O estudo sobre Análise de Similaridade Semântica para Desenvolvimento de Corpus e Detecção de Paráfrases em Árabe pode ser aplicado em repositórios institucionais para melhorar a detecção de plágio, a resposta a perguntas e a sumarização de documentos em árabe. A utilização do modelo proposto, baseado em GloVe e RCNN, pode aprimorar a análise de similaridade semântica entre documentos, contribuindo para a identificação de paráfrases e a preservação da integridade acadêmica nos repositórios institucionais.</p>

				detecção de paráfrases em árabe.	
2021	Semantically-Aware Retrieval of Oceanographic Phenomena Annotated on Satellite Images	<p>O problema de pesquisa abordado neste estudo é a falta de coincidência entre as informações que podem ser extraídas dos dados visuais e a interpretação desses dados por um usuário em uma determinada situação. A pesquisa visa preencher essa lacuna semântica conectando os elementos quantitativos das imagens de satélite de Observação da Terra com informações qualitativas, modelando esse conhecimento em uma ontologia de fenômenos marinhos e desenvolvendo um mecanismo de resposta a perguntas baseado em linguagem natural que permite a recuperação dos dados mais apropriados para as necessidades de cada usuário.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O principal objetivo deste estudo é realizar a busca de imagens de Observação da Terra relacionadas ao domínio marinho de forma baseada em conteúdo, atendendo a necessidades específicas de aplicação. Isso é alcançado por meio da extração de conhecimento semântico de imagens de satélite, modelagem desse conhecimento usando uma geo-ontologia para fenômenos marinhos e aplicação de técnicas de question answering em uma base de conhecimento semanticamente habilitada, permitindo que os usuários expressem suas necessidades e emitam consultas em linguagem natural.</p>	<p>Os resultados da pesquisa incluem o desenvolvimento de algoritmos para a anotação de imagens de Observação da Terra, a formalização da ontologia de fenômenos marinhos, a determinação e implementação de consultas semânticas para o domínio de aplicação que realiza a recuperação baseada em conteúdo, e a coordenação desses componentes para projetar a arquitetura do mecanismo de busca SeMaRe, que realiza o armazenamento, gerenciamento e recuperação semântica do conhecimento extraído. Além disso, o estudo destaca a importância de abordagens multidisciplinares, envolvendo pesquisadores do domínio marinho, engenheiros de ontologia e especialistas em Processamento de Linguagem Natural para integrar o conhecimento e as experiências de diferentes áreas e facilitar a recuperação de dados relevantes para os usuários.</p>	<p>O estudo sobre a recuperação semântica de fenômenos oceanográficos em imagens de satélite pode ser aplicado em repositórios institucionais para melhorar a busca e recuperação de informações relacionadas ao domínio marinho. Ao integrar ontologias de fenômenos marinhos e técnicas de question answering em sistemas de recuperação de informações, os repositórios institucionais podem oferecer aos usuários a capacidade de emitir consultas em linguagem natural e obter resultados mais relevantes e precisos. Isso pode facilitar a descoberta de dados específicos, como imagens de satélite contendo determinados fenômenos marinhos, e aprimorar a eficiência e eficácia da pesquisa dentro desses repositórios.</p>
2021	Structuring clinical text with AI: Old versus new natural language processing techniques evaluated	<p>O problema de pesquisa abordado neste estudo é a necessidade de estruturar e extrair informações significativas de notas clínicas não estruturadas em registros</p>	<p>O objetivo deste estudo é avaliar e comparar técnicas de processamento de linguagem natural antigas e novas para extrair e prever códigos diagnósticos de oito doenças</p>	<p>A pesquisa revelou que técnicas de processamento de linguagem natural foram eficazes na seleção de códigos diagnósticos para oito doenças cardiovasculares comuns em</p>	<p>O estudo sobre a extração de códigos diagnósticos de doenças cardiovasculares a partir de notas clínicas não estruturadas utilizando técnicas de processamento de linguagem</p>

	<p>on eight common cardiovascular diseases</p>	<p>eletrônicos de saúde para melhorar a qualidade dos dados e facilitar a análise de informações clínicas relacionadas a oito doenças cardiovasculares comuns. A pesquisa visa avaliar e comparar técnicas de processamento de linguagem natural antigas e novas para extrair códigos diagnósticos dessas doenças a partir das notas clínicas, a fim de aprimorar a qualidade dos dados e facilitar a recuperação de informações e aplicativos de mineração de dados no contexto da saúde cardiovascular.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>cardiovasculares comuns a partir de notas clínicas não estruturadas em registros eletrônicos de saúde. O estudo busca melhorar a qualidade dos dados, imputar códigos diagnósticos ausentes e corrigir códigos diagnosticados erroneamente, a fim de facilitar a recuperação de informações e aplicativos de mineração de dados no contexto da saúde cardiovascular. Além disso, o estudo visa demonstrar a viabilidade de extrair com precisão códigos diagnósticos estruturados das notas clínicas não estruturadas, fornecendo interpretabilidade dos modelos e avaliando a transferibilidade dos mesmos em diferentes conjuntos de dados, como o MIMIC-III ICU dataset.</p>	<p>notas clínicas não estruturadas. Os modelos desenvolvidos mostraram alta precisão, com capacidade de generalização entre conjuntos de dados diferentes. A interpretabilidade dos modelos foi evidenciada pela identificação de palavras clinicamente relevantes.</p>	<p>natural pode ser aplicado em repositórios institucionais de saúde para melhorar a qualidade dos dados e facilitar a recuperação de informações clínicas. Ao implementar as técnicas de vetorização de palavras, como TF-IDF, os repositórios podem automatizar a extração de informações relevantes dos registros eletrônicos de saúde, corrigir códigos diagnósticos errôneos e imputar códigos ausentes. Isso resulta em dados mais estruturados e de melhor qualidade, permitindo uma análise mais eficiente e aprimorando a tomada de decisões clínicas com base em informações mais precisas e acessíveis.</p>
2021	<p>The NLP Cookbook: Modern Recipes for Transformer Based Deep Learning Architectures</p>	<p>O problema de pesquisa abordado no documento é a eficiência e o desempenho de modelos de aprendizado profundo baseados em arquiteturas Transformer em tarefas de processamento de linguagem natural (NLP). O foco está em explorar como esses modelos podem ser otimizados para alcançar resultados de ponta em diversas tarefas de NLP, levando em consideração questões como tamanho do modelo, eficiência computacional, transferência de</p>	<p>O objetivo do documento é fornecer uma compreensão detalhada e funcional das diferentes arquiteturas baseadas em Transformer utilizadas em tarefas de processamento de linguagem natural. Além disso, o documento busca apresentar uma taxonomia de designs de NLP, avaliações comparativas e direções futuras na área de processamento de linguagem natural. O foco principal é analisar e resumir os modelos de linguagem mais avançados que contribuíram para o estado-da-</p>	<p>O estudo apresentou uma taxonomia única de modelos de processamento de linguagem natural (NLP) e explorou técnicas de otimização computacional, como transfer learning, pruning e knowledge distillation, para melhorar a eficiência e o desempenho dos modelos baseados em arquiteturas Transformer. Destacou a importância de reduzir o tamanho dos modelos sem comprometer o desempenho, além de enfatizar o uso de Knowledge Retrievers</p>	<p>O estudo sobre arquiteturas de aprendizado profundo em processamento de linguagem natural pode ser aplicado em repositórios institucionais para melhorar a eficiência e precisão na extração e organização de informações de grandes conjuntos de dados textuais. A utilização de técnicas como transfer learning, pruning, quantization e knowledge distillation pode ajudar a reduzir o tamanho dos modelos de linguagem sem comprometer o desempenho, tornando a análise</p>

		<p>conhecimento e extração de conhecimento de grandes corpora de texto.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>arte (SOTA) em desempenho de NLP, destacando técnicas como transfer learning, pruning, quantization e knowledge distillation para otimização de modelos.</p>	<p>para extrair informações de grandes conjuntos de dados textuais de forma precisa e eficiente. Esses resultados contribuem para avanços significativos na área de NLP.</p>	<p>de dados mais eficiente e acessível. Além disso, o uso de Knowledge Retrievers pode facilitar a extração de conhecimento de grandes corpora de texto de forma mais precisa e eficiente, contribuindo para a qualidade e relevância das informações disponíveis nos repositórios institucionais.</p>
2021	Two-stage supervised ranking for emotion cause extraction	<p>O problema de pesquisa abordado neste estudo é a extração de causas de emoção em textos, com o objetivo de reconhecer as causas nas sentenças que provocam determinadas emoções. A pesquisa se concentra em desenvolver um método de classificação supervisionada em duas etapas para extrair com precisão as causas de emoção com base em técnicas de recuperação de informação.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é propor um método de classificação supervisionada em duas etapas para a extração precisa das causas de emoção em sentenças, visando reconhecer as causas que provocam determinadas emoções. O método proposto incorpora técnicas de recuperação de informação e avalia a eficácia da abordagem em um conjunto de dados existente para extração de causas de emoção, demonstrando melhor desempenho em relação aos métodos de referência do estado da arte.</p>	<p>Os resultados da pesquisa demonstraram que o método proposto, baseado em um modelo de classificação supervisionada em duas etapas para extração de causas de emoção, superou significativamente os métodos de referência do estado da arte em termos de desempenho. A abordagem mostrou ser eficaz na extração de causas de emoção em sentenças, o que pode beneficiar a análise de emoções em profundidade para uma computação cognitiva eficaz.</p>	<p>O estudo propõe um método de classificação supervisionada em duas etapas para a extração precisa de causas de emoção em textos. Essa abordagem pode ser aplicada em repositórios institucionais para analisar e compreender as emoções expressas em documentos, auxiliando na categorização e organização de conteúdo com base nas causas emocionais identificadas. Isso pode melhorar a recuperação de informações relevantes e a compreensão do contexto emocional subjacente nos documentos armazenados nos repositórios institucionais.</p>
2021	Venue Topic Model-enhanced Joint Graph Modelling for Citation	<p>O problema de pesquisa abordado no artigo é a recomendação de locais de publicação para acadêmicos em</p>	<p>O objetivo do artigo é propor uma abordagem inovadora para recomendação de locais de publicação acadêmica,</p>	<p>O resultado da pesquisa apresentada no artigo demonstrou que a abordagem proposta, baseada em</p>	<p>O estudo pode ser aplicado em repositórios institucionais para aprimorar a recomendação de locais de publicação para</p>

	Recommendation in Scholarly Big Data	<p>um cenário de grande volume de dados acadêmicos. Com o aumento da quantidade de informações acadêmicas disponíveis, os acadêmicos enfrentam dificuldades em selecionar os locais de publicação adequados para seus trabalhos, o que pode impactar negativamente seu sucesso acadêmico a longo prazo. O artigo propõe um modelo de recomendação de locais de publicação que leva em consideração a influência de tópicos específicos dos locais de publicação, a preferência dos acadêmicos e a interação entre os locais de publicação.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>considerando a influência de tópicos específicos dos locais de publicação, a preferência dos acadêmicos e a interação entre os diferentes locais de publicação. A pesquisa visa melhorar a precisão e eficácia das recomendações de locais de publicação para acadêmicos em um ambiente de grande volume de dados acadêmicos, superando as limitações dos métodos tradicionais que não consideram adequadamente a interação entre acadêmicos e locais de publicação.</p>	<p>modelagem conjunta de grafos e modelos de tópicos de locais de publicação, superou várias técnicas de recomendação de locais de publicação existentes. Os experimentos realizados em dois conjuntos de dados do mundo real mostraram que o modelo proposto obteve um desempenho superior em comparação com métodos do estado-da-arte. Isso sugere que a consideração da influência de tópicos específicos dos locais de publicação, juntamente com a preferência dos acadêmicos e a interação entre os locais de publicação, pode resultar em recomendações mais precisas e eficazes para acadêmicos em um ambiente de grande volume de dados acadêmicos.</p>	<p>acadêmicos. Ao considerar a influência de tópicos específicos dos locais de publicação, a preferência dos acadêmicos e a interação entre os diferentes locais de publicação, o modelo proposto pode ajudar os repositórios institucionais a oferecer recomendações mais precisas e personalizadas aos pesquisadores sobre onde publicar seus trabalhos. Isso pode aumentar a visibilidade e impacto das pesquisas, bem como facilitar a tomada de decisões dos acadêmicos em relação aos locais de publicação mais adequados para seus trabalhos.</p>
2021	WabiQA: A Wikipedia-Based Thai Question-Answering System	<p>O problema de pesquisa abordado no documento é a escassez de estudos e desenvolvimento de sistemas de perguntas e respostas (QA) para idiomas de baixo recurso, como o tailandês. Embora tenham sido realizados extensos estudos e avanços em sistemas de QA para idiomas de alto recurso,</p>	<p>O objetivo do estudo é desenvolver um sistema de perguntas e respostas (QA) eficaz para o idioma tailandês, utilizando artigos da Wikipedia em tailandês como fonte de conhecimento. O sistema proposto, chamado WabiQA, visa responder automaticamente a perguntas em tailandês de</p>	<p>O estudo apresentou o sistema WabiQA, um sistema de perguntas e respostas para o idioma tailandês baseado em artigos da Wikipedia. O sistema foi desenvolvido com um recuperador de documentos BM25F e um leitor de documentos LSTM bidirecional, alcançando sucesso na resposta</p>	<p>O estudo sobre o sistema WabiQA, um sistema de perguntas e respostas para o idioma tailandês baseado em artigos da Wikipedia, pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações e o acesso a conhecimento específico. Ao implementar</p>

		<p>como o inglês, há uma lacuna significativa na pesquisa de sistemas de QA para idiomas de baixo recurso, como o tailandês. O objetivo é desenvolver um sistema de QA preciso e aberto para o idioma tailandês, aproveitando os artigos da Wikipedia em tailandês como fonte de conhecimento.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>forma precisa e eficiente, empregando técnicas de recuperação de documentos e leitura de documentos baseadas em redes neurais profundas. Além disso, o estudo busca analisar e comparar diferentes métodos de recuperação de documentos e avaliar o desempenho do sistema WabiQA em relação a outros sistemas de QA para o idioma tailandês.</p>	<p>a perguntas em tailandês. Comparado a outros sistemas de QA para o idioma tailandês, o WabiQA obteve resultados superiores, vencendo um concurso nacional de software e demonstrando sua eficácia na recuperação de documentos e previsão de respostas. Os resultados destacaram a precisão e eficiência do sistema WabiQA em responder a perguntas em tailandês, mostrando sua capacidade de superar sistemas concorrentes em termos de desempenho.</p>	<p>técnicas de recuperação de documentos e leitura de documentos baseadas em redes neurais profundas, repositórios institucionais podem aprimorar a capacidade de responder a perguntas dos usuários de forma mais precisa e eficiente. Isso pode facilitar a busca por informações relevantes, melhorar a experiência do usuário e aumentar a eficácia na disseminação do conhecimento dentro da instituição.</p>
2022	A Comprehensive Guideline for Bengali Sentiment Annotation	<p>O problema de pesquisa abordado neste artigo é a anotação de sentimentos em textos em língua bengali. O foco está na análise de sentimentos em textos escritos nessa língua, com o objetivo de desenvolver recursos e ferramentas para a análise de sentimentos em Bengali. A pesquisa visa aprimorar a compreensão e o processamento de textos em Bengali, especialmente no que diz respeito à identificação e classificação de sentimentos expressos nesses textos.</p> <p><b>Categoria:</b> Recursos de informação</p>	<p>O objetivo principal deste artigo é propor um conjunto claro e conciso de diretrizes para a anotação de sentimentos em língua bengali. Essas diretrizes visam facilitar a criação de conjuntos de dados anotados com sentimentos de alta qualidade em Bengali, o que é essencial para a aplicação de técnicas de aprendizado de máquina contemporâneas na previsão automática de sentimentos. Além disso, o artigo busca reduzir erros e inconsistências durante a criação de conjuntos de dados anotados, fornecendo</p>	<p>O resultado da pesquisa apresentada neste artigo é a elaboração de um conjunto de cinco categorias de rótulos de sentimentos mais adequados aos falantes da língua bengali. Além disso, foi desenvolvido um guia claro e conciso para a realização da tarefa de anotação de sentimentos em bengali para essas cinco categorias desiguais. Este guia é considerado um dos principais focos do artigo e é destinado a auxiliar os anotadores na previsão de rótulos de sentimentos com maior precisão e consistência, reduzindo significativamente o tempo de</p>	<p>O estudo sobre a anotação de sentimentos em língua bengali pode ser aplicado em repositórios institucionais para melhorar a análise de sentimentos em textos escritos nesse idioma. Ao seguir as diretrizes propostas no estudo, os repositórios institucionais podem aprimorar a classificação e compreensão dos sentimentos expressos em documentos em bengali, o que pode ser útil em diversas áreas, como análise de feedback de clientes, avaliação de pesquisas acadêmicas e monitoramento de opiniões públicas. A aplicação dessas diretrizes pode contribuir para</p>

		<b>Subcategoria:</b> Documento	orientações claras para os anotadores seguirem.	treinamento e a propensão a erros. Os autores destacam que, embora existam várias diretrizes de anotação de sentimentos disponíveis para o idioma inglês, este é o primeiro guia proposto especificamente para a anotação de sentimentos em bengali, o que é fundamental para o sucesso na aplicação de modelos de aprendizado de máquina na previsão de sentimentos.	uma análise mais precisa e eficaz dos sentimentos em textos em bengali, beneficiando a tomada de decisões e a compreensão de informações em repositórios institucionais.
2022	A deep learning based end-to-end system (F-Gen) for automated email FAQ generation	<p>O problema de pesquisa abordado neste estudo é a automação da geração de FAQs a partir do conteúdo de e-mails. A pesquisa visa desenvolver um sistema, denominado F-Gen, que possa identificar grupos de perguntas frequentes semelhantes semanticamente em e-mails e gerar FAQs automaticamente a partir desses grupos. O objetivo é criar um sistema que possa transformar eficientemente o conteúdo de e-mails em FAQs válidas e bem formuladas, com potenciais aplicações em serviços de software, suporte de gerenciamento de TI empresarial e resposta a e-mails de consulta.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo do estudo é desenvolver o sistema F-Gen, um sistema especialista que utiliza técnicas de aprendizado profundo para gerar FAQs automaticamente a partir do conteúdo de e-mails. Os principais objetivos incluem modelar FAQs, construir o sistema F-Gen com subsistemas interoperáveis, identificar parâmetros de uma FAQ válida, automatizar a geração de FAQs e validar as FAQs geradas em comparação com as reais.</p>	<p>O resultado do estudo foi a implementação bem-sucedida do sistema F-Gen, que demonstrou a capacidade de gerar FAQs automaticamente a partir do conteúdo de e-mails. Os principais resultados incluíram a modelagem de FAQs e suas características, a construção do sistema F-Gen com subsistemas interoperáveis, a identificação de parâmetros de uma FAQ válida e a realização de experimentos que resultaram em FAQs com um F-Score ROUGE-1 de 74,10% em comparação com as FAQs reais. Esses resultados destacam a eficácia do sistema F-Gen na geração automatizada de FAQs e seu potencial para melhorar a eficiência e reduzir recursos necessários para essa tarefa.</p>	<p>O estudo sobre o sistema F-Gen para geração automatizada de FAQs a partir de e-mails pode ser aplicado em repositórios institucionais para melhorar a eficiência na identificação e geração de FAQs relevantes. Ao implementar o sistema F-Gen em repositórios institucionais, as organizações podem automatizar o processo de transformar o conteúdo de e-mails em FAQs úteis, economizando tempo e recursos. Isso pode facilitar a recuperação de informações importantes, melhorar a comunicação com os clientes e otimizar o suporte ao cliente.</p>

2022	A disponibilização de vocabulário controlado aos usuários para a recuperação da informação	<p>O problema de pesquisa abordado a análise da disponibilização de vocabulário controlado aos usuários no catálogo online das bibliotecas universitárias brasileiras. O estudo visa verificar como esse recurso está sendo apresentado aos usuários e qual impacto essa disponibilização tem na recuperação da informação. Além disso, a pesquisa busca compreender a relação entre a utilização de linguagem de indexação e a eficiência na busca de informação pelos usuários, considerando a importância da padronização e organização dos termos para facilitar o acesso à informação desejada.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo do estudo é investigar a presença e a disposição dos vocabulários controlados nos catálogos online das bibliotecas universitárias participantes da pesquisa. Além disso, busca-se analisar como esses recursos são apresentados aos usuários, visando melhorar a forma de disponibilização e promover a autonomia dos usuários no processo de recuperação da informação. A pesquisa também tem como propósito contribuir para o aprimoramento das estratégias de busca dos usuários, por meio do acesso facilitado aos vocabulários controlados nos sistemas de informação das bibliotecas.</p>	<p>Os catálogos online das bibliotecas muitas vezes apresentam vocabulários controlados de forma pouco perceptível para os usuários comuns, o que os torna mais facilmente identificáveis pelos bibliotecários familiarizados com essas ferramentas. Recomenda-se, portanto, uma melhor exposição e destaque dos vocabulários controlados nos catálogos, além da oferta de tutoriais e treinamentos para os usuários, visando aprimorar suas habilidades de busca e recuperação de informações. Estes resultados ressaltam a importância de tornar os vocabulários controlados mais acessíveis e compreensíveis para os usuários, com o objetivo de melhorar a eficácia e precisão das pesquisas realizadas nos catálogos online das bibliotecas universitárias.</p>	<p>O estudo sobre a disponibilização de vocabulário controlado aos usuários em catálogos online de bibliotecas universitárias pode ser aplicado em repositórios institucionais para aprimorar a recuperação da informação. Para isso, é importante implementar vocabulários controlados nos repositórios, disponibilizá-los de forma visível e acessível aos usuários, oferecer tutoriais sobre seu uso e garantir uma apresentação clara e compreensível. Essas práticas contribuirão para uma busca mais eficiente de informações nos repositórios, promovendo a autonomia dos usuários e melhorando a experiência de busca acadêmica e científica.</p>
2022	Grammar for interpreting geo-analytical questions as concept transformations	<p>O problema de pesquisa abordado no artigo é a necessidade de desenvolver métodos e ferramentas para interpretar perguntas geo-analíticas de forma a transformá-</p>	<p>O objetivo do estudo é propor um método que permita a interpretação automatizada de perguntas geo-analíticas para sistemas de Perguntas e Respostas (QA) em um contexto</p>	<p>O estudo desenvolveu um método automatizado para extrair transformações conceituais de perguntas geo-analíticas, essenciais para criar workflows analíticos eficazes em</p>	<p>O estudo propõe um método para interpretar perguntas geo-analíticas como transformações conceituais, permitindo a geração automatizada de workflows analíticos em</p>

		<p>las em procedimentos que possam gerar conhecimento declarativo a partir de análises geoespaciais. O foco está em entender como as perguntas relacionadas à geografia podem ser traduzidas em workflows analíticos que utilizam informações espaciais para responder a questões complexas que não possuem respostas diretas e factuais, exigindo análises espaciais para estimar as respostas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>indireto. Especificamente, o objetivo é extrair o conhecimento procedimental contido nas perguntas e utilizá-lo para gerar workflows analíticos que possam responder às perguntas de forma relevante e precisa. O estudo visa abordar desafios como a diversidade sintática das perguntas geo-analíticas, a necessidade de derivar a semântica latente dos processos de transformação e a distinção entre frases funcionais e conteúdo auxiliar nas perguntas. Além disso, busca-se tornar explícito o processo de raciocínio necessário para processar informações espaciais em perguntas geo-analíticas, melhorando a interpretabilidade dos sistemas de QA geográfica.</p>	<p>sistemas de Perguntas e Respostas (QA) indiretas. Testado em 134 perguntas, o método alcançou uma cobertura de 75% das transformações conceituais, fornecendo uma base sólida para interpretação de perguntas de QA indireta. Demonstrou-se eficaz na extração de conhecimento procedimental e na geração de workflows analíticos relevantes para questões geoespaciais complexas. Além disso, o modelo desenvolvido incorpora expertise em Sistemas de Informação Geográfica (GIS), melhorando a interpretabilidade dos sistemas de QA geográfica.</p>	<p>sistemas de Perguntas e Respostas (QA) geográficas. Esse método pode ser aplicado em repositórios institucionais para facilitar a extração de conhecimento procedimental a partir de perguntas relacionadas a informações espaciais. Ao integrar essa abordagem em repositórios institucionais, os usuários poderão formular perguntas complexas sobre dados geográficos e obter respostas relevantes por meio da geração automatizada de workflows analíticos. Isso pode melhorar a eficiência na análise de dados espaciais e facilitar a tomada de decisões baseada em informações geográficas.</p>
2022	<p>A machine learning approach to extracting spatial information from geological texts in Chinese</p>	<p>O problema de pesquisa abordado no artigo é a extração de informações espaciais de textos geológicos em chinês. O foco está na aplicação de uma abordagem de aprendizado de máquina para identificar e extrair relações espaciais latentes presentes nesses textos. A pesquisa visa aprimorar a compreensão e extração de informações espaciais em textos geológicos, contribuindo para a análise e interpretação de dados geocientíficos.</p>	<p>O objetivo do estudo é propor e aplicar uma abordagem de aprendizado de máquina para extrair informações espaciais de textos geológicos em chinês. Especificamente, o estudo visa tratar a extração de relações espaciais como um problema de rotulagem de sequência, a fim de identificar e classificar diferentes tipos de relações espaciais presentes nos textos geológicos. Além disso, o estudo busca validar o modelo proposto, construir visualizações das</p>	<p>Os resultados da pesquisa incluem o desenvolvimento de um modelo de aprendizado de máquina para extrair relações espaciais de textos geológicos em chinês, a classificação das relações espaciais em categorias principais e subcategorias, a validação do modelo proposto com um conjunto de dados de teste, a visualização das relações espaciais extraídas em diferentes tópicos e a quantificação da incerteza no</p>	<p>O estudo sobre a extração de informações espaciais de textos geológicos em chinês usando aprendizado de máquina pode ser aplicado em repositórios institucionais para melhorar a análise e interpretação de dados geocientíficos. A abordagem de extração de relações espaciais pode ajudar na organização e recuperação de informações espaciais em textos geológicos, facilitando a pesquisa e o acesso a conhecimentos relevantes para estudos geológicos e</p>

		<p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>relações espaciais extraídas em diferentes tópicos e quantificar a incerteza no processo de reconhecimento de entidades espaciais até a extração de relações espaciais.</p>	<p>processo de extração de relações espaciais. A pesquisa fornece uma representação detalhada das relações espaciais extraídas, o que pode apoiar a resolução de problemas teóricos e práticos relacionados à cognição, previsão, tomada de decisão e avaliação em geociências.</p>	<p>geoespaciais. Além disso, a aplicação desse método em repositórios institucionais pode contribuir para a construção de bases de dados mais ricas e detalhadas, promovendo avanços na compreensão e uso de informações espaciais em contextos geocientíficos.</p>
2022	A Multistage Retrieval System for Health-Related Misinformation Detection	<p>O problema de pesquisa abordado neste estudo é a detecção de desinformação relacionada à saúde na web, com foco especial na detecção de informações incorretas sobre COVID-19. O objetivo é desenvolver um sistema de recuperação de informações capaz de promover documentos confiáveis e filtrar os não confiáveis, visando fornecer informações precisas e confiáveis aos usuários em busca de orientações de saúde online.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é propor um pipeline completo para a detecção de desinformação relacionada à saúde, utilizando a fusão de múltiplos recursos baseados em conteúdo. O sistema desenvolvido visa combater a disseminação de informações incorretas sobre saúde, com foco em duas tarefas principais: recall total e recuperação ad-hoc. O recall total busca identificar documentos que contradizem as respostas de tópicos específicos, enquanto a recuperação ad-hoc visa classificar informações corretas e confiáveis acima das informações incorretas. Além disso, o estudo busca disponibilizar uma solução tecnológica flexível e modular que possa ser facilmente adaptada para apoiar experimentos em ambas as tarefas de busca.</p>	<p>Os resultados da pesquisa incluem a proposta de um pipeline completo para detecção de desinformação relacionada à saúde, com a realização de experimentos para estudar a influência de cada etapa do pipeline. Além disso, o estudo analisou o trade-off entre a recuperação de conteúdos úteis e prejudiciais, destacando a complexidade de encontrar uma combinação que melhore ambos os aspectos. Os pesquisadores também discutiram as limitações do estudo, como a necessidade de expandir a validação empírica para novos conjuntos de dados e um maior número de tópicos de busca.</p>	<p>O estudo propõe um pipeline avançado para detecção de desinformação em saúde, que pode ser aplicado em repositórios institucionais para melhorar a qualidade e confiabilidade das informações disponíveis. Ao integrar múltiplos recursos de processamento de linguagem natural e técnicas de aprendizado de máquina, o sistema pode ajudar a identificar e filtrar conteúdos incorretos, protegendo os usuários de informações prejudiciais. Essa abordagem pode ser implementada em repositórios institucionais para promover a disseminação de informações precisas e confiáveis, contribuindo para a promoção da saúde pública e a prevenção da propagação de desinformação.</p>

2022	A Non-Exclusive Multi-Class Convolutional Neural Network for the Classification of Functional Requirements in AUTOSAR Software Requirement Specification Text	<p>O problema de pesquisa abordado neste estudo é a classificação automatizada de requisitos funcionais em documentos de Especificação de Requisitos de Software (SRS) da AUTOSAR. A complexidade na análise de requisitos devido à interdependência de subsistemas e a necessidade de lidar com a ambiguidade e desequilíbrio de dados nos documentos são os principais desafios enfrentados. A proposta é desenvolver um modelo de rede neural convolucional multi-classe não exclusivo para lidar com esses desafios e melhorar a precisão na classificação dos requisitos funcionais.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é desenvolver um modelo de rede neural convolucional multi-classe não exclusivo para classificar de forma automatizada os requisitos funcionais presentes nos documentos de Especificação de Requisitos de Software (SRS) da AUTOSAR. O foco está em lidar com a ambiguidade, desequilíbrio de dados e a complexidade na análise dos requisitos funcionais, visando melhorar a precisão e eficiência do processo de classificação.</p>	<p>Os resultados da pesquisa indicam que a abordagem proposta, que utiliza um modelo de rede neural convolucional multi-classe não exclusivo para classificar requisitos funcionais em documentos de SRS da AUTOSAR, obteve uma precisão significativa na classificação. Comparando o desempenho do modelo treinado com embeddings de palavras Word2Vec e FastText dos documentos SRS com modelos de embeddings pré-treinados disponíveis online, observou-se que os modelos re-treinados superaram o modelo inicial em termos de precisão. O modelo de classificação CNN re-treinado com vetores de palavras obteve a maior precisão, atingindo 77%.</p>	<p>O estudo propõe um modelo de rede neural convolucional multi-classe não exclusivo para classificar requisitos funcionais em documentos de Especificação de Requisitos de Software (SRS) da AUTOSAR. Essa abordagem pode ser aplicada em repositórios institucionais para automatizar a classificação de requisitos funcionais em documentos de software, melhorando a eficiência e precisão do processo de análise de requisitos. Ao utilizar técnicas de processamento de linguagem natural e aprendizado profundo, o modelo pode ajudar a organizar e categorizar os requisitos de forma mais eficaz, facilitando a recuperação e o gerenciamento de informações em repositórios institucionais de software.</p>
2022	A probabilistic framework for integrating sentence-level semantics via BERT into pseudo-relevance feedback	<p>O problema de pesquisa abordado neste estudo é como integrar informações semânticas em nível de sentença por meio do modelo BERT (Bidirectional Encoder Representations from Transformers) no feedback de</p>	<p>O objetivo deste estudo é investigar o impacto da incorporação de informações semânticas em nível de sentença no framework proposto sobre os resultados de recuperação de informações,</p>	<p>Os resultados da pesquisa indicam que o framework proposto, que integra informações semânticas em nível de sentença por meio do BERT no feedback de pseudo-relevância, resultou em modelos</p>	<p>O estudo propõe um framework que integra informações semânticas em nível de sentença por meio do BERT no feedback de pseudo-relevância para melhorar a expansão de consultas e a recuperação de</p>

		<p>pseudo-relevância (PRF) para expandir consultas de forma mais eficaz e melhorar o desempenho da recuperação de informações.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>analisar a influência do fator de suavização na robustez dos modelos propostos e melhorar o desempenho da recuperação de informações por meio do aumento dos resultados obtidos em métricas como precisão no rank 10 (P@10), média de precisão média (MAP) e ganho cumulativo descontado normalizado (NDCG) dos 1000 principais resultados.</p>	<p>aprimorados (SRoc, SPRoc2 e SKRoc) em comparação com os modelos originais (Roc, PRoc2 e KRoc) com base em várias métricas de avaliação. Esses resultados destacam a viabilidade e eficácia do framework proposto para incorporar informações semânticas em nível de sentença no processo de expansão de consultas e melhorar a recuperação de informações.</p>	<p>informações. Essa abordagem pode ser aplicada em repositórios institucionais para aprimorar a precisão e relevância dos resultados de pesquisa, permitindo uma melhor correspondência entre as consultas dos usuários e os documentos armazenados. Ao incorporar a semântica em nível de sentença, os repositórios institucionais podem oferecer uma experiência de busca mais eficaz e personalizada, aumentando a satisfação dos usuários e a utilidade do sistema de recuperação de informações.</p>
2022	<p>A survey on textual entailment based question answering</p>	<p>O problema de pesquisa abordado no artigo é a aplicação da inferência textual (textual entailment) em sistemas de question answering. Os autores exploram como a textual entailment pode ser utilizada para melhorar a precisão e eficácia dos sistemas de question answering em processamento de linguagem natural. Eles investigam como a textual entailment pode capturar relações semânticas entre textos e, assim, aprimorar a capacidade dos sistemas de responder perguntas de forma mais precisa e contextualizada.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade</p>	<p>O objetivo do artigo é analisar o papel da textual entailment (inferência textual) na área de question answering (resposta a perguntas) e investigar como essa abordagem pode ser aplicada em diferentes subtasks do question answering para melhorar o desempenho dos sistemas. Os autores buscam identificar e discutir as abordagens de pesquisa existentes que utilizam textual entailment em question answering, examinar conjuntos de dados disponíveis, destacar as limitações enfrentadas ao empregar textual entailment nesse contexto, apresentar os resultados alcançados e apontar direções para pesquisas futuras nessa área.</p>	<p>O resultado da pesquisa apresentada inclui uma análise detalhada das abordagens existentes que utilizam textual entailment em question answering, a identificação de conjuntos de dados relevantes para a tarefa de textual entailment, a discussão sobre as subtasks do question answering em que a textual entailment é aplicada, a exposição das limitações enfrentadas ao empregar textual entailment nesse contexto e a apresentação dos resultados obtidos por meio dessas abordagens. Além disso, o artigo destaca a importância da textual entailment na melhoria do desempenho dos sistemas de question answering e aponta</p>	<p>O estudo sobre textual entailment based question answering pode ser aplicado em repositórios institucionais para melhorar a precisão e eficácia dos sistemas de busca e recuperação de informações. Ao incorporar técnicas de textual entailment, os repositórios podem oferecer respostas mais contextuais e relevantes às consultas dos usuários, facilitando a localização de informações específicas e a compreensão de conteúdos complexos. Isso pode resultar em uma experiência de busca mais eficiente e satisfatória para os usuários dos repositórios institucionais.</p>

		percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA		para possíveis direções de pesquisa futura nessa área.	
2022	AI Ekphrasis: Multi-Modal Learning with Foundation Models for Fine-Grained Poetry Retrieval	<p>O problema de pesquisa abordado neste artigo é a recuperação de poesia de alta qualidade de forma fina e detalhada, utilizando modelos de aprendizado multi-modal com base em imagens e texto. A pesquisa visa superar os desafios de reconhecimento de atributos baseados em imagens e recuperação automática de versos poéticos, especialmente para a apreciação de obras de arte por pessoas com deficiência visual.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é proporcionar uma experiência aprimorada de apreciação de obras de arte para pessoas com deficiência visual, utilizando a recuperação automática de ekphrasis para obras de arte e auxiliando em um ambiente de exposição de arte multimídia abundante. O estudo busca fornecer representações multi-modais detalhadas sobre poesia para combinar imagens com poesia, melhorando a compreensão e apreciação das obras de arte.</p>	<p>Os resultados da pesquisa demonstraram que o modelo proposto superou o desempenho de modelos existentes em recuperação de poesia, tanto em cenários de zero-shot, few-shot e totalmente supervisionados. O modelo proposto considerou o reconhecimento de atributos de forma fina para combinar os poemas mais relevantes com uma imagem de consulta, levando em consideração a associação mútua de cenas, sentimentos e objetos sob considerações de ekphrasis de simbolismo e metáforas. Os resultados de recuperação de poesia mostraram melhorias significativas em comparação com o modelo CLIP em diferentes métricas de avaliação, indicando a eficácia do modelo proposto para a tarefa de recuperação de imagem-poema.</p>	<p>O estudo sobre AI Ekphrasis: Multi-Modal Learning with Foundation Models for Fine-Grained Poetry Retrieval pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações poéticas de forma detalhada e precisa. Ao utilizar modelos de aprendizado multi-modal com base em imagens e texto, é possível melhorar a compreensão e apreciação de poesia, tornando a busca e recuperação de poemas mais eficiente e relevante para os usuários dos repositórios. Isso pode enriquecer a experiência de pesquisa e exploração de conteúdo poético em ambientes acadêmicos e culturais.</p>
2022	An adaptable, high-performance relation extraction system for complex sentences	<p>O problema de pesquisa abordado no artigo é a necessidade de extrair relações semânticas de textos em linguagem natural, especialmente em sentenças complexas, pertencentes a</p>	<p>O objetivo do trabalho é desenvolver um sistema de extração de relações que seja capaz de extrair relações significativas de textos judiciais considerados complexos, com a presença de sentenças</p>	<p>Os resultados da pesquisa mostram que o sistema proposto foi testado em e-julgamentos relacionados a casos criminais, com um conjunto de dados composto por 1500 julgamentos, totalizando 5,00,406 sentenças</p>	<p>O estudo apresenta um sistema de extração de relações adaptável e de alto desempenho, especialmente voltado para textos judiciais complexos. Esse sistema pode ser aplicado em repositórios</p>

		<p>domínios fechados, como o domínio judicial. Os autores destacam as limitações e falhas dos sistemas existentes de extração de relações supervisionados e de Extração de Informações Abertas (OIE) e propõem um sistema de extração de relações adaptável e de alto desempenho para lidar com essas questões.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>complicadas e conjunções. O sistema proposto consiste em três módulos principais: Pré-processamento de texto, Extração de Relações e pós-processamento. O foco da extração de relações está em extrair relações de sentenças complexas contendo verbos frasais, conjunções e outras estruturas linguísticas, especialmente no domínio judicial. O sistema combina abordagens baseadas em conhecimento e de aprendizado semi-supervisionado para alcançar esse objetivo.</p>	<p>e 14,79,648 tokens. O sistema foi capaz de extrair relações de diferentes tipos de sentenças, incluindo aquelas com verbos frasais, modificadores clausais de substantivos, modificadores adjetivos, modificadores adverbiais, e conjunções. Além disso, o sistema foi capaz de inferir relações entre entidades consultando a ontologia do domínio. Os resultados mostraram que o sistema foi eficaz na extração de relações complexas em textos judiciais, superando as limitações de sistemas existentes.</p>	<p>institucionais para automatizar a extração de relações semânticas de documentos legais, facilitando a organização e recuperação de informações relevantes. Ao integrar o sistema proposto com repositórios institucionais, é possível melhorar a eficiência na análise e extração de conhecimento de grandes volumes de textos, contribuindo para uma gestão mais eficaz e uma pesquisa mais precisa dentro desses repositórios.</p>
2022	<p>Attention Retrieval Model for Entity Relation Extraction From Biological Literature</p>	<p>O problema de pesquisa abordado neste artigo é a extração de relações entre entidades em textos biológicos, com foco na complexidade e na subjetividade envolvidas nesse processo. Os autores destacam a dificuldade de determinar com precisão as relações entre entidades biológicas devido à diversidade de interpretações possíveis, mesmo entre especialistas. Eles propõem o modelo ARM (Attention Retrieval Model) como uma abordagem inovadora para lidar com esses desafios e melhorar a extração de relações em textos biológicos.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>O objetivo deste estudo é desenvolver e avaliar um modelo de recuperação de atenção (ARM) para extração de relações entre entidades em textos biológicos. O ARM visa melhorar a precisão e a eficiência na identificação e extração de relações complexas entre entidades biológicas, utilizando um mecanismo de autoatenção para aprender associações entre entidades em sequências de texto. Além disso, o ARM busca fornecer uma estrutura flexível que possibilite a personalização do modelo, a integração de conhecimento especializado e a extração de relações não anotadas nos dados de treinamento, contribuindo assim para uma</p>	<p>O estudo revelou que o modelo ARM (Attention Retrieval Model) proposto se mostrou eficiente na extração de relações entre entidades em textos biológicos. Utilizando um mecanismo de autoatenção, o ARM conseguiu aprender associações entre diferentes entidades biológicas, como genes, proteínas e fenótipos, e lidar com relações complexas e subjetivas presentes nesses textos. Além disso, o ARM demonstrou habilidade em extrair relações não anotadas nos dados de treinamento, indicando um potencial para aplicação em cenários de aprendizado de zero-shot.</p>	<p>O estudo sobre o Attention Retrieval Model (ARM) para extração de relações em textos biológicos pode ser aplicado em repositórios institucionais para melhorar a extração de informações e relacionamentos entre entidades biológicas em documentos científicos. Ao utilizar o ARM, os repositórios podem aprimorar a indexação e recuperação de informações relevantes, facilitando a busca e análise de dados biológicos. A flexibilidade e eficácia do ARM permitem uma abordagem personalizada e precisa na extração de relações, contribuindo para uma melhor organização e compreensão dos conteúdos presentes nos repositórios institucionais.</p>

			abordagem mais prática e abrangente de extração de relações em textos biológicos.		
2022	BaNeP: An End-to-End Neural Network Based Model for Bangla Parts-of-Speech Tagging	<p>O problema de pesquisa abordado no artigo é a falta de recursos linguísticos adequados para a marcação de partes do discurso em idiomas de baixo recurso, como o Bangla. O objetivo é desenvolver um modelo neural eficaz que possa lidar com a marcação de partes do discurso em Bangla, superando as limitações dos métodos tradicionais e melhorando a precisão e eficiência dessa tarefa linguística específica.</p> <p><b>Categoria:</b> Recursos de informação  <b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo do estudo apresentado no artigo é propor um modelo neural inovador, chamado BaNeP, para determinar as partes do discurso de palavras em Bangla. O objetivo principal é desenvolver um sistema preciso e eficiente que possa extrair características estruturais e contextuais das palavras em Bangla, a fim de melhorar a marcação de partes do discurso nesse idioma de baixo recurso. O modelo BaNeP visa superar as limitações dos métodos tradicionais e gerar previsões mais precisas das partes do discurso em Bangla.</p>	<p>O resultado da pesquisa apresentada no artigo demonstrou que o modelo BaNeP é altamente capaz de extrair informações contextuais detalhadas que influenciam as partes do discurso de uma palavra em Bangla. A análise da arquitetura do modelo e a avaliação experimental mostraram que o BaNeP supera as abordagens tradicionais, destacando sua eficácia na marcação de partes do discurso em Bangla. O modelo foi projetado para lidar com as características estruturais e contextuais das palavras em Bangla, demonstrando sua capacidade de melhorar a precisão e eficiência da marcação de partes do discurso nesse idioma de baixo recurso.</p>	<p>O estudo sobre o modelo BaNeP para marcação de partes do discurso em Bangla pode ser aplicado em repositórios institucionais para aprimorar a indexação e recuperação de informações em documentos escritos nesse idioma. Ao integrar o BaNeP em sistemas de processamento de linguagem natural de repositórios, é possível melhorar a precisão da marcação de partes do discurso em textos em Bangla, facilitando a busca e organização de conteúdo. Isso pode beneficiar a descoberta de informações relevantes e aprimorar a experiência de usuários que acessam documentos em Bangla nos repositórios institucionais.</p>
2022	BERT-based response selection in dialogue systems using utterance attention mechanisms	<p>O problema de pesquisa abordado neste estudo é a seleção eficaz de respostas em sistemas de diálogo, especialmente em contextos de conversas multi-turno, onde a intenção do interlocutor pode não ser óbvia. O objetivo é melhorar a precisão e a relevância das respostas</p>	<p>O objetivo deste estudo é propor um novo modelo de seleção de respostas que faça uso completo de mecanismos de atenção, incluindo autoatenção e atenção cruzada, para melhor compreender as relações entre palavras e entre enunciados em um contexto de diálogo. O modelo visa aprimorar a seleção de respostas em sistemas de diálogo, proporcionando alta</p>	<p>O estudo apresentou resultados promissores, demonstrando melhorias significativas no desempenho do modelo proposto em comparação com modelos de linha de base em diversos conjuntos de dados de diálogo. Em particular, o modelo proposto obteve um aumento de aproximadamente 2.1%p na métrica hit@1 no conjunto de dados DSTC8 ubuntu. Além</p>	<p>O estudo propõe um modelo avançado de seleção de respostas em sistemas de diálogo, utilizando mecanismos de atenção em nível de palavra e de enunciado. Essa abordagem pode ser aplicada em repositórios institucionais para aprimorar a interação com usuários, fornecendo respostas mais precisas e relevantes em diálogos automatizados. A alta</p>

		<p>selecionadas em sistemas de diálogo, utilizando mecanismos avançados de atenção em nível de palavra e de enunciado, com base em modelos pré-treinados como BERT e RoBERTa.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>portabilidade ao poder ser facilmente combinado com outros modelos existentes e diversos modelos pré-treinados. Além disso, o estudo busca demonstrar melhorias significativas no desempenho do modelo ao ser aplicado em diferentes conjuntos de dados de diálogo.</p>	<p>disso, foram observadas melhorias significativas de desempenho em outros conjuntos de dados avaliados. Esses resultados indicam que a abordagem de atenção em nível de palavra e de enunciado pode contribuir positivamente para a seleção de respostas em sistemas de diálogo, destacando a eficácia do modelo proposto.</p>	<p>portabilidade do modelo permite sua integração fácil com outros sistemas existentes e modelos pré-treinados, o que pode melhorar a eficiência e a qualidade das interações em repositórios institucionais.</p>
2022	Beyond belief: a cross-genre study on perception and validation of health information online	<p>O problema de pesquisa abordado neste estudo é a percepção e validação de informações de saúde online, com foco nas dificuldades enfrentadas ao interpretar e verificar notícias médicas. Os pesquisadores exploram como a interpretação errônea e a representação inadequada impactam a percepção das informações de saúde online, destacando os desafios em identificar as descobertas originais em artigos de newswire relacionados a notícias médicas. Além disso, o estudo busca entender como o conjunto de dados de frases de notícias médicas pode contribuir para melhorar a compreensão e a credibilidade das informações de saúde online.</p>	<p>O objetivo deste estudo é investigar a percepção e validação de informações de saúde online, com foco em identificar o que vale a pena verificar em artigos de saúde online, considerando marcadores de credibilidade, e avaliar se a percepção fornecida pelos marcadores aos leitores em geral é realmente verdadeira. Além disso, busca-se estudar como os marcadores de credibilidade em notícias médicas online podem influenciar a crença dos leitores nas mensagens, desenvolver modelos capazes de identificar casos em que artigos de saúde online podem ser enganosos, mesmo quando fornecem citações explícitas, e investigar a metamorfose da informação entre dois gêneros distintos</p>	<p>Os resultados da pesquisa destacam a importância dos marcadores de credibilidade, como as citações a pesquisas revisadas por pares, na percepção da credibilidade das informações de saúde online. Além disso, evidenciam a dificuldade em identificar as descobertas originais em artigos de newswire relacionados a notícias médicas, o que pode levar a interpretações errôneas e representações inadequadas. A pesquisa ressalta a necessidade de modelos capazes de identificar casos em que artigos de saúde online podem ser enganosos, mesmo quando fornecem citações explícitas. Também enfatiza a relevância de estudar a metamorfose da informação entre diferentes gêneros (notícias e pesquisas</p>	<p>O estudo sobre percepção e validação de informações de saúde online pode ser aplicado em repositórios institucionais para melhorar a credibilidade e confiança nas informações disponíveis. Isso pode ser feito através da implementação de marcadores de credibilidade, como citações a pesquisas revisadas por pares, para ajudar os usuários a avaliar a confiabilidade das informações. Além disso, os repositórios podem desenvolver modelos de verificação de fatos para identificar e corrigir possíveis informações enganosas, mesmo quando acompanhadas de citações explícitas. Essas práticas podem contribuir para uma melhor avaliação e interpretação das informações de saúde disponíveis nos</p>

		<p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>(notícias e pesquisas médicas) e examinar a verificação de fatos relacionados à saúde em ambos os contextos. O estudo também visa contribuir para a compreensão da credibilidade, confiança e representação inadequada de informações médicas, visando ajudar os leitores a avaliar com mais discernimento as reivindicações médicas que normalmente seriam consideradas como verdadeiras.</p>	<p>médicas) para compreender a verificação de fatos relacionados à saúde. Por fim, destaca a importância de ajudar os leitores a avaliar com mais discernimento as reivindicações médicas, considerando a credibilidade, confiança e representação adequada das informações de saúde online.</p>	<p>repositórios institucionais, promovendo uma compreensão mais crítica e precisa por parte dos usuários.</p>
2022	<p>Building siamese attention-augmented recurrent convolutional neural networks for document similarity scoring</p>	<p>O problema de pesquisa abordado neste estudo é a medição automática da similaridade entre documentos em processamento de linguagem natural, utilizando o modelo S-ARCNN. O objetivo é desenvolver uma abordagem eficaz para identificar se pares de documentos são duplicados ou não, o que é essencial em diversas aplicações, como detecção de plágio, recomendação de livros e auxílio em fóruns de ajuda online.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo deste estudo é desenvolver e avaliar um modelo de rede neural denominado Siamese Attention-augmented Recurrent Convolutional Neural Network (S-ARCNN) para pontuar pares de documentos em sua similaridade. O foco é avaliar o desempenho do S-ARCNN em identificar se questões duplicadas são semelhantes ou não, comparando-o com outros modelos como S-CNN, S-LSTM e regressão logística. O estudo busca melhorar a precisão na classificação de pares de documentos e explorar a combinação de diferentes arquiteturas de redes neurais para essa tarefa.</p>	<p>Os resultados da pesquisa indicaram que o modelo S-ARCNN proposto obteve um desempenho superior em comparação com outros modelos, como S-CNN, S-LSTM e regressão logística, na tarefa de pontuação de similaridade entre pares de documentos. Especificamente, o S-ARCNN demonstrou melhor desempenho em identificar a similaridade entre questões duplicadas com sentenças mais longas, ou seja, com um comprimento total de mais de 50 palavras. Isso sugere que o S-ARCNN pode ser especialmente eficaz ao lidar com documentos extensos, como resumos de artigos científicos, o que tem implicações práticas em áreas como identificação de colaboradores com interesses de pesquisa semelhantes,</p>	<p>O estudo sobre o modelo S-ARCNN para pontuação de similaridade entre documentos pode ser aplicado em repositórios institucionais para melhorar a identificação de documentos semelhantes ou duplicados. Ao implementar o S-ARCNN nesses repositórios, é possível automatizar a detecção de plágio, facilitar a busca por documentos relacionados e aprimorar a organização e recuperação de informações. Isso pode beneficiar instituições acadêmicas, bibliotecas digitais e centros de pesquisa ao aumentar a eficiência na gestão de conteúdo e na garantia da originalidade dos documentos disponibilizados.</p>

				seleção de revisores e correspondência entre currículos e descrições de empregos.	
2022	CBIR-DSS: Business Decision Oriented Content-Based Recommendation Model for E-Commerce	<p>O problema de pesquisa abordado neste artigo é a necessidade de desenvolver um modelo de recomendação baseado em conteúdo orientado para decisões de negócios no contexto do comércio eletrônico. O objetivo é superar as limitações dos sistemas tradicionais de suporte à decisão que não possuem algoritmos de correspondência baseados em índices de pesquisa replicáveis para corresponder e recuperar consultas exatas entregues por meio de entradas na forma de respostas e imagens.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>O objetivo deste estudo é propor um modelo híbrido de recuperação de imagem baseado em conteúdo, integrado a um Sistema de Suporte à Decisão (DSS) e um sistema de recomendação para a indústria têxtil, seja offline ou online. O modelo visa demonstrar como um DSS pode ser integrado a um sistema capaz de segmentar clientes com base em suas características pessoais para personalizar recomendações de produtos usando análises comportamentais. Isso é realizado por meio do treinamento em dados de personalidade MBTI e dados de EEG Deep contendo várias ondas cerebrais pré-treinadas.</p>	<p>Os resultados da pesquisa indicam que o modelo proposto alcançou uma precisão máxima em comparação com outros modelos propostos do estado da arte, devido à sua análise qualitativa. Na primeira parte da análise, foi utilizado um algoritmo de aprendizado profundo para treinar o modelo CBIR com base em diferentes classificadores, como VGG-net, Inception-Net e U-net, alcançando uma precisão de 98,2% com uma taxa de erro minimizada de 2%. Os resultados foram validados usando diferentes métricas de desempenho, como F-score, F-weight, Precisão e Recall. A segunda parte do modelo foi testada em diferentes algoritmos de aprendizado de máquina com uma taxa de precisão de 89,9%. Assim, o modelo completo foi treinado, validado e testado separadamente para obter a máxima eficiência.</p>	<p>O estudo propõe um modelo híbrido de recuperação de imagem baseado em conteúdo, integrado a um Sistema de Suporte à Decisão (DSS) e um sistema de recomendação para a indústria têxtil. Esse modelo pode ser aplicado em repositórios institucionais para melhorar a personalização de recomendações de produtos com base nas características pessoais dos clientes, utilizando análises comportamentais e dados de personalidade. Isso pode resultar em uma experiência de compra mais personalizada e eficaz para os clientes, aumentando a satisfação e fidelidade à marca.</p>
2022	Characteristics of Malay translated hadith corpus	<p>O problema de pesquisa abordado no documento está relacionado à análise e anotação de documentos de hadith traduzidos para o malaio, com foco em questões de processamento de linguagem</p>	<p>O objetivo do estudo sobre as características do corpus de hadith traduzido para o malaio é analisar e anotar os documentos de hadith em malaio, especificamente em termos de marcação de partes do discurso</p>	<p>O resultado da pesquisa sobre as características do corpus de hadith traduzido para o malaio inclui a produção de uma análise e benchmark do corpus em termos de marcação de partes do discurso e entidades</p>	<p>O estudo sobre as características do corpus de hadith traduzido para o malaio pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações e a análise de</p>

		<p>natural e marcação de entidades nomeadas. O estudo visa melhorar a precisão e eficácia da recuperação de informações, bem como contribuir para o desenvolvimento de ferramentas de processamento de linguagem natural para o idioma malaio.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>e entidades nomeadas. O estudo visa criar uma análise e um benchmark do corpus de hadith traduzido para o malaio, seguindo a distribuição da Lei de Zipf. Isso contribui para a pesquisa em processamento de linguagem natural relacionada aos hadiths, melhorando as técnicas de extração e recuperação de informações, além de fornecer uma base para futuras pesquisas no campo.</p>	<p>nomeadas, seguindo a distribuição da Lei de Zipf. Esses resultados fornecem uma base sólida para estudos futuros no campo do processamento de linguagem natural relacionado aos hadiths em malaio. A análise e anotação do corpus contribuem para a melhoria das técnicas de recuperação e extração de informações, bem como para o desenvolvimento de ferramentas de processamento de linguagem natural mais precisas e eficazes para o idioma malaio.</p>	<p>documentos em língua malaia. A análise e anotação do corpus podem melhorar a precisão da indexação e categorização de documentos, facilitando a busca e recuperação de informações relevantes. Além disso, as técnicas desenvolvidas no estudo, como a marcação de partes do discurso e entidades nomeadas, podem ser utilizadas para aprimorar a organização e a acessibilidade de conteúdos em repositórios institucionais, beneficiando pesquisadores, estudantes e demais usuários que buscam informações em malaio.</p>
2022	Clone-Seeker: Effective Code Clone Search Using Annotations	<p>O problema de pesquisa abordado no artigo é a busca eficaz por clones de código, com foco na recuperação de métodos de clones. O objetivo é desenvolver uma abordagem que possa recuperar métodos de clones de forma eficaz com base em consultas de pesquisa, tanto em termos de termos de código-fonte quanto de linguagem natural. A pesquisa visa melhorar a eficácia da busca de clones de código, utilizando estratégias de anotação para construir metadados para os métodos de clones.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de</p>	<p>O objetivo do estudo é propor uma abordagem inovadora para a recuperação eficaz de métodos de clones de código. O foco principal é demonstrar a eficácia do Clone-Seeker na recuperação de métodos de clones, especialmente os clones semânticos (Tipo-4), em comparação com abordagens existentes. Além disso, o estudo visa mostrar como o Clone-Seeker pode buscar clones de código relevantes ao aplicar consultas em linguagem natural.</p>	<p>O estudo apresentou resultados promissores na busca por clones de código, com destaque para a eficácia do Clone-Seeker na recuperação de clones semânticos (Tipo-4) em comparação com abordagens existentes. Além disso, a pesquisa demonstrou a capacidade do Clone-Seeker de buscar com precisão clones de código relevantes ao aplicar consultas em linguagem natural. Esses resultados indicam o potencial do Clone-Seeker como uma ferramenta eficaz para a busca de clones de código, especialmente os clones semânticos, e para a aplicação de consultas em linguagem natural nesse contexto.</p>	<p>O estudo pode ser aplicado em repositórios institucionais de código-fonte de software para melhorar a busca e recuperação de clones de código. A abordagem proposta, Clone-Seeker, utiliza estratégias de anotação para gerar metadados em forma de documentos em linguagem natural para métodos de clones, permitindo uma busca mais eficaz com consultas em termos de código-fonte e linguagem natural. Ao implementar o Clone-Seeker em repositórios institucionais, os desenvolvedores podem identificar e explorar implementações semelhantes de funcionalidades, facilitando a reutilização de código e a</p>

		informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA			manutenção do software. Essa abordagem pode melhorar a eficiência e a precisão da busca por clones de código, contribuindo para o desenvolvimento de software de alta qualidade e aprimorando a produtividade dos desenvolvedores.
2022	Clustering of Relevant Documents Based on Findability Effort in Information Retrieval	<p>O problema de pesquisa abordado neste artigo é a dificuldade que os usuários enfrentam ao tentar encontrar informações relevantes nos resultados de busca recuperados. Os autores exploram a questão da eficácia da recuperação de informações e propõem técnicas de clustering para melhorar a capacidade dos usuários de encontrar informações relevantes de forma mais eficiente.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo do estudo é utilizar as características de "findability" para determinar a quantidade de esforço necessária para encontrar informações em documentos relevantes, por meio de uma abordagem de aprendizado de máquina. Além disso, busca-se demonstrar as alterações no desempenho dos sistemas de recuperação de informações quando o esforço é considerado na avaliação.</p>	<p>Os resultados da pesquisa indicam que os documentos relevantes podem ser agrupados com sucesso utilizando a abordagem de clustering k-means com base nas características de "findability". A inclusão do esforço de findability na avaliação dos sistemas de recuperação de informações resultou em variações de desempenho de até 23%, em média. Isso sugere que considerar o esforço necessário para encontrar informações relevantes pode impactar significativamente a eficácia dos sistemas de recuperação de informações.</p>	<p>O estudo sobre clustering de documentos relevantes com base no esforço de findability pode ser aplicado em repositórios institucionais para melhorar a eficiência da recuperação de informações. Ao considerar o esforço necessário para encontrar informações relevantes, os repositórios podem ser organizados de forma a facilitar a localização de conteúdo importante pelos usuários, resultando em uma experiência de busca mais eficaz e satisfatória.</p>
2022	Comprehensive Framework of Early and Late Fusion for Image–Sentence Retrieval	<p>O problema de pesquisa abordado no artigo é a melhoria da correspondência entre imagens e sentenças em sistemas de recuperação de informações multimodais. O</p>	<p>O objetivo do estudo apresentado no artigo é propor um framework abrangente que combina técnicas de fusão precoce e tardia para a tarefa de correspondência entre imagens</p>	<p>Os resultados da pesquisa demonstraram que o framework abrangente de fusão precoce e tardia (CFELF) proposto apresentou um desempenho superior em comparação com</p>	<p>O estudo sobre o framework abrangente de fusão precoce e tardia para recuperação de imagem-sentença pode ser aplicado em repositórios institucionais para melhorar a</p>

		<p>objetivo é superar as limitações dos métodos tradicionais de fusão tardia e explorar o potencial da fusão precoce para aprimorar a precisão e a eficiência da recuperação de informações entre modalidades diferentes, como visão e linguagem.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>e sentenças. O objetivo principal é melhorar a eficácia da recuperação de informações multimodais, promovendo a comunicação cruzada entre modalidades e capturando informações salientes dentro de cada modalidade, ao mesmo tempo em que se concentra no alinhamento intermodal.</p>	<p>outros módulos de fusão precoce e modelos de fusão tardia avançados. A abordagem CFELF conseguiu melhorar significativamente as taxas de recall em comparação com os modelos de fusão tardia existentes, especialmente em métodos de codificação independente para as modalidades de imagem e texto. Além disso, a fusão precoce no CFELF mostrou melhorias mais significativas nas pontuações de recuperação de imagens em comparação com as pontuações de recuperação de sentenças, indicando a eficácia da abordagem na comunicação cruzada entre modalidades e na aprendizagem intramodal.</p>	<p>correspondência entre imagens e descrições de itens no acervo. Ao implementar a abordagem CFELF, os repositórios podem aprimorar a eficiência e precisão da recuperação de informações multimodais, facilitando a busca e navegação dos usuários por meio de consultas visuais e textuais. Isso pode resultar em uma experiência de usuário aprimorada, com resultados de pesquisa mais relevantes e precisos, promovendo a descoberta e uso eficaz dos recursos disponíveis nos repositórios institucionais.</p>
2022	<p>Conversational Voice Assistants and a Case Study of Long-Term Users: A Human Information Behaviours Perspective</p>	<p>O problema de pesquisa abordado neste estudo é a investigação dos comportamentos de informação humanos de usuários de longo prazo de assistentes de voz conversacionais. O estudo busca compreender como esses usuários interagem com os assistentes de voz, explorando suas experiências, desafios e percepções ao buscar e receber informações por meio dessa tecnologia.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de</p>	<p>O objetivo deste estudo é analisar como os usuários de longo prazo de assistentes de voz conversacionais utilizam essa tecnologia para atender às suas necessidades de informação. Os pesquisadores buscam compreender como a interação com os assistentes de voz afeta os comportamentos de informação dos usuários, incluindo como eles buscam, utilizam e interagem com as informações por meio desses dispositivos. Além disso, o estudo explora como os usuários aprendem a utilizar os assistentes de voz, como as</p>	<p>Os resultados da pesquisa indicam que os usuários de longo prazo de assistentes de voz enfrentam obstáculos relacionados à persona humanizada desses dispositivos e à adaptação à interação por voz. Os participantes do estudo demonstraram uma variedade de comportamentos de informação ao utilizar os assistentes de voz, incluindo a busca de informações, a solicitação de suporte ao cliente e a realização de transações. Além disso, os usuários se adaptaram às interfaces de voz e aprenderam a interagir com os</p>	<p>O estudo sobre os comportamentos de informação de usuários de assistentes de voz pode ser aplicado em repositórios institucionais para melhorar a experiência do usuário ao buscar e acessar informações. Com base nas descobertas do estudo, os repositórios institucionais podem considerar a implementação de interfaces de voz para facilitar a recuperação de informações, adaptando-se às preferências e comportamentos dos usuários. Além disso, a compreensão dos padrões de interação dos usuários com assistentes de voz</p>

		<p>informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>relações com esses dispositivos influenciam suas interações com outras pessoas e como eles se adaptam a novas formas de busca e navegação de informações.</p>	<p>assistentes ao longo do tempo. O estudo destaca a importância de considerar os comportamentos de informação humanos ao projetar e desenvolver assistentes de voz, bem como a necessidade de explorar mais a fundo o impacto dessas tecnologias nas interações sociais e na dinâmica familiar.</p>	<p>pode orientar a personalização de serviços e a criação de experiências mais intuitivas e eficazes nos repositórios institucionais.</p>
2022	Dealing with textual noise for robust and effective BERT re-ranking	<p>O problema de pesquisa abordado no documento é a investigação do impacto da presença de ruído textual na eficácia e robustez dos modelos de reclassificação BERT em tarefas de recuperação de informações. Os autores exploram como perturbações textuais imperceptíveis podem afetar a qualidade da classificação dos modelos BERT em cenários do mundo real, onde o texto ruidoso é comum. Eles buscam compreender se a eficácia relatada dos modelos de reclassificação BERT ainda se mantém em textos ruidosos, dada a natureza contextualizada desses modelos.  <b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Relevance Feedback, expansão da</p>	<p>O objetivo do estudo é avaliar a capacidade dos modelos de reclassificação BERT de lidar com ruído textual dentro de documentos e investigar maneiras de melhorar a robustez e eficácia desses modelos em cenários do mundo real, nos quais o texto ruidoso é prevalente. Os autores realizam experimentos exploratórios no conjunto de dados MS MARCO para examinar se os modelos de reclassificação BERT ainda conseguem ter um bom desempenho ao classificar textos com ruído. Além disso, eles propõem um modelo denominado De-Ranker, que é projetado para lidar de forma mais eficaz com o ruído textual, resultando em melhorias de desempenho em comparação com os modelos BERT tradicionais.</p>	<p>Os resultados da pesquisa indicam que o modelo BERT re-ranker é afetado negativamente pela presença de ruído textual dentro de documentos, com uma degradação significativa na eficácia de classificação ao lidar com diferentes tipos de ruído sintético. No entanto, ao introduzir o modelo De-Ranker, que é treinado para minimizar a distância entre o texto ruidoso e sua versão limpa original, observou-se uma melhoria de desempenho de 3% a 4% em comparação com o BERT re-ranker convencional. Além disso, experimentos de transferência zero em 18 conjuntos de dados de recuperação de informações demonstraram que o De-Ranker não só consegue lidar com ruído textual sintético, mas também melhora em média 1,32% em termos de capacidade de generalização entre domínios. Esses resultados sugerem que o</p>	<p>O estudo sobre a eficácia e robustez dos modelos de reclassificação BERT em lidar com ruído textual pode ter aplicações significativas em repositórios institucionais. Ao considerar a presença comum de ruído textual em documentos acadêmicos e científicos, a implementação de modelos como o De-Ranker pode melhorar a precisão e relevância dos resultados de busca em repositórios institucionais. Essa abordagem pode ajudar a mitigar os efeitos negativos do ruído textual, garantindo uma recuperação de informações mais eficaz e precisa para os usuários que acessam esses repositórios em busca de conteúdo relevante.</p>

		consulta e serviços de recomendação		De-Ranker é mais eficaz na presença de ruído textual, tanto sintético quanto do mundo real, em comparação com os modelos BERT tradicionais.	
2022	Deep learning-based approach for Arabic open domain question answering	<p>O problema de pesquisa abordado neste estudo é a questão de responder a perguntas em árabe em um domínio aberto (OpenQA) utilizando técnicas de aprendizado profundo. O foco está em desenvolver um sistema que possa extrair respostas precisas de grandes volumes de texto não estruturado em árabe, como a Wikipedia, por meio de métodos avançados de recuperação de documentos e compreensão de leitura.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é melhorar o desempenho do sistema de OpenQA em árabe por meio da implementação de uma arquitetura de duas etapas (Retriever-Reader) utilizando técnicas de aprendizado profundo. Os pesquisadores buscam aprimorar a recuperação de informações e a extração de respostas em árabe, comparando o desempenho do sistema em diferentes conjuntos de dados de referência. A meta é demonstrar como o uso de modelos de aprendizado profundo, como o Dense Passage Retrieval (DPR) e o AraELECTRA, pode aprimorar significativamente a capacidade de responder a perguntas em árabe em um contexto de domínio aberto.</p>	<p>O estudo resultou na implementação de um sistema de OpenQA em árabe que utiliza o Dense Passage Retrieval (DPR) para a tarefa de recuperação de passagens e o AraELECTRA para a tarefa de compreensão de leitura. Os resultados mostraram que o DPR superou abordagens tradicionais, como o Term Frequency-Inverse Document Frequency (TF-IDF), em termos de precisão na recuperação das 20 melhores passagens. Além disso, o sistema de OpenQA desenvolvido melhorou significativamente a capacidade de responder a perguntas em árabe em dois conjuntos de dados de referência. A pesquisa demonstrou que a combinação de técnicas de aprendizado profundo para recuperação de informações e compreensão de leitura pode levar a melhorias substanciais no desempenho de sistemas de OpenQA em árabe.</p>	<p>O estudo sobre o uso de técnicas de aprendizado profundo para aprimorar sistemas de OpenQA em árabe pode ser aplicado em repositórios institucionais para melhorar a eficiência na recuperação de informações e na resposta a perguntas em árabe. Ao implementar uma arquitetura de duas etapas com um modelo de recuperação de passagens denso e um modelo de compreensão de leitura, os repositórios institucionais podem oferecer respostas mais precisas e relevantes aos usuários que buscam informações em árabe. Isso pode melhorar a experiência do usuário, aumentar a eficácia da busca e facilitar o acesso a conhecimentos específicos em árabe dentro do contexto institucional.</p>
2022	Deep Neural Network and Pseudo Relevance Feedback Based Query	<p>O problema de pesquisa abordado no estudo é o desajuste de palavras (word mismatch problem) em sistemas de recuperação de informações,</p>	<p>O objetivo do estudo é minimizar o problema de desajuste de palavras (word mismatch problem) ao aplicar uma combinação de feedback de</p>	<p>Os resultados da pesquisa indicam que o modelo proposto, que combina feedback de relevância pseudo e um método baseado em redes neurais</p>	<p>O estudo propõe um modelo que combina feedback de relevância pseudo e redes neurais profundas para expandir consultas em sistemas de</p>

		<p>onde a correspondência exata entre os termos da consulta do usuário e os termos nos documentos relevantes pode ser insuficiente, levando à recuperação de documentos irrelevantes juntamente com os relevantes.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>relevância pseudo e um método baseado em redes neurais profundas para expandir consultas em sistemas de recuperação de informações.</p>	<p>profundas para expandir consultas, apresentou um desempenho superior em comparação com outros modelos existentes. Houve uma melhoria significativa na média de precisão (MAP) em relação à consulta original, ao modelo BM25 e à expansão de consulta com o modelo Chi-Square. Além disso, o modelo proposto conseguiu recuperar um número maior de documentos relevantes em comparação com a consulta original e os outros modelos, melhorando assim o valor de recall.</p>	<p>recuperação de informações. Essa abordagem pode ser aplicada em repositórios institucionais para melhorar a precisão da recuperação de documentos relevantes em resposta às consultas dos usuários. Ao expandir as consultas com termos relevantes identificados a partir dos documentos inicialmente recuperados, o modelo pode ajudar a reduzir o desajuste de palavras e aprimorar a correspondência entre as consultas dos usuários e os documentos recuperados, melhorando assim a eficácia e a relevância das buscas realizadas nos repositórios institucionais.</p>
2022	Detecting geospatial location descriptions in natural language text	<p>O problema de pesquisa abordado no artigo é a detecção automática de descrições de localização geoespaciais relativas em textos em linguagem natural, diferenciando-as de descrições de localização não geográficas. O desafio principal é identificar e distinguir essas expressões geoespaciais relativas de descrições de localização não geográficas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade</p>	<p>O objetivo do estudo é desenvolver métodos automatizados para detectar expressões de localização geoespaciais relativas em textos em linguagem natural e distingui-las de descrições de localização não geográficas. A detecção dessas expressões é crucial para facilitar o processo de georreferenciamento de recursos textuais, permitindo uma identificação mais eficiente e precisa de informações geográficas contidas em textos diversos, como mídias sociais e páginas da web.</p>	<p>O estudo implementou métodos automatizados para detectar expressões de localização geoespaciais em textos em linguagem natural. Os pesquisadores testaram diversas abordagens, incluindo modelos de aprendizado de máquina baseados em características como saco de palavras, embeddings de palavras e padrões de linguagem sintática. Após avaliar os métodos em três conjuntos de dados de teste, observaram que abordagens de ensemble e meta-classificadores proporcionaram a melhor</p>	<p>O estudo sobre a detecção de expressões de localização geoespaciais em textos em linguagem natural pode ser aplicado em repositórios institucionais para melhorar a indexação e recuperação de informações geográficas contidas em documentos acadêmicos e científicos. Ao implementar os métodos automatizados desenvolvidos no estudo, os repositórios institucionais podem aprimorar a identificação e categorização de conteúdo relacionado a locais geográficos, facilitando a busca e a análise de informações</p>

		percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA		medida F de 0,90 para a detecção de expressões de localização geoespaciais. Isso evidencia a eficácia dos métodos propostos, contribuindo para avanços na área de processamento de linguagem natural e geoinformática.	especiais por pesquisadores, estudantes e profissionais que utilizam esses repositórios para acessar conhecimento geoespacial relevante.
2022	DynamicRetriever: A Pre-training Model-based IR System with Neither Sparse nor Dense Index	<p>O problema de pesquisa é aprimorar a eficiência e a precisão dos sistemas de recuperação de informações (IR) ao substituir os índices tradicionais, como índices esparsos ou densos, por um modelo pré-treinado. A pesquisa visa explorar como um sistema baseado em modelo pode capturar informações sem a necessidade de índices explícitos, melhorando assim a qualidade da recuperação de documentos e a relevância dos resultados para os usuários.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo do estudo apresentado no artigo é propor e investigar um novo paradigma de recuperação de informações que não dependa de índices esparsos ou densos, mas sim de um modelo pré-treinado. O principal objetivo é desenvolver o DynamicRetriever, um sistema de IR baseado em modelo que incorpora informações em nível de token e em nível de documento em seus parâmetros, permitindo a geração direta de identificadores de documentos para consultas específicas. A pesquisa visa demonstrar a eficácia e o potencial desse novo paradigma para a recuperação de informações, especialmente em termos de qualidade de classificação e eficiência de busca.</p>	<p>Os resultados da pesquisa demonstraram que o DynamicRetriever superou significativamente os baselines, como o modelo BM25 e o modelo de duas torres baseado em BERT, em todos os subconjuntos de dados avaliados. O modelo D-OverDense obteve os melhores resultados, superando todos os baselines com significância estatística. O DynamicRetriever foi capaz de capturar informações em nível de token e em nível de documento, como autoridade e popularidade, para cada documento, mantendo um índice dinâmico de todos os documentos que pode ser atualizado durante o treinamento do modelo. Esses resultados confirmam a eficácia e escalabilidade do DynamicRetriever em comparação com abordagens tradicionais de recuperação de informações.</p>	<p>O estudo sobre o DynamicRetriever, um sistema de recuperação de informações baseado em modelo, pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de documentos. Ao incorporar informações em nível de token e em nível de documento em seus parâmetros, o DynamicRetriever pode melhorar a precisão e eficiência da busca, permitindo uma recuperação mais relevante e personalizada para os usuários dos repositórios institucionais. Além disso, a capacidade do DynamicRetriever de manter um índice dinâmico de todos os documentos pode facilitar a atualização e expansão contínua dos repositórios, garantindo que os usuários tenham acesso a informações atualizadas e pertinentes.</p>
2022	Effective Natural Language Processing	O problema de pesquisa abordado neste estudo é a	O objetivo deste estudo é propor um método de estruturação de	Os resultados da pesquisa indicam que o método proposto	O estudo sobre a estruturação de relatórios de tumores

	<p>and Interpretable Machine Learning for Structuring CT Liver-Tumor Reports</p>	<p>necessidade de estruturar relatórios de tumores hepáticos em tomografias computadorizadas (CT) de forma eficaz e interpretável. Anteriormente, os resultados de reconhecimento de tumores eram relatados manualmente em um formato não estruturado, o que dificultava a rápida compreensão e extração de padrões. O objetivo é desenvolver um método de estruturação em duas etapas, utilizando Processamento de Linguagem Natural (NLP) e Aprendizado de Máquina (ML), para identificar padrões e facilitar a geração de relatórios estruturados e de fácil interpretação. <b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>relatórios de tumores hepáticos em tomografias computadorizadas (CT) que integre efetivamente o Processamento de Linguagem Natural (NLP) e o Aprendizado de Máquina (ML). O método visa extrair conhecimento valioso dos relatórios não estruturados de forma eficiente e interpretável, a fim de facilitar a avaliação de riscos de doenças, o reconhecimento de doenças e a recomendação de tratamentos. O foco está em melhorar a eficácia e a eficiência da organização e interpretação dos relatórios de tumores hepáticos para benefício da prática médica.</p>	<p>para estruturar relatórios de tumores hepáticos em tomografias computadorizadas (CT) utilizando Processamento de Linguagem Natural (NLP) e Aprendizado de Máquina (ML) é ligeiramente mais eficaz e muito mais eficiente do que as modernas redes neurais baseadas em Bert. A avaliação experimental demonstrou a eficácia, eficiência e interpretabilidade do método, mostrando que ele é promissor em termos de estruturação de relatórios de forma mais organizada e compreensível. Além disso, o método gera uma estrutura de dados hierárquica que inclui características comuns e refinadas em diferentes níveis, contribuindo para a análise e interpretação aprimoradas dos relatórios.</p>	<p>hepáticos em tomografias computadorizadas (CT) utilizando Processamento de Linguagem Natural (NLP) e Aprendizado de Máquina (ML) pode ser aplicado em repositórios institucionais de saúde para melhorar a organização e interpretação de relatórios médicos não estruturados. Ao integrar técnicas avançadas de NLP e ML, é possível extrair conhecimento valioso de relatórios clínicos, facilitando a avaliação de riscos de doenças, o reconhecimento de doenças e a recomendação de tratamentos. Essa abordagem pode aprimorar a eficácia e eficiência da análise de dados médicos, contribuindo para uma prática clínica mais informada e assertiva.</p>
2022	<p>English–Vietnamese cross-language paraphrase identification using hybrid feature classes</p>	<p>O problema de pesquisa abordado neste estudo é a identificação de paráfrases entre sentenças em inglês e vietnamita, utilizando classes de recursos híbridos. O objetivo é desenvolver um método eficaz para detectar a similaridade entre sentenças em diferentes idiomas, visando aprimorar a precisão e a qualidade da identificação de paráfrases em pares de idiomas cruzados.</p>	<p>O objetivo deste estudo é propor um método para identificar casos de paráfrases entre sentenças em inglês e vietnamita, utilizando classes de recursos híbridos. O foco é melhorar a precisão na identificação de paráfrases em pares de idiomas cruzados, visando aplicações em processamento de linguagem natural, como tradução automática, recuperação de informações bilíngues e</p>	<p>O resultado da pesquisa foi a criação de um modelo de identificação de paráfrases entre sentenças em inglês e vietnamita, utilizando classes de recursos híbridos. O modelo combinou métodos fuzzy e modelos siameses recorrentes para calcular as classes de recursos e obter um resultado final por meio de uma fórmula matemática. Os resultados experimentais demonstraram que o modelo alcançou uma</p>	<p>O estudo sobre a identificação de paráfrases entre sentenças em inglês e vietnamita usando classes de recursos híbridos pode ser aplicado em repositórios institucionais para melhorar a detecção de similaridade entre textos multilíngues. Ao implementar o modelo proposto, os repositórios institucionais podem aprimorar a busca por informações, a detecção de plágio e a recuperação de documentos em</p>

		<p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>detecção de plágio. O estudo busca contribuir para o avanço na identificação de paráfrases em contextos multilíngues, com ênfase na combinação de métodos fuzzy e modelos siameses recorrentes para alcançar resultados mais precisos e eficazes.</p>	<p>precisão de 87,4% de F-measure. Isso indica que o método proposto foi eficaz na identificação de paráfrases em pares de idiomas cruzados, contribuindo para aprimorar a detecção de similaridade semântica entre sentenças em inglês e vietnamita.</p>	<p>diferentes idiomas. Isso contribui para uma melhor organização e acesso aos conteúdos disponíveis nos repositórios, facilitando a pesquisa e a análise de documentos em contextos multilíngues.</p>
2022	<p>Enhancing pretrained language models with structured commonsense knowledge for textual inference</p>	<p>O problema de pesquisa abordado no documento é a melhoria do desempenho em tarefas de inferência textual através da integração de conhecimento estruturado de senso comum em modelos de linguagem pré-treinados. Os pesquisadores propõem o uso do framework Mix-Sci para incorporar informações relacionais em modelos de linguagem pré-treinados, visando aprimorar a compreensão e inferência de texto em diversas aplicações de processamento de linguagem natural.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>O objetivo do estudo apresentado no documento é desenvolver e avaliar o framework Mix-Sci, que combina estratégias de recuperação e geração de informações estruturadas de conhecimento de senso comum para enriquecer modelos de linguagem pré-treinados. O principal objetivo é melhorar o desempenho desses modelos em tarefas de inferência textual, como previsão de eventos e compreensão de leitura de máquinas, integrando efetivamente conhecimento estruturado de senso comum em forma de grafos de eventos e grafos de conhecimento.</p>	<p>Os resultados da pesquisa indicam que o framework Mix-Sci, que integra dois tipos de conhecimento estruturado (Event Graph e Knowledge Graph) em modelos de linguagem pré-treinados, demonstrou melhorias significativas no desempenho em tarefas de inferência textual, como previsão de eventos e compreensão de leitura de máquinas. Comparado a métodos de referência, o Mix-Sci conseguiu aprimorar consistentemente o desempenho dos modelos de linguagem pré-treinados, mostrando a eficácia da integração de conhecimento estruturado de senso comum nessas tarefas.</p>	<p>O estudo sobre a integração de conhecimento estruturado de senso comum em modelos de linguagem pré-treinados, como proposto no framework Mix-Sci, pode ser aplicado em repositórios institucionais para aprimorar a compreensão e inferência de texto em documentos acadêmicos. Ao integrar informações de grafos de eventos e grafos de conhecimento, os repositórios institucionais podem utilizar essa abordagem para melhorar a recuperação de informações, a análise de texto e a geração de insights a partir dos conteúdos armazenados. Isso pode resultar em uma melhor organização, busca e compreensão de documentos acadêmicos, beneficiando pesquisadores, estudantes e profissionais que utilizam esses repositórios para acesso a informações relevantes.</p>

2022	Event-related microblog retrieval in Turkish	<p>O problema de pesquisa abordado é a busca por microblogs relacionados a eventos específicos em turco. O estudo investiga os desafios associados à recuperação de microblogs relevantes em meio à natureza dinâmica e à extensão limitada dessas mensagens. O objetivo é encontrar métodos eficazes para representar artigos de notícias como consultas e microblogs como documentos, a fim de identificar microblogs relacionados a eventos em turco e analisar redes sociais online para compreender a opinião pública sobre esses eventos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo é encontrar microblogs relacionados a eventos em turco, utilizando artigos de notícias como consultas e microblogs como documentos. O estudo visa examinar métodos de codificação para representar artigos de notícias e microblogs, incluindo abordagens tradicionais como "bag-of-words" e embeddings de palavras fornecidos por modelos de linguagem pré-treinados, como BERT e FastText. Além disso, o objetivo é medir a similaridade ou relacionamento entre o artigo de notícias codificado e o microblog, classificando os microblogs de acordo com sua relevância para a consulta de entrada. Os resultados experimentais visam comparar a eficácia dos diferentes métodos de codificação e seleção de consulta, bem como avaliar o impacto do pré-processamento de texto em microblogs turcos.</p>	<p>Os resultados do estudo indicam que o modelo baseado em BERT superou outras abordagens de codificação em turco, enquanto o modelo "bag-of-words" com a medida de similaridade de Dice enfrentou desafios ao lidar com textos curtos. Além disso, o uso do título de notícias como consulta mostrou ser mais eficaz do que o texto completo ou o resumo. O pré-processamento de microblogs em turco teve um impacto positivo nos modelos "bag-of-words" e embeddings FastText, enquanto os embeddings BERT mostraram robustez ao ruído em turco, tornando o pré-processamento de texto menos essencial para esses modelos. Esses resultados destacam a importância de modelos de linguagem pré-treinados e técnicas de pré-processamento para a recuperação de microblogs em turco relacionados a eventos.</p>	<p>O estudo sobre recuperação de microblogs relacionados a eventos em turco pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de informações relevantes. Utilizando métodos de codificação como BERT e FastText, juntamente com técnicas de seleção de consulta e pré-processamento de texto, os repositórios institucionais podem melhorar a eficiência na identificação de microblogs relacionados a eventos específicos. Isso pode resultar em uma recuperação mais precisa e relevante de informações em turco, beneficiando a análise de redes sociais, opinião pública e eventos atuais nos repositórios institucionais.</p>
2022	Exploring intrinsic information content models for addressing the issues of traditional semantic measures to evaluate verb similarity	<p>O problema de pesquisa abordado neste trabalho é a avaliação da similaridade semântica de pares de verbos, com foco na superação das limitações dos modelos tradicionais de medidas semânticas. Os autores exploram abordagens baseadas em conteúdo de informação</p>	<p>O objetivo deste trabalho é investigar e propor estratégias baseadas em modelos de conteúdo de informação intrínseco (IIC) para avaliar a similaridade semântica de pares de verbos. Os autores buscam preencher a lacuna na pesquisa relacionada à avaliação da similaridade semântica de</p>	<p>Os resultados da pesquisa incluem a proposição de estratégias combinadas e não lineares para estimar a similaridade semântica de verbos, a análise extensiva de modelos baseados em conteúdo de informação (IC) e modelos baseados em conteúdo de informação intrínseco (IIC) em</p>	<p>O estudo sobre a avaliação da similaridade semântica de pares de verbos, com foco em modelos de conteúdo de informação intrínseco (IIC), pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações e a análise de conteúdo textual. Ao utilizar as estratégias propostas</p>

		<p>intrínseco (IIC) para lidar com questões como a falta de cobertura de hiperônimos e hipônimos em verbos, a inadequação dos modelos existentes para lidar com conjuntos de sinônimos de verbos e a necessidade de estratégias inovadoras para melhorar a correlação com julgamentos humanos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>verbos, melhorando a precisão e eficácia das medidas tradicionais de similaridade semântica. Eles propõem abordagens combinadas e não lineares, bem como estratégias inovadoras para lidar com verbos, visando aprimorar a correlação com julgamentos humanos e aperfeiçoar a compreensão das tarefas de processamento de linguagem natural.</p>	<p>pares de verbos, e a melhoria da precisão das medidas de similaridade semântica existentes. Além disso, o estudo aborda questões como a manipulação de pares de verbos sem caminho de similaridade, a geração de conjuntos de sinônimos com base em partes do discurso e a proposição de estratégias não lineares utilizando informações normalizadas de caminho, profundidade e conteúdo de informação. Os autores também realizam comparações entre diferentes modelos de incorporação de palavras e avaliam o desempenho das estratégias propostas em conjuntos de dados de pares de verbos.</p>	<p>no estudo, os repositórios podem melhorar a precisão na busca por informações relacionadas a verbos, facilitando a identificação de relações semânticas entre termos e aprimorando a organização e recuperação de dados. Isso pode beneficiar a indexação, a categorização e a busca de documentos em repositórios institucionais, tornando a recuperação de informações mais eficiente e precisa.</p>
2022	<p>“FabNER”: information extraction from manufacturing process Science domain literature using named entity recognition</p>	<p>O problema de pesquisa abordado neste trabalho é a extração de informações do domínio da ciência dos processos de fabricação a partir de textos usando Reconhecimento de Entidades Nomeadas (NER), com foco na categorização de palavras-chave relacionadas à ciência dos processos de fabricação em textos científicos. A pesquisa visa desenvolver um modelo de NER para identificar palavras-chave específicas dentro de frases em resumos científicos e mapeá-las para categorias</p>	<p>O objetivo deste trabalho é desenvolver um modelo de Reconhecimento de Entidades Nomeadas (NER) para a extração de informações em larga escala da literatura científica sobre processos de fabricação. Esse modelo de NER é projetado para analisar mais de 500 mil resumos científicos relevantes e classificar entidades reconhecíveis em 12 categorias específicas relacionadas à fabricação. Além disso, o objetivo é demonstrar como esse modelo de NER treinado</p>	<p>O resultado da pesquisa inclui o desenvolvimento de um modelo de NER para a extração de informações da literatura científica sobre processos de fabricação. O modelo foi treinado em mais de 1200 resumos científicos anotados manualmente e alcançou uma precisão geral de 88%. O modelo foi capaz de identificar entidades nomeadas em várias categorias relacionadas à fabricação, como material, processo, equipamento, aplicação, propriedades mecânicas, entre outras. Além</p>	<p>O estudo sobre Reconhecimento de Entidades Nomeadas (NER) para a extração de informações da literatura científica sobre processos de fabricação pode ser aplicado em repositórios institucionais para melhorar a indexação e recuperação de informações. Ao implementar o modelo de NER desenvolvido, os repositórios podem automatizar a categorização de documentos relacionados à fabricação, facilitando a busca e organização de conteúdo. Isso pode aumentar a eficiência na descoberta de conhecimento e</p>

		<p>específicas de tópicos de fabricação.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>pode ser utilizado para extrair tópicos de um determinado parágrafo de texto sobre processos de fabricação e agrupar palavras semelhantes com base em um termo de consulta de interesse.</p>	<p>disso, o estudo demonstrou a aplicação prática do modelo em dois casos de uso específicos: resumo de literatura e identificação de tópicos de artigos. Os resultados indicam que o modelo de NER desenvolvido é eficaz na extração de informações relevantes do texto científico sobre processos de fabricação.</p>	<p>na análise de tendências dentro do domínio da fabricação, beneficiando pesquisadores, profissionais e estudantes que utilizam esses repositórios para acessar informações relevantes.</p>
2022	<p>Geographic Named Entity Recognition by Employing Natural Language Processing and an Improved BERT Model</p>	<p>O problema de pesquisa abordado no artigo é a identificação e reconhecimento de entidades nomeadas geográficas em textos em chinês. O estudo propõe um método híbrido de rede neural para o reconhecimento de nomes de lugares chineses, visando superar desafios como a representação de características em nível de palavra, extração de recursos semânticos contextuais e geração de sequências de rótulos ótimas. O objetivo é melhorar o desempenho do reconhecimento de toponímia em diferentes conjuntos de dados públicos, contribuindo para estudos de reconhecimento de entidades nomeadas em diversas áreas geográficas.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>O objetivo do estudo é desenvolver e avaliar um modelo aprimorado de Reconhecimento de Entidades Nomeadas (NER) para o domínio geográfico em chinês, utilizando o modelo de pré-treinamento ALBERT e a arquitetura BiLSTM-CRF. O foco é melhorar a precisão na identificação de entidades geográficas em textos, contribuindo para avanços no processamento de linguagem natural e geoinformática.</p>	<p>Os resultados da pesquisa indicam que o modelo proposto ALBERT-BiLSTM-CRF para o reconhecimento de entidades nomeadas geográficas em chinês obteve um desempenho significativamente superior em comparação com outros modelos de aprendizado profundo. O método demonstrou bom desempenho em todos os índices de avaliação, destacando a importância da diversidade de dados no treinamento do modelo e mostrando melhorias significativas em conjuntos de dados de teste específicos. Esses resultados evidenciam a eficácia do modelo no reconhecimento de entidades geográficas em textos em chinês, com potencial para aplicações em diversas áreas da geoinformática.</p>	<p>O estudo sobre Reconhecimento de Entidades Nomeadas Geográficas utilizando Processamento de Linguagem Natural e um Modelo BERT Aprimorado pode ser aplicado em repositórios institucionais para aprimorar a extração e organização de informações geográficas em textos. Ao implementar o modelo ALBERT-BiLSTM-CRF proposto, os repositórios podem automatizar a identificação de nomes de lugares em documentos, facilitando a indexação e recuperação de informações geográficas relevantes. Isso pode melhorar a precisão na busca por conteúdo geográfico em repositórios institucionais, tornando a descoberta de dados espaciais mais eficiente e precisa.</p>

2022	Global and local interaction matching model for knowledge-grounded response selection in retrieval-based chatbots	<p>O problema de pesquisa abordado neste estudo é a seleção de respostas fundamentadas no conhecimento em sistemas de diálogo, com foco na interação entre o contexto da conversa, o conhecimento disponível e a resposta adequada. O objetivo é melhorar a capacidade dos chatbots de selecionar respostas relevantes e coerentes, considerando tanto informações globais (interação entre seqüências de contexto e conhecimento) quanto locais (interação entre pares de enunciados).</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é propor um modelo de interação global e local para a seleção de respostas fundamentadas no conhecimento em chatbots de recuperação, visando melhorar a correspondência entre uma resposta e o par contexto-conhecimento. O modelo GLIMM busca integrar informações de diferentes perspectivas (global e local) para fornecer dados complementares e aprimorar o processo de seleção de respostas em sistemas de diálogo.</p>	<p>Os resultados da pesquisa demonstraram que o modelo proposto, GLIMM, superou os baselines do estado-da-arte em termos de desempenho em dois conjuntos de dados públicos. O GLIMM mostrou melhorias significativas em relação a modelos como RSM-DCK e FIRE, mesmo sem a seleção prévia de informações relevantes de contexto/conhecimento. Além disso, o GLIMM superou modelos como DGMN e FIRE, indicando a utilidade dos processos de interação global e local. A análise de ablação também destacou a importância tanto da interação global quanto da local para a seleção de respostas fundamentadas no conhecimento.</p>	<p>O estudo sobre o modelo de interação global e local para seleção de respostas fundamentadas no conhecimento em chatbots pode ser aplicado em repositórios institucionais para melhorar a correspondência entre as consultas dos usuários e as respostas fornecidas pelo sistema. Ao integrar informações de contexto e conhecimento de forma mais eficaz, o modelo GLIMM pode aprimorar a recuperação de informações relevantes e precisas em repositórios institucionais, resultando em interações mais significativas e satisfatórias para os usuários.</p>
2022	Improving word embeddings in Portuguese: increasing accuracy while reducing the size of the corpus	<p>O problema de pesquisa abordado neste estudo é a melhoria dos embeddings de palavras em português, com foco em aumentar a precisão enquanto reduz o tamanho do corpus utilizado. A pesquisa visa aprimorar a descrição de conteúdo multimídia adicionando tags semanticamente relacionadas, utilizando uma técnica de</p>	<p>O objetivo do estudo é aprimorar os embeddings de palavras em português, visando aumentar a precisão dessas representações semânticas enquanto reduz o tamanho do corpus utilizado para treinamento. Além disso, o estudo busca melhorar a descrição de conteúdo multimídia adicionando tags semanticamente relacionadas,</p>	<p>Os resultados da pesquisa indicam um desempenho superior na identificação de tags relacionadas em comparação com soluções semelhantes desenvolvidas não apenas para o português, mas também para o inglês e outras línguas. Além disso, o estudo demonstra que o tamanho do corpus pode ser significativamente reduzido sem afetar o desempenho do modelo,</p>	<p>O estudo sobre a melhoria dos embeddings de palavras em português pode ser aplicado em repositórios institucionais para aprimorar a descrição de conteúdo multimídia adicionando tags semanticamente relacionadas. Isso pode facilitar a recuperação de informações mais precisas e relevantes para os usuários que buscam conteúdo específico nos</p>

		embedding de palavras baseada na rede neural Word2Vec. <b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Informação	utilizando a técnica de Word2Vec.	sem a necessidade de criar conjuntos de dados específicos do domínio.	repositórios. Além disso, a redução do tamanho do corpus necessário para treinamento do modelo pode tornar a implementação mais viável em ambientes com recursos limitados, mantendo um alto nível de desempenho na identificação de tags relacionadas.
2022	Information is essential for competitive and cost-effective public procurement	O problema de pesquisa abordado neste estudo é a análise dos efeitos econômicos da informação nas práticas de contratação pública dentro da União Europeia. Os pesquisadores investigaram como o nível de informação presente nos avisos de contratação pública afeta a competição, os preços dos contratos e a eficácia dos processos de contratação pública. Eles buscaram quantificar o conteúdo de informação dos avisos de contratação pública da UE e examinar como esse conteúdo impacta os resultados da contratação pública. Além disso, o estudo procurou identificar potenciais economias que os governos da UE poderiam obter se todos os avisos de contratação pública contivessem informações detalhadas o suficiente. <b>Categoria:</b> Recursos de informação	O objetivo deste estudo foi analisar os efeitos econômicos da informação na contratação pública dentro da União Europeia. Os pesquisadores buscaram quantificar o conteúdo de informação presente nos avisos de contratação pública da UE, avaliar o impacto desse conteúdo nos resultados da contratação pública, como competição, preços dos contratos e eficácia dos processos, e identificar potenciais economias que os governos da UE poderiam obter se todos os avisos de contratação pública contivessem informações detalhadas.	Os resultados da pesquisa revelaram que a inclusão de mais informações nos avisos de contratação pública tem efeitos positivos significativos. Aumentar o conteúdo informativo está associado a um aumento médio de 20% no número de empresas que participam de licitações públicas da UE. Além disso, empresas estrangeiras têm mais chances de ganhar contratos em avisos mais detalhados, e a probabilidade de manipulação diminui à medida que o conteúdo informativo aumenta. Os preços dos contratos também diminuem, em média, entre 6% e 8% quando os avisos são mais informativos. Em termos de economia, os governos da UE poderiam economizar até 80 bilhões de euros se todos os avisos de contratação pública fossem suficientemente detalhados. Esses resultados ressaltam a importância de regular e monitorar o conteúdo	O estudo sobre os efeitos econômicos da informação na contratação pública pode ser aplicado em repositórios institucionais para promover a transparência e eficiência nos processos de contratação. Ao disponibilizar os dados e códigos utilizados no estudo, os repositórios podem facilitar a reprodução dos resultados e permitir que pesquisadores e formuladores de políticas conduzam análises adicionais. Além disso, a análise baseada em texto dos avisos de contratação pública pode ser utilizada para melhorar os processos de contratação pública, identificar padrões de competição e manipulação, e otimizar os resultados econômicos. Dessa forma, os repositórios institucionais podem se tornar fontes valiosas de informações para aprimorar as práticas de contratação pública e promover a eficiência no uso dos recursos públicos.

		<b>Subcategoria:</b> Informação		informativo nos avisos de contratação pública para promover a competição e a eficácia nos processos de contratação.	
2022	Information Retrieval Systems: Between Morphological Analyzers and Stemming Algorithms	<p>O problema de pesquisa abordado no documento é a comparação do impacto da análise morfológica com o dos algoritmos de stemming em sistemas de recuperação de informações (IRS) para a língua árabe. A questão central é investigar qual abordagem resulta em melhor desempenho na recuperação de informações, especialmente em relação à precisão, recall e MAP (Mean Average Precision) para um determinado número de documentos retornados em consultas curtas e longas.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Representação do documento</p>	<p>O objetivo do estudo é contribuir para a melhoria do desempenho de sistemas de recuperação de informações (IRS) em língua árabe, integrando a análise morfológica em comparação com os algoritmos de stemming. O foco principal é avaliar como a análise morfológica pode impactar a precisão, recall e MAP do IRS, especialmente em termos da qualidade dos resultados retornados para consultas de diferentes comprimentos.</p>	<p>Os resultados da pesquisa indicam que a integração da análise morfológica para a língua árabe em um IRS de código aberto resultou em melhorias significativas no desempenho em termos de precisão/recall, MAP e precisão para um número específico de documentos retornados. Em comparação com o uso de algoritmos de stemming, a análise morfológica mostrou um aumento na precisão para consultas curtas e longas. Além disso, houve uma melhoria na MAP ao utilizar a análise morfológica.</p>	<p>O estudo sobre a integração da análise morfológica em sistemas de recuperação de informações (IRS) para a língua árabe pode ser aplicado em repositórios institucionais para melhorar a precisão e eficiência na recuperação de informações. Ao implementar técnicas de análise morfológica, os repositórios podem aprimorar a indexação e a correspondência de consultas, resultando em uma seleção mais precisa e relevante de documentos para os usuários. Isso pode melhorar a experiência de busca e recuperação de informações em repositórios institucionais, tornando o acesso aos recursos mais eficiente e eficaz.</p>
2022	Intelligent Information Retrieval for Reducing Missed Cancer and Improving the Healthcare System	<p>O problema de pesquisa abordado neste estudo é a necessidade de reduzir os casos de câncer não detectados e melhorar o sistema de saúde, especialmente no contexto do câncer de mama. A pesquisa visa desenvolver um sistema de recuperação de informações inteligente que possa extrair</p>	<p>O objetivo deste estudo é desenvolver um sistema de recuperação de informações inteligente que possa fornecer suporte em tempo real aos radiologistas para prever com precisão a malignidade, reduzindo assim o número de falsos negativos e minimizando erros humanos no processo de</p>	<p>O método proposto neste estudo alcançou um score F1 de 0.981 para classificar os relatórios de mamografia em classes benignas e malignas. Tanto a precisão quanto o recall foram superiores aos do sistema baseado em regras existente. O desempenho do modelo proposto foi comparado com o</p>	<p>O estudo propõe um sistema de recuperação de informações inteligente para auxiliar na detecção de câncer de mama, integrando dados estruturados e semi-estruturados. Esse sistema pode ser aplicado em repositórios institucionais de saúde para automatizar a extração de informações</p>

		<p>dados úteis de conjuntos de dados de câncer de mama e utilizá-los para construir um modelo de classificação eficaz. O objetivo é fornecer suporte em tempo real aos radiologistas para prever com precisão a malignidade, reduzindo o número de falsos negativos e minimizando erros humanos no processo de análise radiológica e patológica.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>análise radiológica e patológica. O sistema proposto visa integrar o suporte à decisão no processo de análise dos radiologistas e patologistas, sem a necessidade de processos equivalentes de entrada de dados.</p>	<p>de outros grupos de pesquisa e superou os sistemas baseados em regras de ponta. O modelo proposto conseguiu extrair informações mais abrangentes dos relatórios do que o sistema baseado em regras descrito anteriormente por Bozkurt et al. Além disso, o modelo reduziu a taxa de falsos positivos e aumentou o valor de precisão. No entanto, o estudo reconhece algumas limitações, como o viés potencial de desempenho de classificação devido ao uso de conjuntos de dados institucionais únicos e o tamanho limitado do conjunto de dados de relatórios de mamografia, que pode levar ao ajuste e resultados enganosos.</p>	<p>relevantes de relatórios de mamografia e FNA, auxiliando os profissionais de saúde na tomada de decisões mais precisas e rápidas. A integração de técnicas de processamento de linguagem natural e aprendizado de máquina permite a análise eficiente de grandes conjuntos de dados, melhorando a precisão do diagnóstico e potencialmente reduzindo erros humanos.</p>
2022	<p>KEST: A graph-based keyphrase extraction technique for tweets summarization using Markov Decision Process</p>	<p>O problema de pesquisa abordado no artigo é a identificação de frases-chave específicas de interesse em tweets para resumir informações relevantes sobre acontecimentos atuais. O estudo destaca a necessidade de desenvolver técnicas eficazes de extração de frases-chave em tweets, considerando a natureza dinâmica e em constante evolução dos dados de mídias sociais, como o Twitter.</p> <p><b>Categoria:</b> Representação do problema do usuário</p>	<p>O principal objetivo do estudo apresentado no artigo é propor uma abordagem eficaz e simples para a extração de frases-chave específicas de interesse em tweets, visando a sumarização de informações relevantes sobre acontecimentos atuais. A técnica KEST utiliza um processo de Decisão de Markov para gerar cadeias dinâmicas de palavras interconectadas a partir de redes de concorrência de palavras, com o intuito de melhorar a pontuação ROUGE-2 ao conectar palavras vizinhas. Além disso, o estudo visa</p>	<p>Os resultados da pesquisa apresentada no artigo demonstraram que a técnica KEST superou as técnicas existentes de sumarização de multi-tweets em termos de precisão, recall e medida F para a avaliação de desempenho de correspondência exata de strings de frases-chave. A análise comparativa com outras técnicas mostrou que a informatividade das sumarizações resultantes com o KEST foi superior. Além disso, a pesquisa destacou a importância de considerar tópicos menos</p>	<p>O estudo sobre a técnica KEST para extração de frases-chave em tweets pode ser aplicado em repositórios institucionais para melhorar a sumarização de informações relevantes em textos curtos, como resumos de artigos, notícias e publicações acadêmicas. Ao utilizar o processo de Decisão de Markov e redes de concorrência de palavras, é possível aprimorar a identificação de frases-chave específicas de interesse, facilitando a busca e recuperação de informações importantes em repositórios</p>

		(necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA	disponibilizar dois conjuntos de dados de tweets (UTweet10 e BTweet10) publicamente para uso futuro pela comunidade de pesquisa.	discutidos em redes de concorrência de palavras para melhorar a qualidade das frases-chave extraídas.	institucionais. Essa abordagem pode contribuir para uma melhor organização e acesso aos conteúdos disponíveis, tornando a pesquisa mais eficiente e eficaz para os usuários.
2022	Knowledge Graph and Deep Learning-based Text-to-GraphQL Model for Intelligent Medical Consultation Chatbot	O problema de pesquisa abordado neste estudo é a tradução de consultas em linguagem natural para consultas em formato de grafo (Graph + QL) em um contexto de sistemas de consulta médica inteligente. A pesquisa visa desenvolver um modelo eficaz que possa transformar perguntas em linguagem natural em consultas compreensíveis por bancos de dados em formato de grafo, permitindo uma comunicação mais eficiente e precisa em sistemas de consulta médica.  <b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Representação do documento	O objetivo deste estudo é propor um modelo baseado em Knowledge Graph e Deep Learning para realizar a tradução de consultas em linguagem natural para consultas em formato de grafo (Graph + QL) em um contexto de sistemas de consulta médica inteligente. O modelo visa melhorar a comunicação entre humanos e máquinas em consultas médicas, permitindo uma interação mais eficiente e precisa, aproveitando a estrutura flexível e expansível dos bancos de dados em formato de grafo.	Os resultados da pesquisa demonstraram a eficácia do modelo proposto para a tradução de consultas em linguagem natural para consultas em formato de grafo (Graph + QL) no contexto de sistemas de consulta médica inteligente. O desempenho do modelo foi avaliado em conjuntos de dados como ATIS e GeoQuery, mostrando uma melhoria significativa em comparação com outros métodos existentes. A abordagem proposta foi capaz de alcançar uma precisão de cerca de 80% em múltiplos conjuntos de dados de domínio único, demonstrando sua utilidade e eficácia na tradução de consultas médicas complexas em consultas compreensíveis por bancos de dados em formato de grafo.	O estudo sobre o modelo de Text-to-GraphQL baseado em Knowledge Graph e Deep Learning para sistemas de consulta médica inteligente pode ser aplicado em repositórios institucionais para melhorar a eficiência e precisão das consultas realizadas. Ao integrar a capacidade de traduzir consultas em linguagem natural para consultas em formato de grafo, os repositórios institucionais podem oferecer uma experiência de busca mais avançada e interativa para os usuários, permitindo uma recuperação de informações mais precisa e abrangente. Isso pode resultar em uma melhor organização e acesso aos dados armazenados nos repositórios, facilitando a pesquisa e a tomada de decisões com base em informações mais claras e estruturadas.
2022	Latent Space Semantic Supervision Based on Knowledge Distillation for Cross-Modal Retrieval	O problema de pesquisa abordado neste trabalho é a melhoria da correspondência fina entre espaços latentes de imagem e texto para recuperação cruzada de	O objetivo deste estudo é propor um modelo de supervisão semântica em espaços latentes baseado em destilação de conhecimento (L3S-KD) para recuperação cruzada de	O resultado da pesquisa foi a proposta do modelo L3S-KD, que demonstrou consistentemente um desempenho superior em comparação com os métodos de	O estudo sobre Latent Space Semantic Supervision Based on Knowledge Distillation for Cross-Modal Retrieval propõe um modelo inovador, L3S-KD, que melhora a correspondência

		<p>modalidades. O objetivo é superar as limitações dos métodos existentes que se concentram principalmente na correspondência global, resultando em uma representação onde conceitos secundários de fundo são frequentemente ignorados. A proposta visa aprimorar a interpretabilidade e a precisão da correspondência entre imagem e texto, considerando a interação fina entre visão e linguagem.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>imagem-texto. O modelo visa supervisionar a correspondência fina entre os espaços latentes de imagem e texto, melhorando a precisão da inferência de relações intra-modais e do alinhamento de informações cruzadas. Através da introdução de classificadores de objetos e atributos nos espaços latentes de imagem e texto, e da transferência de conhecimento semântico por meio da destilação de conhecimento, o objetivo é alinhar melhor as características de regiões de imagem com as características de contexto de texto, resultando em uma recuperação mais precisa e eficaz de pares imagem-texto.</p>	<p>última geração para correspondência imagem-texto. Através da supervisão semântica em espaços latentes e da destilação de conhecimento, o L3S-KD foi capaz de aprender similaridades semânticas mais precisas para fragmentos locais em pares imagem-texto. Os experimentos extensivos realizados nos conjuntos de dados MS-COCO e Flickr30K validaram a eficácia do modelo proposto, destacando sua capacidade de superar os métodos existentes em tarefas de correspondência imagem-texto.</p>	<p>entre espaços latentes de imagem e texto para recuperação cruzada mais precisa. Esse modelo pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de informações multimodais, como imagens e textos, permitindo uma melhor organização e acesso aos conteúdos disponíveis. Ao utilizar a supervisão semântica em espaços latentes e a destilação de conhecimento, o L3S-KD pode melhorar a precisão e relevância dos resultados de pesquisa, facilitando a descoberta de informações relevantes de forma mais eficiente e eficaz.</p>
2022	<p>Legal Information Retrieval systems: State-of-the-art and open issues</p>	<p>O problema de pesquisa abordado no documento está focado no estado atual e nas questões em aberto relacionadas aos sistemas de recuperação de informações legais. Os autores discutem as abordagens de ponta, desafios e questões não resolvidas no campo da recuperação de informações legais, incluindo a classificação de documentos legais, a extração de informações relevantes de documentos legais, a</p>	<p>O objetivo do documento fornece uma visão abrangente do estado atual da recuperação de informações legais, destacando as tecnologias e abordagens mais recentes utilizadas nesse campo. Além disso, o documento visa identificar e discutir as questões em aberto e os desafios enfrentados na recuperação de informações legais, com o intuito de orientar futuras pesquisas e desenvolvimentos nessa área. Os autores buscam fornecer</p>	<p>O resultado da pesquisa inclui uma análise aprofundada do estado atual dos sistemas de recuperação de informações legais, destacando as tecnologias e abordagens mais recentes utilizadas nesse campo. Além disso, são discutidos os desafios e questões em aberto relacionados à recuperação de informações legais, como a classificação de documentos legais, a extração de informações relevantes, a</p>	<p>O estudo pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações legais armazenadas nessas plataformas. A utilização de tecnologias como inteligência artificial, processamento de linguagem natural e ontologias pode melhorar a eficiência na busca e extração de informações relevantes em documentos legais. Isso pode beneficiar instituições ao facilitar o acesso a dados jurídicos</p>

		<p>compreensão semântica, a robustez dos sistemas de recuperação de informações legais e a busca de documentos legais relevantes para profissionais do direito.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>insights valiosos para profissionais do direito, pesquisadores e desenvolvedores interessados em aprimorar os sistemas de recuperação de informações legais.</p>	<p>compreensão semântica e a busca de documentos relevantes para profissionais do direito. Os autores também fornecem insights sobre as tendências emergentes, como o uso de inteligência artificial, processamento de linguagem natural e ontologias para aprimorar a recuperação de informações legais. O documento destaca a importância de desenvolver sistemas mais robustos e sensíveis ao contexto para atender às necessidades dos profissionais do direito e aborda a relevância contínua da pesquisa nesse campo em constante evolução.</p>	<p>importantes, agilizar a pesquisa e análise de casos, e garantir uma recuperação mais precisa e contextualizada de informações legais para profissionais do direito que utilizam os repositórios institucionais.</p>
2022	Leveraging statistical information in fine-grained financial sentiment analysis	<p>O problema de pesquisa abordado neste artigo é a análise de sentimentos financeiros em detalhes finos, especialmente a falta de um conjunto de dados rotulado em grande escala e a natureza dependente do domínio, que dificultam a aplicação de métodos de aprendizado de máquina tradicionais. A pesquisa visa desenvolver um modelo de regressão leve que incorpora informações estatísticas sobre a distribuição de termos em uma faixa de polaridade para melhorar a análise de sentimentos</p>	<p>O objetivo deste estudo é propor um método eficaz que aproveite informações estatísticas em nível de corpus para a tarefa de análise de sentimentos financeiros em detalhes finos. O foco é dividir a faixa contínua de pontuação de sentimento em intervalos e incorporar a ocorrência de cada termo nesses intervalos como uma característica estatística para melhorar o desempenho da regressão. O estudo visa validar a viabilidade e eficácia desses métodos por meio de experimentos extensivos com modelos de regressão baseados em CNN e RNN.</p>	<p>Os resultados da pesquisa demonstraram que a incorporação de informações estatísticas em nível de corpus na análise de sentimentos financeiros em detalhes finos resultou em melhorias significativas no desempenho do modelo de regressão. Os experimentos realizados com modelos baseados em CNN e RNN validaram a eficácia do método proposto. A relação entre os termos e as etiquetas de sentimento mostrou-se informativa, e a inclusão do conhecimento estatístico foi considerada flexível e fácil de aplicar. Além disso, a</p>	<p>O estudo sobre a utilização de informações estatísticas na análise de sentimentos financeiros em detalhes finos pode ser aplicado em repositórios institucionais para aprimorar a compreensão e previsão de orientações sentimentais em textos financeiros. Ao incorporar dados estatísticos em modelos de regressão, as instituições podem melhorar a precisão na avaliação de sentimentos em documentos financeiros, auxiliando na tomada de decisões de investimento e monitoramento de retornos anormais. Essa abordagem</p>

		financeiros em um nível mais refinado. <b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Representação do documento		importância das estatísticas dos sinais na análise de sentimentos financeiros foi reconhecida como benéfica para o desempenho geral do modelo.	pode ser útil para analisar e interpretar grandes volumes de informações financeiras de forma mais eficaz e eficiente.
2022	MedGraph: A semantic biomedical information retrieval framework using knowledge graph embedding for PubMed	O problema de pesquisa abordado neste artigo é a busca por uma solução eficaz para a recuperação de informações semânticas em bibliotecas digitais biomédicas, especificamente no contexto do PubMed. O objetivo é desenvolver e avaliar o framework MedGraph, baseado em incorporação de grafos de conhecimento, para melhorar a relevância semântica na recuperação e classificação de literatura biomédica indexada no PubMed. <b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Representação do documento	O objetivo deste estudo é introduzir e avaliar o framework MedGraph, que utiliza a incorporação de grafos de conhecimento para fornecer recuperação e classificação de relevância semântica para a literatura biomédica indexada no PubMed. O estudo visa comparar o desempenho do MedGraph com um algoritmo tradicional baseado em TF-IDF, avaliar a eficácia do MedGraph usando consultas predefinidas com resultados classificados como verdadeiros e explorar abordagens semânticas para a busca e relevância em bibliotecas digitais biomédicas.	Os resultados da pesquisa indicam que o framework MedGraph apresenta um desempenho comparável ao algoritmo BestMatch do PubMed e supera o método tradicional baseado em TF-IDF na recuperação de informações biomédicas. A avaliação do MedGraph utilizando métricas diversas demonstrou sua eficácia na classificação de relevância semântica. Além disso, o estudo fornece evidências de que abordagens semânticas baseadas em grafos de conhecimento oferecem resultados de busca mais relevantes em comparação com métodos tradicionais, conforme avaliado por métricas objetivas.	O estudo sobre o framework MedGraph, que utiliza incorporação de grafos de conhecimento para recuperação semântica de informações biomédicas, pode ser aplicado em repositórios institucionais para aprimorar a busca e classificação de conteúdo acadêmico. Ao adotar abordagens semânticas baseadas em grafos de conhecimento, os repositórios institucionais podem melhorar a relevância e eficiência na recuperação de informações, beneficiando pesquisadores, estudantes e profissionais que utilizam esses repositórios para acessar conhecimento científico.
2022	MS-Transformer: Introduce multiple structural priors into a unified transformer for encoding sentences	O problema de pesquisa abordado no estudo é a limitação dos modelos Transformer convencionais em capturar informações estruturais entre palavras em uma sentença de forma precisa. O objetivo é integrar múltiplos tipos de informações estruturais em um único modelo Transformer,	O objetivo da pesquisa é introduzir múltiplos tipos de informações estruturais em um modelo Transformer unificado, chamado MS-Transformer, para aprimorar a representação de sentenças e melhorar o desempenho em tarefas de processamento de linguagem natural.	O estudo demonstrou que o MS-Transformer, que incorpora múltiplos tipos de informações estruturais em um único modelo Transformer, obteve melhorias significativas em relação a outras abordagens em três tarefas de processamento de linguagem natural: inferência de linguagem natural, identificação	O estudo sobre o MS-Transformer, que introduz múltiplos tipos de informações estruturais em um modelo Transformer unificado, pode ser aplicado em repositórios institucionais para melhorar a representação e a compreensão de textos. Ao incorporar essas informações estruturais, o

		<p>chamado MS-Transformer, para melhorar a representação de sentenças e a capacidade do modelo de capturar dependências entre as palavras de maneira mais eficaz.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>		<p>de paráfrases e classificação de sentimentos. Isso sugere que a introdução de múltiplos tipos de informações estruturais pode aprimorar a capacidade do modelo de capturar dependências entre as palavras e melhorar a representação de sentenças.</p>	<p>modelo pode ajudar na organização e recuperação de informações, bem como na análise de documentos acadêmicos e científicos. Isso pode resultar em uma melhor indexação, busca e recomendação de conteúdo nos repositórios institucionais, facilitando o acesso e a utilização eficaz dos recursos disponíveis.</p>
2022	Multi-task framework based on feature separation and reconstruction for cross-modal retrieval	<p>O problema de pesquisa abordado neste artigo é a melhoria do desempenho da recuperação cruzada de imagens e legendas (cross-modal image-caption retrieval) por meio da introdução de um novo framework chamado feature separation and reconstruction (mFSR). O objetivo é lidar com a assimetria de informações entre diferentes modalidades (imagem e texto) e melhorar a qualidade da recuperação cruzada de informações visuais e textuais.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>O objetivo deste trabalho é propor e desenvolver um framework de múltiplas tarefas baseado em separação e reconstrução de características (mFSR) para a recuperação cruzada de informações visuais e textuais. O foco principal é melhorar o desempenho da recuperação cruzada de imagens e legendas, abordando a assimetria de informações entre diferentes modalidades e utilizando técnicas de separação e reconstrução de características para aprimorar a representação e a correspondência entre dados visuais e textuais.</p>	<p>Os resultados da pesquisa demonstram que o framework mFSR, que combina a separação e reconstrução de características para a recuperação cruzada de imagens e legendas, apresenta melhorias significativas no desempenho em comparação com métodos de linha de base em diversos conjuntos de dados, incluindo MS-COCO e Flickr30K. A introdução da separação de características e da reconstrução de informações específicas mostrou-se eficaz na melhoria da recuperação cruzada de informações visuais e textuais, resultando em um avanço promissor na área de recuperação cruzada de multimídia.</p>	<p>O estudo sobre o framework mFSR, baseado em separação e reconstrução de características para recuperação cruzada de informações visuais e textuais, pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de conteúdo multimídia. Ao utilizar técnicas avançadas de separação e reconstrução de características, é possível melhorar a correspondência entre imagens e legendas, facilitando a indexação e recuperação de informações em repositórios que contenham uma variedade de tipos de dados, como imagens, textos e vídeos. Essa abordagem inovadora pode aprimorar a experiência de busca e navegação em repositórios institucionais,</p>

					tornando mais eficiente a localização de conteúdo multimodal relevante.
2022	Natural Language Semantic Answering Applied to Medicinal Plant and Coronavirus	<p>O problema de pesquisa abordado no documento é a criação e avaliação de um sistema de resposta a perguntas em linguagem natural, chamado PryRe, que utiliza propriedades semânticas e análise de ontologias para fornecer respostas precisas sobre plantas medicinais e coronavírus. O objetivo é desenvolver uma abordagem eficaz para interpretar perguntas em espanhol e fornecer respostas relevantes com base em um conhecimento estruturado e dedutivo.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo do estudo é apresentar e avaliar o sistema de resposta a perguntas em linguagem natural PryRe, que integra processamento de linguagem natural, ontologias e métodos de recuperação de informações. O sistema foi desenvolvido para interpretar perguntas em espanhol e encontrar respostas com base em uma ontologia do domínio de interesse. O foco principal é fornecer respostas precisas e não triviais para perguntas relacionadas a plantas medicinais e coronavírus, com uma precisão aceitável de 82%.</p>	<p>O resultado da pesquisa foi a criação e avaliação do sistema PryRe, que demonstrou uma precisão de 82% na resposta a perguntas em linguagem natural sobre plantas medicinais e coronavírus. O sistema foi capaz de interpretar perguntas em espanhol, transformá-las em triplas semânticas e realizar análises em uma ontologia para identificar as respostas corretas. Além disso, PryRe mostrou ser uma ferramenta eficaz para responder a perguntas em linguagem natural com base em um conhecimento estruturado e dedutivo.</p>	<p>O estudo sobre o sistema PryRe, que integra processamento de linguagem natural e ontologias para responder a perguntas em linguagem natural, pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações. Ao utilizar a abordagem de PryRe, os repositórios podem oferecer respostas mais precisas e relevantes às consultas dos usuários, facilitando a busca por informações específicas dentro do contexto institucional. Essa aplicação pode aprimorar a eficiência e a usabilidade dos repositórios, tornando a busca por informações mais intuitiva e eficaz.</p>
2022	Natural Language Processing Applied to Forensics Information Extraction With Transformers and Graph Visualization	<p>O problema de pesquisa abordado no artigo é como aplicar técnicas de Processamento de Linguagem Natural (PLN), como reconhecimento de entidades nomeadas e extração de relações, para a extração de informações em investigações</p>	<p>O objetivo do estudo é demonstrar como as técnicas de Processamento de Linguagem Natural, especialmente o reconhecimento de entidades nomeadas e a extração de relações, podem ser aplicadas com sucesso na extração de informações em investigações</p>	<p>Os resultados da pesquisa demonstraram que a combinação de modelos de reconhecimento de entidades nomeadas (NER) e extração de relações (RE), aprimorados por arquiteturas de redes neurais e transformadores, pode melhorar a eficiência da extração de</p>	<p>O estudo demonstra como técnicas avançadas de Processamento de Linguagem Natural, como reconhecimento de entidades nomeadas e extração de relações, podem ser aplicadas na extração de informações em investigações forenses digitais. Essas</p>

		<p>forenses digitais. Os autores buscam desenvolver modelos de PLN aprimorados por arquiteturas de redes neurais e transformadores, visando automatizar a análise de extração de informações em diferentes contextos, incluindo forense digital, investigação digital, análise de big data e inteligência empresarial.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>forenses digitais. Os autores buscam desenvolver modelos aprimorados por arquiteturas de redes neurais e transformadores, com o intuito de automatizar a análise de extração de informações em diversos cenários, incluindo forense digital, investigação digital, análise de big data e inteligência empresarial. A pesquisa visa melhorar a eficiência e a precisão na análise de grandes volumes de dados heterogêneos, com foco na segurança cibernética e na luta contra crimes digitais.</p>	<p>informações em investigações forenses digitais. Os modelos propostos apresentaram melhorias significativas na precisão e no desempenho em comparação com abordagens tradicionais. Por meio da construção de um sistema de recuperação de informações baseado em conhecimento, os pesquisadores conseguiram automatizar a análise de extração de informações, gerando grafos interativos que facilitam a visualização e compreensão dos dados extraídos. Esses resultados indicam o potencial das técnicas de PLN e transformadores para aprimorar a análise de dados em contextos forenses e de investigação digital, bem como em outras aplicações de análise de dados em larga escala e inteligência empresarial.</p>	<p>técnicas, combinadas com arquiteturas de transformadores, podem ser utilizadas em repositórios institucionais para automatizar a análise de dados, melhorar a precisão na extração de informações e facilitar a visualização de dados por meio de grafos interativos. Isso pode otimizar a gestão de informações, acelerar processos de análise e fortalecer a segurança cibernética em ambientes institucionais.</p>
2022	<p>NLP-Based Bi-Directional Recommendation System: Towards Recommending Jobs to Job Seekers and Resumes to Recruiters</p>	<p>O problema de pesquisa abordado no artigo é a melhoria do processo de recomendação de empregos para os candidatos e de currículos para os recrutadores, utilizando um sistema de recomendação bi-direcional baseado em Processamento de Linguagem Natural (NLP). O objetivo é aprimorar a eficiência e a precisão das recomendações feitas entre candidatos em busca de emprego e recrutadores,</p>	<p>O objetivo do estudo é desenvolver e avaliar um sistema de recomendação bi-direcional baseado em NLP para melhorar a correspondência entre candidatos e ofertas de emprego. Especificamente, o sistema visa recomendar empregos relevantes para os candidatos com base em seus perfis e currículos, bem como recomendar perfis de candidatos adequados para os recrutadores com base nas descrições das</p>	<p>O resultado da pesquisa demonstrou que o sistema de recomendação proposto, baseado em NLP, obteve uma precisão geral de 0.8. Isso significa que o sistema foi capaz de fazer recomendações corretas em 80% dos casos avaliados, melhorando significativamente a correspondência entre candidatos e empregos, bem como entre currículos e recrutadores. Além disso, o</p>	<p>O estudo sobre o sistema de recomendação bi-direcional baseado em NLP pode ser aplicado em repositórios institucionais para melhorar a correspondência entre os currículos dos candidatos e as vagas de emprego disponíveis. Ao implementar esse sistema em repositórios institucionais, é possível automatizar e aprimorar o processo de seleção de candidatos, tornando as recomendações mais precisas e</p>

		<p>facilitando o processo de correspondência entre habilidades e requisitos de emprego.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>	<p>vagas de emprego. O foco está em aprimorar a eficácia do processo de recrutamento e seleção, proporcionando benefícios tanto para os candidatos quanto para os recrutadores.</p>	<p>estudo mostrou que o sistema foi capaz de melhorar a precisão das recomendações em diferentes áreas, como Engenheiro de Dados, Desenvolvedor Java e Desenvolvedor Back-end, com valores de recall de 0.4, 0.8 e 0.8, respectivamente. Esses resultados confirmam a eficácia do sistema de recomendação bi-direcional proposto no contexto de recrutamento e seleção de pessoal.</p>	<p>eficientes. Isso pode beneficiar tanto os candidatos, ao receberem sugestões de empregos mais alinhadas com seus perfis, quanto os recrutadores, ao identificarem candidatos mais adequados para as posições em aberto.</p>
2022	On cross-lingual retrieval with multilingual text encoders	<p>O problema de pesquisa abordado neste estudo é a avaliação do desempenho de codificadores de texto multilíngues de última geração em tarefas de recuperação de documentos e frases em um contexto de recuperação cruzada entre idiomas. O estudo investiga como esses codificadores pré-treinados lidam com a recuperação de informações em diferentes idiomas e a eficácia de sua aplicação em tarefas de recuperação de informações multilíngues.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>O objetivo deste estudo é realizar uma análise sistemática do desempenho dos codificadores de texto multilíngues de última geração em tarefas de recuperação de documentos e frases em vários pares de idiomas diversos. O estudo visa preencher lacunas na compreensão da adequação desses modelos para tarefas de recuperação de informações multilíngues, especialmente em comparação com modelos anteriores baseados em espaços de incorporação de palavras multilíngues. Além disso, o estudo explora o impacto do ajuste fino supervisionado e contrastivo na qualidade da classificação em tarefas de recuperação cruzada de informações.</p>	<p>Os codificadores multilíngues são eficazes na recuperação de frases em nível de sentença, mas não superam modelos anteriores para a recuperação de documentos em nível de documento sem supervisão. O ajuste fino supervisionado geralmente não melhora o desempenho dos transformadores multilíngues como classificadores não supervisionados. Apenas o ajuste fino contrastivo no mesmo domínio, mas com transferência de idioma, melhora a qualidade da classificação. Esses resultados evidenciam diferenças entre recuperação cruzada e transferência de recuperação de informações para recuperação monolíngue em idiomas-alvo.</p>	<p>O estudo destaca a eficácia de codificadores de texto multilíngues em tarefas de recuperação de informações em diferentes idiomas. Esses modelos podem ser aplicados em repositórios institucionais para melhorar a recuperação de documentos em vários idiomas, permitindo uma busca mais eficaz e abrangente. Além disso, a técnica de ajuste fino contrastivo no mesmo domínio, com transferência de idioma, pode ser utilizada para melhorar a qualidade da classificação em repositórios institucionais multilíngues, garantindo resultados mais precisos e relevantes para os usuários.</p>

2022	<p>On the Effectiveness of Pre-Trained Language Models for Legal Natural Language Processing: An Empirical Study</p>	<p>O problema de pesquisa abordado no estudo é a avaliação da eficácia de modelos de linguagem pré-treinados em tarefas de processamento de linguagem natural no contexto jurídico. O estudo investiga como esses modelos se comparam a métodos tradicionais e como podem ser aplicados para melhorar a recuperação de informações legais e outras tarefas relacionadas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>O objetivo do estudo é analisar e comparar o desempenho de modelos de linguagem pré-treinados em tarefas específicas de processamento de linguagem natural no domínio jurídico. Além disso, o estudo busca identificar as vantagens e limitações desses modelos em comparação com abordagens tradicionais, visando contribuir para o avanço da pesquisa em NLP jurídico e fornecer insights sobre como esses modelos podem ser aplicados de forma eficaz nesse contexto.</p>	<p>Os resultados da pesquisa indicam que os modelos de linguagem pré-treinados demonstraram um desempenho superior em várias tarefas de processamento de linguagem natural no domínio jurídico em comparação com abordagens tradicionais. Além disso, o estudo revelou que modelos de linguagem pré-treinados específicos do domínio podem ter limitações em termos de transferibilidade dentro do mesmo domínio. Esses resultados destacam a eficácia e as potenciais limitações dos modelos de linguagem pré-treinados em aplicações jurídicas e fornecem insights valiosos para pesquisas futuras e desenvolvimentos nesse campo.</p>	<p>O estudo sobre a eficácia de modelos de linguagem pré-treinados em NLP jurídico pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações legais, classificação de documentos, resposta a perguntas e resumos automáticos. A utilização desses modelos pode melhorar a eficiência na análise de documentos legais, facilitando a pesquisa e extração de informações relevantes em ambientes jurídicos.</p>
2022	<p>Optimal Query Expansion Based on Hybrid Group Mean Enhanced Chimp Optimization Using Iterative Deep Learning</p>	<p>O problema de pesquisa abordado neste estudo é a necessidade de aprimorar os sistemas de recuperação de informações para lidar com a vasta quantidade de informações disponíveis na internet. A automação desses sistemas é essencial para ajudar os usuários a filtrar os documentos de acordo com suas necessidades, exibindo apenas os documentos relevantes. Um dos principais</p>	<p>O objetivo deste estudo é desenvolver uma abordagem de expansão de consulta otimizada com base em uma combinação de técnicas, incluindo o aprimoramento do grupo médio híbrido e a otimização do chimpanzé, utilizando aprendizado profundo iterativo. Essa abordagem visa melhorar a precisão e eficácia dos sistemas de recuperação de informações, permitindo aos usuários obter</p>	<p>Os resultados da pesquisa demonstraram que a técnica proposta de expansão de consulta otimizada apresentou melhorias significativas em termos de métricas de avaliação, como o ganho cumulativo descontado normalizado, a precisão média e a classificação média recíproca. Esses resultados indicam que a abordagem híbrida utilizada neste estudo foi eficaz na recuperação de informações</p>	<p>O estudo sobre a expansão de consulta otimizada com base em técnicas híbridas pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações relevantes. Ao utilizar a abordagem proposta, os repositórios podem melhorar a precisão e eficácia das buscas realizadas pelos usuários, garantindo resultados mais relevantes e precisos. Isso pode contribuir significativamente</p>

		<p>desafios enfrentados pelos sistemas de recuperação de informações é a questão das diferenças de vocabulário, que muitas vezes resultam de ambiguidades e problemas contextuais.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>resultados mais relevantes e precisos em suas buscas.</p>	<p>relevantes, superando métodos tradicionais de recuperação de informações sem expansão de consulta. Portanto, a pesquisa evidenciou a eficiência da metodologia proposta na obtenção de respostas apropriadas para a recuperação de informações, com potenciais aplicações em mecanismos de busca e sistemas de recuperação de informações.</p>	<p>para a eficiência na localização de documentos e informações dentro dos repositórios institucionais, facilitando o acesso e a utilização desses recursos de forma mais eficaz e produtiva.</p>
2022	<p>Plain Template Insertion: Korean-Prompt-Based Engineering for Few-Shot Learners</p>	<p>O problema de pesquisa abordado no documento é a escassez de dados em ambientes de aprendizado de poucas amostras (few-shot learning) e a necessidade de desenvolver métodos eficazes para lidar com essa limitação. O estudo destaca a importância de encontrar maneiras de superar a escassez de dados de treinamento, especialmente em contextos onde apenas um pequeno número de exemplos está disponível para treinar modelos de aprendizado de máquina.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade</p>	<p>O objetivo do estudo sobre Korean-Prompt-Based Engineering for Few-Shot Learners é desenvolver e avaliar métodos de aprendizado baseados em prompts para lidar com a escassez de dados em ambientes de poucas amostras. O foco está em melhorar o desempenho de modelos de aprendizado de máquina em situações em que há disponibilidade limitada de dados de treinamento, buscando soluções eficazes para otimizar o aprendizado com poucas amostras.</p>	<p>Os resultados da pesquisa sobre Korean-Prompt-Based Engineering for Few-Shot Learners demonstraram que o método de inserção de modelos de template (PTI) é eficaz e robusto em lidar com a escassez de dados em ambientes de aprendizado de poucas amostras. O PTI foi capaz de alcançar melhorias significativas de desempenho em tarefas de aprendizado de poucas amostras, como KLUE-NLI, STS e TC. Além disso, o estudo mostrou que o PTI é aplicável tanto em ajustes de prompts "hard" quanto "soft" e é capaz de maximizar a eficiência dos dados disponíveis de forma prática.</p>	<p>O estudo sobre Korean-Prompt-Based Engineering for Few-Shot Learners apresenta um método inovador de inserção de modelos de template (PTI) para lidar com a escassez de dados em ambientes de aprendizado de poucas amostras. Esse método pode ser aplicado em repositórios institucionais para otimizar o aprendizado de máquina com poucas amostras, melhorando o desempenho de modelos em tarefas específicas. Ao utilizar o PTI, os repositórios podem maximizar a eficiência dos dados disponíveis, sem a necessidade de aumentar o número de exemplos de treinamento, o que pode ser especialmente útil em contextos</p>

		percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA			com restrições de coleta de dados.
2022	Relation-Aggregated Cross-Graph Correlation Learning for Fine-Grained Image–Text Retrieval	<p>O problema de pesquisa abordado neste estudo é a dificuldade em aprender a correspondência semântica entre diferentes modalidades, como imagens e texto, de forma finamente detalhada para melhorar o desempenho da recuperação de informações entre essas modalidades heterogêneas. A lacuna de heterogeneidade, a grande discrepância intramodal e a fraca dependência intermodal representam desafios significativos para a recuperação de informações entre imagem e texto. O objetivo é desenvolver um modelo inovador, o RACG, que explore as relações intramodais e intermodais para guiar o processo de aprendizado de correspondência de recursos e melhorar a correlação semântica.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>O objetivo do estudo sobre o modelo RACG é desenvolver uma abordagem eficiente de aprendizado de correlação cruzada entre grafos para melhorar a correspondência semântica finamente detalhada entre imagens e texto. O modelo visa explorar e integrar relações intramodais e intermodais para aprimorar a recuperação de informações multimodais, com foco em otimizar a precisão e relevância da correspondência entre diferentes modalidades de dados.</p>	<p>Os experimentos realizados mostraram que o RACG superou significativamente outros métodos de correspondência global e local, bem como métodos de correspondência baseados em grafos. A abordagem do RACG foi capaz de melhorar a precisão da correspondência semântica entre imagem e texto, resultando em um desempenho superior em comparação com métodos concorrentes. Os resultados da pesquisa confirmaram a eficácia do modelo RACG na recuperação de informações entre imagem e texto, destacando sua capacidade de aprender relações intramodais e intermodais de forma integrada para obter uma correspondência semântica mais precisa e detalhada.</p>	<p>O estudo sobre o modelo RACG para recuperação de imagem-texto pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de informações multimodais. Ao integrar relações intramodais e intermodais, o RACG pode melhorar a correspondência semântica entre imagens e texto, facilitando a indexação e recuperação de conteúdo visual e textual de forma mais precisa e detalhada. Isso pode beneficiar repositórios institucionais ao tornar a busca mais eficiente e relevante, permitindo aos usuários encontrar informações específicas com maior facilidade e precisão.</p>
2022	Retrieval-based language model adaptation for handwritten Chinese text recognition	<p>O problema de pesquisa abordado neste estudo é a melhoria do reconhecimento de textos chineses manuscritos por meio da adaptação de um modelo de linguagem baseado</p>	<p>O objetivo deste estudo é apresentar um método inovador de adaptação de modelo de linguagem baseado em recuperação para o reconhecimento offline de textos</p>	<p>Os resultados da pesquisa demonstraram que o método de adaptação de modelo de linguagem baseado em recuperação proposto neste estudo resultou em melhorias no</p>	<p>O estudo sobre a adaptação de modelo de linguagem baseado em recuperação para o reconhecimento de textos chineses manuscritos pode ser aplicado em repositórios</p>

		<p>em recuperação. A questão central é como adaptar dinamicamente um modelo de linguagem para melhorar a precisão do reconhecimento de textos chineses manuscritos, levando em consideração a variabilidade e a complexidade desses textos.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>chineses manuscritos. O principal objetivo é melhorar o desempenho de reconhecimento de textos chineses manuscritos não restritos, utilizando uma estratégia de reconhecimento em duas etapas que combina modelos de linguagem comuns com modelos adaptativos gerados dinamicamente a partir de conteúdos relacionados obtidos da Internet. A meta é demonstrar a eficácia desse método de adaptação de modelo de linguagem na melhoria do desempenho de reconhecimento de textos chineses manuscritos.</p>	<p>desempenho do reconhecimento de textos chineses manuscritos. A combinação de modelos de linguagem comuns com modelos adaptativos gerados dinamicamente a partir de conteúdos relacionados da Internet levou a uma maior precisão no reconhecimento de textos chineses manuscritos não restritos. Os experimentos realizados com dois conjuntos de dados de referência, CASIA-HWDB e ICDAR-2013, mostraram que a abordagem de adaptação de modelo de linguagem baseada em recuperação resultou em melhorias significativas no desempenho de reconhecimento em comparação com métodos existentes.</p>	<p>institucionais para aprimorar a precisão do reconhecimento de textos manuscritos em chinês. Ao utilizar uma abordagem de adaptação de modelo de linguagem que combina modelos comuns com modelos adaptativos gerados dinamicamente a partir de conteúdos relacionados da Internet, é possível melhorar a eficiência e a precisão do reconhecimento de textos manuscritos não restritos. Isso pode beneficiar repositórios institucionais ao facilitar a indexação e recuperação de documentos manuscritos em chinês, tornando o acesso e a pesquisa mais eficazes para os usuários.</p>
2022	Shallow Neural Network and Ontology-Based Novel Semantic Document Indexing for Information Retrieval	<p>O problema de pesquisa abordado no documento é a melhoria do desempenho dos sistemas de recuperação de informações por meio da indexação semântica de documentos. A abordagem SNNOntoSDI proposta combina modelos de redes neurais rasas e ontologias de domínio para identificar conceitos em documentos e superar as limitações dos sistemas tradicionais de recuperação de informações.</p> <p><b>Categoria:</b> Recursos de informação</p>	<p>O objetivo do estudo é projetar e avaliar uma abordagem inovadora de indexação semântica de documentos, denominada SNNOntoSDI, que integra redes neurais rasas, processamento de linguagem natural e ontologias de domínio. A proposta visa melhorar o desempenho dos sistemas de recuperação de informações, superando a lacuna de vocabulário e permitindo a identificação de conceitos relevantes em documentos com base em características estatísticas, semânticas e de</p>	<p>O resultado da pesquisa demonstrou que a abordagem SNNOntoSDI superou os métodos de ponta existentes, com uma melhoria média de 29% na precisão e 25% no F-measure em comparação com esses métodos. A avaliação empírica realizada em cinco conjuntos de dados de referência mostrou que a abordagem proposta obteve resultados superiores, destacando sua eficácia na indexação semântica de documentos para recuperação de informações.</p>	<p>O estudo propõe uma abordagem inovadora de indexação semântica de documentos, o SNNOntoSDI, que combina redes neurais rasas, processamento de linguagem natural e ontologias de domínio para melhorar a recuperação de informações. Essa abordagem pode ser aplicada em repositórios institucionais para aprimorar a indexação e recuperação de documentos, permitindo uma busca mais precisa e eficiente com base em conceitos semânticos relevantes. Ao</p>

		<b>Subcategoria:</b> Documento	entidades nomeadas científicas. A avaliação empírica da abordagem é realizada em cinco conjuntos de dados de referência para demonstrar sua eficácia em comparação com métodos de ponta existentes.		integrar recursos externos, como ontologias de domínio, o SNNOntoSDI pode ajudar a superar desafios de vocabulário e melhorar a organização e recuperação de informações em repositórios institucionais.
2022	Short text matching model with multiway semantic interaction based on multi-granularity semantic embedding	O problema de pesquisa abordado neste artigo é a correspondência de texto curto em chinês, levando em consideração a diferença na granularidade dos caracteres em chinês em comparação com as palavras em inglês. A pesquisa propõe um modelo de correspondência de texto curto com interação semântica multiway baseada em incorporação semântica de multi-granularidade para lidar com essa distinção e melhorar a precisão da correspondência de sentenças curtas em chinês. <b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Representação do documento	O objetivo deste estudo é desenvolver e avaliar um modelo de interação semântica multiway baseado em incorporação semântica de multi-granularidade (MSIM) para melhorar a correspondência de texto curto em chinês. O modelo proposto visa capturar de forma mais eficaz as nuances semânticas presentes em sentenças curtas em chinês, levando em consideração a granularidade dos caracteres chineses e a complexidade da estrutura gramatical e expressão de contexto do idioma. A pesquisa busca demonstrar a eficácia do MSIM em comparação com abordagens existentes para a correspondência de sentenças curtas.	Os resultados da pesquisa demonstram que o modelo MSIM proposto supera outras abordagens existentes na correspondência de texto curto em chinês. A análise dos dados experimentais revela que o MSIM alcança a maior precisão ao utilizar a função de ativação ReLU em comparação com LReLU e PReLU. Além disso, o modelo MSIM se destaca em termos de precisão, perda e medida F1 em comparação com experimentos comparativos anteriores. Esses resultados indicam que a combinação de interação semântica multiway com incorporação semântica de multi-granularidade permite obter informações semânticas de sentenças curtas em chinês com maior precisão, melhorando a correspondência e generalização dessas sentenças.	O estudo sobre o modelo de correspondência de texto curto com interação semântica multiway baseada em incorporação semântica de multi-granularidade (MSIM) pode ser aplicado em repositórios institucionais para melhorar a precisão e eficiência na correspondência de informações curtas em chinês. Ao utilizar o MSIM, os repositórios podem aprimorar a identificação de informações relevantes, como na recuperação de informações, perguntas e respostas, e identificação de paráfrases. Isso pode resultar em uma melhor organização e recuperação de dados em repositórios institucionais, contribuindo para uma melhor experiência do usuário e uma maior eficácia na busca e análise de informações.
2022	STCM-Net: A symmetrical one-stage network for temporal language localization in videos	O problema de pesquisa abordado no artigo "STCM-Net: A symmetrical one-stage network for temporal language localization in videos" é a	O objetivo do estudo apresentado no artigo "STCM-Net: A symmetrical one-stage network for temporal language localization in videos" é	Os resultados da pesquisa apresentada no artigo "STCM-Net: A symmetrical one-stage network for temporal language localization in videos"	O estudo sobre STCM-Net, uma rede neural para localização temporal de linguagem em vídeos, pode ser aplicado em repositórios institucionais para

		<p>localização temporal de linguagem em vídeos. O objetivo é desenvolver uma rede neural capaz de extrair conceitos temporais contidos em frases descritivas de forma a melhorar a precisão na localização de segmentos de vídeo com base nessas informações temporais. O estudo visa resolver os desafios de extração de semântica de frases e integração contextual de vídeos para aprimorar o desempenho de localização em vídeos.</p> <p><b>Categoria:</b> Recursos de informação  <b>Subcategoria:</b> Documento</p>	<p>desenvolver uma rede neural eficaz, denominada STCM-Net, que possa extrair e utilizar de forma precisa os conceitos temporais contidos em frases descritivas para melhorar a localização temporal de segmentos de vídeo. A proposta visa superar os desafios relacionados à extração de semântica de frases e à integração contextual de vídeos, com o intuito de aprimorar significativamente o desempenho em tarefas de localização em vídeos.</p>	<p>demonstraram que a abordagem proposta, STCM-Net, obteve melhorias significativas na precisão de localização em vídeos em comparação com modelos de referência do estado da arte. A extração de conceitos temporais das frases descritivas mostrou ter um efeito positivo na melhoria da precisão de localização. Além disso, a rede neural desenvolvida conseguiu superar os modelos de referência em três conjuntos de dados públicos: ActivityNet Captions, TACoS e Charades-STA. Os resultados indicam que a abordagem proposta é eficaz na extração e utilização de informações temporais para otimizar a localização de segmentos de vídeo com base em descrições de linguagem natural.</p>	<p>aprimorar a indexação e recuperação de vídeos com base em descrições de linguagem natural. A capacidade da STCM-Net de extrair e utilizar conceitos temporais de frases pode melhorar a precisão na localização de segmentos de vídeo, facilitando a busca e navegação em grandes coleções de vídeos. Isso pode beneficiar repositórios institucionais ao tornar o acesso e a organização de vídeos mais eficientes e precisos, proporcionando uma experiência de usuário aprimorada.</p>
2022	<p>Summarization of Text and Image Captioning in Information Retrieval Using Deep Learning Techniques</p>	<p>O problema de pesquisa abordado neste estudo é a melhoria da eficiência e precisão na recuperação de informações e na sumarização de texto e legendas de imagens utilizando técnicas de aprendizado profundo (deep learning). O objetivo é desenvolver um modelo inovador que possa lidar com a crescente quantidade de dados textuais disponíveis na internet e produzir resumos precisos e relevantes de</p>	<p>O objetivo deste estudo é aplicar técnicas avançadas de aprendizado profundo (deep learning) para aprimorar a recuperação de informações, a geração de resumos de texto e a criação de legendas para imagens. O foco está em desenvolver um modelo que possa extrair informações significativas de grandes volumes de texto, gerar resumos precisos e relevantes, e produzir legendas descritivas para</p>	<p>Os resultados da pesquisa indicaram que o modelo proposto, que combina o uso de bidirectional long short term memory (BiLSTM) para recuperação de informações, deep belief network (DBN) para geração de resumos de texto e legendas de imagens, superou os métodos comparativos em termos de precisão, recall e F-score. A avaliação do desempenho do modelo foi realizada utilizando o corpus Gigaword e o corpus DUC,</p>	<p>O estudo propõe um modelo avançado de aprendizado profundo para recuperação de informações e geração de resumos de texto e legendas de imagens. Essa abordagem pode ser aplicada em repositórios institucionais para automatizar a extração de informações relevantes de documentos, facilitando a busca e recuperação de conteúdo. Além disso, a geração de resumos precisos e legendas descritivas para imagens pode melhorar a</p>

		documentos e legendas de imagens. <b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento	imagens de forma automatizada e eficiente.	demonstrando a eficácia e superioridade do modelo proposto em relação às abordagens anteriores.	acessibilidade e compreensão do conteúdo armazenado nos repositórios, tornando a pesquisa mais eficiente e eficaz para os usuários.
2022	Multi-head attention graph convolutional network model: End-to-end entity and relation joint extraction based on multi-head attention graph convolutional network	O problema de pesquisa abordado neste estudo é a extração conjunta de entidades e relações em texto não estruturado. A pesquisa visa desenvolver um modelo inovador baseado em redes neurais convolucionais com atenção multi-cabeça para realizar essa tarefa de forma eficiente e precisa. O objetivo é superar as limitações dos métodos tradicionais de extração de entidades e relações, especialmente em cenários onde as relações entre entidades podem se sobrepor, resultando em desafios adicionais para os sistemas de processamento de linguagem natural. <b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento	O objetivo deste estudo é propor um modelo de rede neural convolucional com atenção multi-cabeça para realizar a extração conjunta de entidades e relações em texto não estruturado. O modelo visa superar as limitações dos métodos tradicionais, especialmente em cenários onde as relações entre entidades são complexas e podem se sobrepor. A pesquisa busca melhorar a precisão e eficiência da extração de entidades e relações, contribuindo para avanços significativos em tarefas de processamento de linguagem natural relacionadas à extração de informações.	Os resultados da pesquisa demonstraram que o modelo proposto, baseado em redes neurais convolucionais com atenção multi-cabeça, superou outros métodos existentes em termos de desempenho na extração conjunta de entidades e relações em texto não estruturado. Os experimentos realizados mostraram que o modelo desenvolvido obteve resultados superiores, e também foi evidenciado que a matriz de adjacência extraída pelo modelo apresentou uma precisão de reconhecimento melhor do que ferramentas externas. Esses resultados destacam a eficácia e a relevância do modelo proposto para a tarefa de extração de entidades e relações de forma integrada e eficiente.	O estudo sobre o modelo de rede neural convolucional com atenção multi-cabeça para extração conjunta de entidades e relações em texto não estruturado pode ser aplicado em repositórios institucionais para aprimorar a extração de informações de documentos acadêmicos e científicos. Ao utilizar esse modelo avançado, os repositórios institucionais podem melhorar a precisão na identificação de entidades e relações relevantes nos textos, facilitando a indexação, recuperação e análise de conteúdo. Isso pode resultar em uma melhor organização e acessibilidade dos dados armazenados nos repositórios, contribuindo para uma pesquisa mais eficiente e eficaz.
2022	The effect of co-opinion on the cocitation-based information retrieval systems' effectiveness evaluated by semantic similarity	O problema de pesquisa abordado neste estudo é a avaliação da eficácia dos sistemas de recuperação de informações baseados em cocitação, considerando o impacto da co-opinião nas relações de cocitação. Os	O objetivo deste estudo foi investigar o impacto da medida de co-opinião na eficácia dos sistemas de recuperação de informações baseados em cocitação. Os pesquisadores buscaram avaliar como a co-opinião entre os documentos	Os resultados da pesquisa demonstraram que a medida de co-opinião, que avalia a similaridade das opiniões entre os documentos citantes em relação aos documentos cocitados, teve um impacto significativo na eficácia dos	O estudo sobre o impacto da co-opinião em sistemas de recuperação de informações baseados em cocitação pode ser aplicado em repositórios institucionais para melhorar a precisão e a relevância dos resultados de busca. Ao

		<p>pesquisadores investigaram como a co-opinião entre os documentos citados pode influenciar a eficácia dos sistemas de recuperação de informações, especialmente em termos de classificação e ranking de relevância. Eles exploraram como a análise de opinião e a similaridade semântica podem ser utilizadas para melhorar a precisão e a qualidade dos resultados de recuperação de informações baseados em cocitação.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>citantes pode influenciar a classificação e o ranking de relevância dos documentos cocitados em sistemas de recuperação de informações. Eles também procuraram comparar a eficácia da co-opinião com outras medidas de similaridade, como a similaridade semântica distribucional e a similaridade de termos do Medical Subject Headings (MeSH), para avaliar a melhoria nos resultados de recuperação de informações.</p>	<p>sistemas de recuperação de informações baseados em cocitação. A co-opinião foi capaz de melhorar a classificação e o ranking de relevância dos documentos cocitados, resultando em uma melhoria na precisão e na qualidade dos resultados de recuperação de informações. Os pesquisadores compararam a eficácia da co-opinião com outras medidas de similaridade, como a similaridade semântica distribucional e a similaridade de termos do Medical Subject Headings (MeSH). Eles observaram que a co-opinião foi uma medida eficaz para melhorar os resultados de recuperação, destacando a importância de considerar as opiniões dos documentos citantes ao avaliar a relevância dos documentos cocitados em sistemas de recuperação de informações baseados em cocitação.</p>	<p>considerar a similaridade das opiniões entre os documentos citantes em relação aos documentos cocitados, os repositórios podem aprimorar a classificação e o ranking dos documentos recuperados, proporcionando aos usuários resultados mais relevantes e alinhados com suas necessidades de informação. Isso pode resultar em uma experiência de busca mais eficaz e satisfatória para os usuários dos repositórios institucionais.</p>
2022	<p>The HoPE Model Architecture: a Novel Approach to Pregnancy Information Retrieval Based on Conversational Agents</p>	<p>O problema de pesquisa abordado no estudo do modelo HoPE (Healthcare Obstetric in PrEgnancy) é a necessidade de desenvolver uma abordagem inovadora para a recuperação de informações relacionadas à gravidez, utilizando agentes de conversação. O objetivo é criar um modelo que seja capaz de fornecer respostas assertivas e</p>	<p>O objetivo do estudo do modelo HoPE é propor uma arquitetura de conversational agent inovadora e eficaz, denominada HoPE (Healthy Obstetrician for Pregnancy), para auxiliar na recuperação de informações relacionadas à gravidez. Esse modelo visa fornecer respostas assertivas e rápidas, utilizando uma combinação de arquiteturas</p>	<p>O resultado da pesquisa sobre o modelo HoPE (Healthcare Obstetric in PrEgnancy) foi a proposição de uma arquitetura inovadora de conversational agent para recuperação de informações relacionadas à gravidez. O modelo HoPE utiliza uma combinação de redes Sentence-BERT ajustadas a diretrizes de gravidez e</p>	<p>O estudo do modelo HoPE pode ser aplicado em repositórios institucionais de saúde para melhorar a recuperação de informações relacionadas à gravidez. A arquitetura de conversational agent HoPE, que combina redes Sentence-BERT e ontologias, pode ser implementada para oferecer suporte personalizado e eficaz</p>

		<p>rápidas, especificamente adaptadas para dados de gravidez, superando as limitações de velocidade de modelos baseados em arquiteturas de transformadores tradicionais. A pesquisa visa avaliar e propor um modelo que possa recuperar informações de forma eficiente e precisa, atendendo às necessidades de saúde pública e informação clínica relacionadas à gravidez.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>híbridas, como o uso de redes Sentence-BERT ajustadas a diretrizes de gravidez e modelagem baseada em ontologias. O principal propósito é oferecer suporte às mulheres grávidas, ajudando-as a obter informações confiáveis durante o período de gravidez e pós-parto, contribuindo assim para a saúde materna e neonatal. Além disso, o estudo busca avaliar a eficácia do modelo HoPE em comparação com outros modelos de recuperação de informações, visando aprimorar a qualidade e a precisão das informações disponibilizadas para as gestantes.</p>	<p>estruturas baseadas em ontologias para oferecer suporte às mulheres grávidas na obtenção de informações confiáveis durante o período de gestação. Os resultados indicaram que o modelo HoPE obteve um desempenho superior em comparação com outras abordagens testadas, alcançando um F1-Score de 0.89. Além disso, o modelo BERTimbau, utilizado no HoPE, demonstrou ser capaz de recuperar informações de forma rápida e precisa, superando outros modelos pré-treinados em língua portuguesa. A avaliação do modelo HoPE incluiu a avaliação da capacidade de prever respostas assertivas em comparação com outros modelos de recuperação de informações.</p>	<p>às mulheres grávidas que buscam informações confiáveis durante a gestação. Ao integrar o modelo HoPE em repositórios institucionais, é possível aprimorar a acessibilidade e a qualidade das informações disponíveis, contribuindo para a promoção da saúde materna e neonatal por meio de respostas assertivas e rápidas.</p>
2022	Towards retrieval-based conversational recommendation	<p>O problema de pesquisa abordado é a melhoria e compreensão dos sistemas de recomendação conversacional, com foco em abordagens baseadas em recuperação de informações. O estudo investiga os desafios enfrentados por esses sistemas, a comparação de desempenho entre abordagens baseadas em geração e recuperação, e a análise qualitativa das expressões de preferência dos</p>	<p>O objetivo do estudo é propor e avaliar uma abordagem baseada em recuperação de informações para sistemas de recomendação conversacional. O estudo visa entender como essa abordagem se compara às abordagens baseadas em geração de linguagem, bem como explorar a forma como os usuários expressam suas preferências durante as interações de recomendação. Além disso, o estudo busca fornecer insights</p>	<p>Os resultados da pesquisa indicam que abordagens baseadas em recuperação de informações podem ser uma alternativa promissora ou complemento às abordagens baseadas em geração de linguagem em sistemas de recomendação conversacional. O estudo comparativo realizado mostrou que a abordagem de recuperação proposta obteve resultados promissores em relação às abordagens de</p>	<p>O estudo pode ser aplicado em repositórios institucionais para aprimorar a interação com os usuários, fornecendo recomendações personalizadas de conteúdo de forma mais eficaz. A abordagem de recuperação de informações proposta pode ser utilizada para melhorar a qualidade das respostas fornecidas aos usuários com base em interações anteriores. Além disso, a combinação de técnicas</p>

		usuários durante as interações de recomendação.  <b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA	sobre o potencial das abordagens híbridas que combinam técnicas de recuperação e geração de linguagem.	geração de linguagem recentes. Além disso, o estudo destacou a importância de considerar a combinação de abordagens de recuperação e geração de linguagem em sistemas híbridos para melhorar a capacidade de resposta e confiabilidade dos sistemas de recomendação conversacional.	de recuperação e geração de linguagem em sistemas híbridos pode aumentar a relevância e a confiabilidade das recomendações feitas aos usuários nos repositórios institucionais. Essa abordagem pode resultar em uma experiência de usuário mais satisfatória e personalizada, contribuindo para a eficácia e eficiência dos sistemas de recomendação em repositórios institucionais.
2022	Unifying Relational Sentence Generation and Retrieval for Medical Image Report Composition	O problema de pesquisa abordado no documento é a geração de relatórios de imagens médicas de forma precisa e diversificada. Os métodos atuais muitas vezes geram as sentenças mais comuns devido ao viés do conjunto de dados para casos individuais, independentemente de as sentenças capturarem adequadamente as entidades-chave e as relações. Isso limita severamente a aplicabilidade e a capacidade de generalização desses métodos na composição de relatórios médicos, onde as sentenças mais críticas estão nas descrições de doenças anormais que são relativamente raras. A entrelaçamento de termos médicos em um relatório, como sintomas associados a uma doença específica, também é um desafio a ser superado.	O objetivo do estudo é propor um framework inovador que unifique a recuperação de modelos e a geração de sentenças para lidar tanto com anormalidades comuns quanto raras, garantindo a coerência semântica entre os termos médicos detectados. Especificamente, o framework visa explorar o co-raciocínio de conhecimento híbrido, incluindo as relações explícitas entre todos os termos médicos anormais para induzir a aprendizagem de atenção visual e a codificação de representação de tópicos para descrições de sintomas orientadas por tópicos. Além disso, o objetivo é introduzir um Gerador Adaptativo que alterna entre a recuperação de modelos e a geração de sentenças de acordo com um codificador de	Os resultados da pesquisa demonstraram a superioridade do framework proposto em termos de avaliação humana e métricas em dois conjuntos de dados de relatórios médicos. O estudo conseguiu superar os trabalhos do estado da arte, alcançando um desempenho atraente sob avaliação humana. A abordagem unificada de recuperação e geração de tópicos mostrou-se eficaz na geração de relatórios médicos mais precisos e profissionais, superando as limitações dos métodos atuais que tendem a gerar as sentenças mais comuns devido ao viés do conjunto de dados. Os resultados experimentais destacaram a capacidade do framework de lidar com anormalidades comuns e raras, garantindo a coerência semântica entre os	O estudo propõe um framework inovador que unifica a recuperação de modelos e a geração de sentenças para melhorar a composição de relatórios médicos. Essa abordagem pode ser aplicada em repositórios institucionais para aprimorar a geração automática de relatórios de imagens médicas, garantindo precisão e diversidade nas descrições de diagnósticos e achados. Ao incorporar o co-raciocínio de conhecimento híbrido e a geração adaptativa de sentenças, o framework pode melhorar a eficiência e a qualidade dos relatórios gerados, beneficiando instituições de saúde e profissionais médicos.

		<p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	tópicos contextuais, visando a produção de relatórios médicos mais precisos e profissionais.	termos médicos detectados e melhorando a composição de relatórios médicos.	
2022	Urdu Wikification and Its Application in Urdu News Recommendation System	<p>O problema de pesquisa abordado neste estudo é a falta de utilização de grafos de conhecimento (KG) e técnicas de vinculação de entidades (EL) para o processamento de linguagem natural (NLP) em idioma Urdu. O objetivo é estabelecer uma base para a vinculação de entidades e o uso de KG para aplicações de NLP em Urdu, desenvolvendo um mecanismo automático para vincular texto em Urdu com KG existentes, além de criar um sistema de recomendação de notícias em Urdu.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	Os objetivos principais deste estudo são estabelecer a base para a vinculação de entidades e o uso de grafos de conhecimento (KG) para o processamento de linguagem natural (NLP) em Urdu, desenvolver um mecanismo automático para vincular texto em Urdu com KG existentes, destacar a importância dos grafos de conhecimento (KG) em aplicações de NLP em Urdu e desenvolver um sistema de recomendação de notícias em Urdu. Esses objetivos visam preencher a lacuna entre a pesquisa existente em NLP em Urdu e as metodologias de ponta em NLP, aproveitando o potencial dos grafos de conhecimento para melhorar as aplicações em Urdu.	Os resultados da pesquisa incluem a criação de um conjunto de dados de vinculação de entidades em Urdu (SKEL) para a wikificação em Urdu por meio do Wikidata, o desenvolvimento de uma ferramenta que extrai entidades do texto em Urdu e as vincula ao grafo de conhecimento do Wikidata, a geração de um conjunto de dados de recomendação de notícias em Urdu (SKRS) e a avaliação da ferramenta proposta tanto intrinsecamente, utilizando o conjunto de dados SKEL, quanto extrinsecamente, aplicando-a ao sistema de recomendação de notícias em Urdu e comparando os resultados com métodos de recomendação de notícias padrão. Os resultados destacam a importância das incorporações de KG e extração de entidades para o desempenho do sistema de recomendação de notícias em Urdu, demonstrando a aplicação bem-sucedida do	O estudo sobre a wikificação em Urdu e sua aplicação em um sistema de recomendação de notícias em Urdu pode ser aplicado em repositórios institucionais para melhorar a organização e recuperação de informações. A vinculação de entidades em texto em Urdu com grafos de conhecimento pode enriquecer metadados e facilitar a busca e navegação em repositórios. Além disso, o sistema de recomendação de notícias em Urdu pode ser adaptado para recomendar conteúdo relevante aos usuários com base em seus interesses e interações, melhorando a experiência de descoberta de informações nos repositórios institucionais.

				conhecimento de grafos em NLP em Urdu.	
2022	Why is a document relevant? Understanding the relevance scores in cross-lingual document retrieval	<p>O problema de pesquisa abordado neste documento é a necessidade de desenvolver um modelo eficaz de recuperação de documentos em diferentes idiomas que seja capaz de calcular a relevância do documento para a consulta com alta precisão e também fornecer insights sobre por que um documento é relevante. O objetivo é superar as limitações dos modelos tradicionais de recuperação de informações que não explicam adequadamente a relevância dos documentos, especialmente em um contexto multilíngue.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>O objetivo principal deste trabalho é desenvolver um modelo de recuperação de documentos multilíngue que possa calcular a relevância do documento para a consulta com alta precisão e, ao mesmo tempo, fornecer insights sobre por que um documento é relevante. Além disso, o estudo visa avaliar o desempenho do modelo em conjuntos de dados com consultas e documentos em diferentes idiomas, analisar a interpretabilidade dos escores de relevância do documento fornecidos pelo modelo e investigar o impacto da função de perda objetiva utilizada durante o treinamento no desempenho dos modelos de recuperação de documentos multilíngues.</p>	<p>Os resultados da pesquisa indicam que o modelo proposto, LM-EMD, apresenta um desempenho semelhante ao dos melhores modelos comparativos em idiomas com recursos abundantes, porém é menos eficaz em idiomas com recursos limitados. Além disso, o modelo fornece insights sobre por que um documento é relevante para a consulta, o que pode auxiliar tanto no desenvolvimento de modelos de recuperação de documentos quanto na utilização desses modelos. A análise comparativa das funções de perda objetiva revelou que treinar os modelos com a função de classificação por pares resulta em um melhor desempenho em comparação com a entropia cruzada.</p>	<p>O estudo propõe um modelo de recuperação de documentos multilíngue, LM-EMD, que pode ser aplicado em repositórios institucionais para melhorar a precisão da busca e fornecer insights sobre a relevância dos documentos em diferentes idiomas. Ao utilizar o LM-EMD, os repositórios institucionais podem aprimorar a experiência de busca dos usuários, especialmente em ambientes multilíngues, permitindo uma compreensão mais profunda de por que um documento é considerado relevante para uma determinada consulta. Isso pode resultar em uma recuperação mais eficaz e eficiente de informações relevantes para os usuários dos repositórios institucionais.</p>
2022	Manuscript version: Author's Accepted Manuscript	<p>O problema de pesquisa abordado no documento é a recuperação de informações multimodais entre imagens e texto, conhecido como o problema de recuperação de</p>	<p>O objetivo do estudo é desenvolver um modelo de rede neural que possa melhorar a recuperação de informações multimodais entre imagens e texto, visando aprimorar a</p>	<p>Os resultados da pesquisa demonstram que o modelo proposto, denominado Stacked Multi-Modal Attention Network (SMAN), é eficaz na recuperação de informações</p>	<p>O estudo sobre o Stacked Multi-Modal Attention Network (SMAN) pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações multimodais, como</p>

		<p>imagem-texto cruzado. O desafio reside na dificuldade de medir as relevâncias semânticas entre instâncias multimodais devido à separação dos dados em espaços de características heterogêneas. O objetivo é melhorar a discriminação dos embeddings latentes para alinhar dados heterogêneos em um espaço semântico comum.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>capacidade de representação de imagens e textos e a consistência semântica entre as modalidades. O modelo proposto utiliza mecanismos de atenção para selecionar componentes significativos de ambas as modalidades e explorar a complementaridade semântica entre os dados heterogêneos, a fim de melhorar a precisão da recuperação de informações cruzadas entre imagens e texto.</p>	<p>entre imagens e texto. O SMAN utiliza um mecanismo de atenção empilhado para processar características contextuais intra-modais e gerar representações discriminativas para medir a similaridade entre dados multimodais. Isso resulta em uma melhoria na capacidade de capturar associações complexas entre imagens e texto, levando a um desempenho aprimorado na recuperação de informações multimodais.</p>	<p>imagens e texto, em documentos acadêmicos. Ao implementar mecanismos de atenção empilhados para processar e relacionar dados heterogêneos, os repositórios podem melhorar a precisão na busca e recuperação de conteúdo multimodal, facilitando a descoberta de informações relevantes e aprimorando a experiência do usuário ao acessar e explorar os recursos disponíveis no repositório.</p>
2022	<p>Building a Technology Recommender System Using Web Crawling and Natural Language Processing Technology</p>	<p>O problema de pesquisa abordado neste artigo é a dificuldade enfrentada por especialistas em determinado domínio em acompanhar as tecnologias recentes de forma eficiente e eficaz. Além disso, uma vez que a informação sobre uma tecnologia é encontrada, é desafiador e demorado compreendê-la e avaliá-la de forma aproximada para decidir se merece uma investigação mais detalhada. Portanto, o objetivo é investigar a possibilidade de automatizar a tarefa de acompanhar as tecnologias recentes, como</p>	<p>O objetivo deste projeto de pesquisa é explorar se a combinação de web crawling, web scrapping e sumarização automática de texto com tecnologia de processamento de linguagem natural pode resultar em um protótipo adequado para fornecer recomendações sobre tecnologias em tendência para um especialista em determinado propósito de aplicação. O estudo visa demonstrar como essas tecnologias podem ser utilizadas em conjunto para fornecer uma solução adequada para extração automática de informações e descoberta de conhecimento no</p>	<p>O resultado da pesquisa indicou que, embora tenha sido possível desenvolver um programa para auxiliar na utilização de mecanismos de busca com a capacidade adicional de gerar resumos extrativos curtos dos resultados da pesquisa, o programa não alcançou resultados satisfatórios em termos de qualidade das recomendações. Isso se deve principalmente à extração relativamente aleatória de texto das fontes de conteúdo, resultando em resumos com frases completas, mas não necessariamente organizadas</p>	<p>O estudo sobre a construção de um Sistema de Recomendação de Tecnologia utilizando Web Crawling e Processamento de Linguagem Natural pode ser aplicado em repositórios institucionais para automatizar a busca e sumarização de informações relevantes sobre tecnologia. Isso pode facilitar a descoberta de novas tecnologias, auxiliar na tomada de decisões e agilizar o processo de pesquisa para os usuários dos repositórios, tornando a recuperação de informações mais eficiente e eficaz.</p>

		<p>hardware, software ou outras soluções de engenharia, por meio do uso de rastreamento da web em combinação com algoritmos de processamento de linguagem natural.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>contexto da aplicação considerada, destacando os benefícios potenciais por meio de experimentos de avaliação.</p>	<p>de forma significativa. Além disso, alguns resumos não correspondiam muito bem aos termos de pesquisa. No entanto, o artefato desenvolvido mostrou valor em um novo caso de uso identificado, oferecendo ao usuário uma maneira única de descobrir em quais contextos o termo de pesquisa também é usado, permitindo encontrar produtos e serviços complementares de forma eficiente.</p>	
2023	A Comparative Study on R Packages for Text Mining	<p>O problema de pesquisa abordado no estudo é a comparação entre diferentes pacotes R para mineração de texto, com o objetivo de avaliar suas funcionalidades, desempenho e aplicabilidade em diversas áreas de análise de texto. A pesquisa visa identificar as vantagens e limitações de cada pacote, fornecendo insights para profissionais e pesquisadores interessados em utilizar ferramentas de mineração de texto no ambiente R.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>O objetivo do estudo é realizar uma análise comparativa e de benchmarking dos pacotes R para técnicas de mineração de texto, com foco no desempenho, tempo de execução e uso de memória. A pesquisa busca fornecer uma avaliação abrangente dos pacotes disponíveis, destacando suas capacidades em áreas específicas da mineração de texto, como processamento de linguagem natural, representação de conhecimento, extração de informações e análise de sentimentos. O estudo visa auxiliar os usuários do R na escolha dos melhores pacotes para suas necessidades de análise de texto, contribuindo para o avanço do conhecimento nessa área.</p>	<p>Os resultados da pesquisa indicaram diferenças significativas no desempenho, tempo de execução e uso de memória entre os pacotes R analisados para mineração de texto. Foram identificados pacotes que se destacaram em técnicas específicas, como o pacote tm para Tokenization e Stemming, o fastTextR para Topic Modeling e Normalization, e o textir para Term Frequency-Inverse Document Frequency (TF-IDF). Além disso, foram observadas variações na eficiência dos pacotes em relação ao tempo de execução e uso de memória, com alguns pacotes apresentando melhores resultados em um aspecto do que em outro. Esses resultados fornecem insights valiosos para os usuários do R na escolha dos pacotes mais adequados para</p>	<p>O estudo comparativo dos pacotes R para mineração de texto pode ser aplicado em repositórios institucionais para auxiliar pesquisadores e profissionais na seleção dos melhores pacotes para suas análises de texto. Ao disponibilizar informações sobre o desempenho, tempo de execução e uso de memória dos pacotes, os repositórios podem orientar os usuários na escolha das ferramentas mais adequadas para suas necessidades específicas de mineração de texto. Isso pode contribuir para a eficiência e precisão das análises realizadas nos repositórios institucionais, melhorando a qualidade e relevância dos resultados obtidos.</p>

				suas tarefas de mineração de texto, contribuindo para a otimização do desempenho e eficácia das análises realizadas.	
2023	A fast local citation recommendation algorithm scalable to multi-topics	<p>O problema de pesquisa abordado neste artigo é a recomendação de citações locais em documentos acadêmicos. O foco está em desenvolver um algoritmo eficiente e competitivo para recomendar citações com base em contextos locais de citação, utilizando informações como o texto ao redor da citação, títulos, resumos e metadados dos artigos citados. O objetivo é melhorar a precisão e utilidade das recomendações de citações, tornando o processo de escrita acadêmica mais eficiente e eficaz para os pesquisadores.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>Neste estudo, o principal objetivo é propor um sistema inovador de recomendação de citações locais, denominado FLCR (Fast Local Citation Recommendation), que visa melhorar a eficiência e precisão das recomendações de citações em documentos acadêmicos. Além disso, o estudo busca explorar uma nova área de pesquisa em recomendação de citações multi-tópicas, estendendo o sistema base para lidar com essa complexidade (FLCRM). Os pesquisadores almejam avaliar o desempenho do sistema em um amplo conjunto de dados, demonstrando sua superioridade em relação a abordagens anteriores. Adicionalmente, disponibilizarão um conjunto de dados em grande escala e um sistema de demonstração online para facilitar a aplicação prática do sistema por pesquisadores em suas atividades acadêmicas.</p>	<p>O estudo apresenta um sistema de recomendação de citações locais e multi-tópicas, demonstrando que o modelo proposto (FLCR e FLCRM) supera métodos anteriores em termos de desempenho. O FLCRM consegue distinguir diferentes tópicos de forma eficaz, proporcionando recomendações relevantes e diversificadas. Os resultados indicam que o sistema pode ser aplicado em repositórios institucionais para melhorar a precisão e eficiência das recomendações de citações em documentos acadêmicos.</p>	<p>O estudo propõe um sistema inovador de recomendação de citações locais e multi-tópicas, que pode ser aplicado em repositórios institucionais para melhorar a precisão e eficiência das recomendações de citações em documentos acadêmicos. Ao implementar o FLCR e FLCRM em repositórios, os pesquisadores e acadêmicos podem obter sugestões relevantes de referências com base em contextos locais de citação e em múltiplos tópicos, facilitando a identificação de trabalhos relacionados e ampliando a diversidade de referências disponíveis. Isso pode aprimorar a qualidade das pesquisas, acelerar o processo de escrita acadêmica e auxiliar na descoberta de novos conhecimentos.</p>
2023	A high-performance speech BioHashing retrieval algorithm based on audio segmentatio	<p>O problema de pesquisa abordado neste estudo é a necessidade de melhorar a precisão e eficiência da recuperação de fala em um</p>	<p>O objetivo deste estudo é propor um algoritmo de recuperação de fala de alto desempenho baseado em BioHashing e segmentação de áudio.</p>	<p>O resultado da pesquisa foi a proposição de um algoritmo de recuperação de fala de alto desempenho baseado em BioHashing e segmentação de</p>	<p>O estudo sobre o algoritmo de recuperação de fala baseado em BioHashing e segmentação de áudio pode ser aplicado em repositórios institucionais para</p>

		<p>contexto de grandes volumes de dados, alta redundância de dados e métodos complexos de indexação que ocorrem no processo de recuperação de fala. O objetivo é desenvolver um algoritmo de recuperação de fala de alto desempenho baseado em BioHashing que se baseia na segmentação de áudio para otimizar a validade dos dados de fala, reduzir o espaço de armazenamento de dados e melhorar a eficiência da recuperação.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>Especificamente, o objetivo é melhorar a precisão e eficiência da recuperação de fala, lidando com desafios como grandes volumes de dados de fala, alta redundância de conteúdo e métodos complexos de indexação. O algoritmo visa reduzir o espaço de armazenamento de dados, remover partes redundantes da fala e otimizar o desempenho de recuperação, proporcionando uma solução eficaz para a recuperação de fala em ambientes diversos.</p>	<p>áudio. Este algoritmo demonstrou melhorias significativas na precisão e eficiência da recuperação de fala, superando desafios como grandes volumes de dados de fala, alta redundância de conteúdo e complexidade nos métodos de indexação. Os resultados experimentais mostraram que o algoritmo alcançou uma precisão de recuperação de 100% para os cliques de fala originais, com um tempo médio de recuperação de apenas 0,0157 segundos. Além disso, a remoção de segmentos redundantes silenciosos resultou em sequências de hash reconstruídas com melhor robustez e capacidade de discriminação. Esses resultados indicam que o algoritmo proposto possui um desempenho de recuperação eficaz e atende às necessidades de recuperação de fala em diversos ambientes.</p>	<p>aprimorar a busca e recuperação de conteúdo de fala de forma eficiente e precisa. Ao implementar esse algoritmo, os repositórios institucionais podem melhorar a organização e indexação de dados de fala, reduzir redundâncias, otimizar o espaço de armazenamento e aprimorar a capacidade de recuperação de informações relevantes. Isso resultaria em uma experiência de busca mais eficaz para os usuários, facilitando o acesso a conteúdos específicos de fala dentro do repositório.</p>
2023	A Method to Detect Chorus Sections in Lyrics Text	<p>O problema de pesquisa abordado neste estudo é a detecção de seções de refrão em letras de músicas em inglês e japonês. Os autores propõem um modelo baseado em redes neurais para rotulagem de sequência, visando identificar linhas líricas repetidas e características linguísticas que indicam seções de refrão. Eles</p>	<p>O objetivo deste estudo é desenvolver um modelo supervisionado capaz de detectar seções de refrão em letras de músicas em inglês e japonês. Os autores propõem utilizar características estruturais que representam padrões de linhas líricas repetidas, bem como características linguísticas calculadas a partir de técnicas</p>	<p>Os resultados da pesquisa demonstram que o modelo proposto pelos autores superou os modelos basais alternativos na detecção de seções de refrão em letras de músicas em inglês e japonês. A combinação de características estruturais e linguísticas contribuiu para um melhor desempenho na tarefa. Além disso, os experimentos</p>	<p>O estudo sobre a detecção de seções de refrão em letras de músicas, utilizando um modelo baseado em redes neurais, pode ser aplicado em repositórios institucionais de diversas formas. Por exemplo, a implementação desse modelo pode auxiliar na análise e organização de coleções de letras de músicas, permitindo a</p>

		<p>enfrentam o desafio de trabalhar em um cenário onde há poucos trabalhos anteriores sobre essa tarefa específica, o que destaca a novidade e a relevância do problema abordado no estudo.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>como word2vec e BERT. Eles buscam investigar quais modelos e características são eficazes para a detecção de seções de refrão, com a finalidade de superar métodos basais alternativos e melhorar o desempenho na tarefa.</p>	<p>mostraram que o modelo treinado com o grande conjunto de dados gerado obteve resultados promissores, mesmo detectando seções de refrão em letras de músicas em inglês de forma surpreendentemente eficaz. Os padrões de linhas líricas repetidas foram considerados independentes do idioma, destacando a robustez do modelo proposto.</p>	<p>identificação automática de seções de refrão. Isso pode facilitar a indexação e recuperação de informações musicais, bem como contribuir para pesquisas em áreas como processamento de linguagem natural e recuperação de informações musicais. Além disso, a abordagem de combinar características estruturais e linguísticas pode ser adaptada para melhorar a análise e categorização de conteúdo textual em repositórios institucionais, proporcionando uma maior eficiência na organização e busca de informações.</p>
2023	<p>A novel algorithm for the construction of fast English sentence retrieval model using a combination of ontology and advanced machine learning techniques</p>	<p>O problema de pesquisa abordado no documento é a construção de um modelo de recuperação rápida de frases em inglês que seja eficiente e preciso. O estudo propõe a utilização de uma combinação de ontologia e técnicas avançadas de aprendizado de máquina para aprimorar a compreensão de linguagem natural e a recuperação de informações com maior precisão e rapidez.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>O objetivo do estudo é desenvolver um modelo de recuperação rápida de frases em inglês que integre ontologia e técnicas avançadas de aprendizado de máquina. Esse modelo visa melhorar a precisão e a eficiência na compreensão da linguagem natural, permitindo a tradução de expressões em linguagem natural para entidades conceituais associadas, resultando em uma recuperação mais precisa e relevante de informações para os usuários.</p>	<p>O resultado da pesquisa foi a criação de um modelo de recuperação rápida de frases em inglês que utiliza uma combinação de ontologia e técnicas avançadas de aprendizado de máquina. Esse modelo demonstrou um impressionante índice de recuperação de 98,5%, superando modelos existentes em comparação. Além disso, o algoritmo proposto apresentou um desempenho superior ao algoritmo TF-IDF, com uma melhoria de 7,52% na precisão. Quando aplicado a um corpus com poucos rótulos e um corpus não rotulado relativamente grande, o algoritmo melhorou o</p>	<p>O estudo sobre o modelo de recuperação rápida de frases em inglês, que combina ontologia e técnicas avançadas de aprendizado de máquina, pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de informações de forma mais precisa e eficiente. Ao integrar a ontologia e o aprendizado de máquina, o modelo pode facilitar a correspondência semântica e a tradução de consultas em linguagem natural para entidades conceituais, melhorando a relevância dos resultados recuperados para os usuários dos repositórios institucionais.</p>

				desempenho do classificador em 12,6%. Esses resultados indicam que o modelo proposto reduz a influência do processamento de sinônimos no desempenho geral, mantendo a precisão e a qualidade dos resultados de cálculo.	
2023	A Novel Artificial Social Cognitive Agent Model for Teaching the Concepts	<p>O problema de pesquisa abordado neste estudo é a necessidade de desenvolver um modelo de agente cognitivo artificial que possa extrair conhecimento de documentos fornecidos por professores, utilizando técnicas híbridas de Aprendizado de Máquina e Sistemas Baseados em Regras. Este agente deve ser capaz de compreender perguntas dos usuários, processar informações de forma significativa e fornecer respostas relevantes, superando as limitações dos sistemas de recuperação de informações existentes.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>O objetivo deste estudo é propor e avaliar um modelo de agente cognitivo artificial que utilize a cognição social para ensinar conceitos. O modelo visa extrair conhecimento de documentos por meio de técnicas híbridas de Aprendizado de Máquina e Sistemas Baseados em Regras, permitindo que o agente tome decisões na vida real. Além disso, o modelo considera os aspectos comportamentais e as restrições sociais do agente, com foco na avaliação de desempenho no domínio de Motores Elétricos. A proposta é estender esse modelo para outras áreas, como a tecnologia da informação, e potencialmente desenvolver um humanoide baseado no agente cognitivo artificial para futuros trabalhos.</p>	<p>Os resultados da pesquisa indicam que o modelo de agente cognitivo artificial proposto, que utiliza a cognição social, obteve resultados satisfatórios ao ser avaliado no domínio de Motores Elétricos. O agente demonstrou capacidade de tomar decisões na vida real e considerar aspectos comportamentais e restrições sociais. A avaliação do desempenho do agente foi realizada com base em diversos parâmetros de avaliação, e os resultados obtidos foram considerados razoáveis. Além disso, o estudo sugere que o modelo pode ser estendido para outras áreas, como a tecnologia da informação, e que há potencial para o desenvolvimento de um humanoide baseado nesse agente para trabalhos futuros.</p>	<p>O estudo propõe um modelo de agente cognitivo artificial que pode ser aplicado em repositórios institucionais para melhorar a extração de conhecimento de documentos e fornecer respostas relevantes aos usuários. Ao utilizar técnicas de cognição social e processamento de linguagem natural, o agente pode ajudar na recuperação de informações de forma mais eficaz e na tomada de decisões com base nesse conhecimento. Isso pode aprimorar a experiência dos usuários ao interagir com os repositórios, tornando a busca por informações mais precisa e eficiente.</p>
2023	A Novel Joint Extraction Model for Entity Relations Using Interactive Encoding and Visual Attention	<p>O problema de pesquisa abordado neste estudo é a extração conjunta de entidades e relações em textos, com foco na modelagem das interações bidirecionais entre sujeitos e</p>	<p>O objetivo deste estudo é propor um novo modelo de extração conjunta de entidades e relações que seja capaz de capturar as interações bidirecionais entre sujeitos e objetos, bem como</p>	<p>Os resultados da pesquisa demonstraram que o modelo proposto obteve um desempenho superior em comparação com outros modelos de referência em várias</p>	<p>O estudo sobre o modelo de extração conjunta de entidades e relações pode ser aplicado em repositórios institucionais para aprimorar a extração de informações de textos</p>

		<p>objetos, entidades e relações. O objetivo é superar as limitações dos modelos existentes, que muitas vezes não capturam adequadamente as informações de interação entre sujeitos e objetos, nem exploram as relações de localização espacial entre entidades. A proposta é desenvolver um modelo inovador que possa melhorar a precisão da extração de entidades e relações, além de aumentar a robustez do modelo treinado.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>entre entidades e relações em textos. O modelo visa melhorar a precisão da extração de entidades e relações, além de aumentar a robustez do modelo treinado, especialmente ao explorar as informações de interação e as relações de localização espacial entre entidades. A proposta inclui a utilização de uma unidade de codificação de texto inovadora e a introdução de um mecanismo de atenção visual para aprimorar as relações espaciais entre triplas de relacionamento, visando aprimorar a precisão na extração de relações.</p>	<p>bases de dados publicamente disponíveis. Através de experimentos de ablação realizados em cada componente do modelo, foi possível avaliar a eficácia de cada parte do modelo. Os resultados indicaram que a unidade de filtragem de recursos paralelos e o mecanismo de atenção Shuffle contribuíram significativamente para a precisão da extração de informações de interação entre entidades, relações, sujeitos e objetos. Além disso, a utilização das camadas intermediárias do modelo de linguagem pré-treinado para extração de recursos mostrou-se eficaz em comparação com a utilização apenas da última camada.</p>	<p>acadêmicos e científicos. Ao utilizar a unidade de codificação de texto inovadora e o mecanismo de atenção visual propostos, os repositórios institucionais podem melhorar a precisão na identificação de entidades e relações em documentos, facilitando a organização e recuperação de informações relevantes. Isso pode resultar em uma melhor indexação e categorização de conteúdo, tornando mais eficiente a busca e análise de dados em repositórios acadêmicos e científicos.</p>
2023	<p>AI-Assisted Deep NLP-Based Approach for Prediction of Fake News From Social Media Users</p>	<p>O problema de pesquisa abordado neste estudo é a detecção e prevenção da propagação de notícias falsas (fake news) originadas de usuários de mídias sociais. O foco está em desenvolver uma abordagem baseada em inteligência artificial e processamento de linguagem natural profundo para identificar e classificar notícias falsas com alta precisão. A pesquisa visa combater a disseminação de informações enganosas que</p>	<p>O objetivo do estudo é desenvolver uma abordagem baseada em inteligência artificial e processamento de linguagem natural profundo para atingir dois principais objetivos: detectar artigos de notícias falsas por meio de uma rede NLP-integrada de CNN Bi-LSTM profunda com atenção (N-DCBL) para evitar sua propagação adicional e verificar a credibilidade das contas de usuários de mídias sociais e dos editores por meio de uma rede</p>	<p>O resultado da pesquisa foi a proposição de um modelo de detecção de notícias falsas integrando processamento de linguagem natural (NLP) e aprendizado profundo. O modelo, denominado N-DCBL, foi avaliado utilizando três conjuntos de dados: Buzzface, FakeNewsNet e Twitter. Os resultados da simulação demonstraram que o modelo proposto obteve uma precisão média de 99,72% e um escore F1 de 98,33%, superando outros</p>	<p>O estudo propõe um modelo de detecção de notícias falsas baseado em inteligência artificial e processamento de linguagem natural profundo. Esse modelo pode ser aplicado em repositórios institucionais para verificar a veracidade das informações compartilhadas. Ao integrar o NLP e o aprendizado profundo, o modelo N-DCBL pode ajudar a identificar e classificar notícias falsas, evitando a propagação de desinformação. Isso pode ser</p>

		<p>podem afetar a reputação de personalidades conhecidas e prejudicar a reputação de um país.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>de atenção profunda baseada em estrutura multiheaded habilitada para aprendizado profundo. Esses objetivos visam aprimorar a detecção e identificação de notícias falsas e rumores que se espalham nas redes sociais, contribuindo para a redução da disseminação de desinformação.</p>	<p>métodos existentes de detecção de notícias falsas. Além disso, o modelo N-DCBL foi capaz de prever a credibilidade de usuários e editores de mídias sociais, bloqueando contas consideradas não confiáveis. Esses resultados destacam a eficácia do modelo proposto na detecção precoce e prevenção da propagação de notícias falsas, contribuindo para a mitigação da disseminação de desinformação.</p>	<p>especialmente útil em repositórios institucionais para garantir a precisão e confiabilidade das informações disponibilizadas aos usuários.</p>
2023	<p>An Alternative Application of Natural Language Processing to Express a Characteristic Feature of Diseases in Japanese Medical Records</p>	<p>O problema de pesquisa abordado neste estudo é a identificação e expressão de características específicas de doenças em registros médicos japoneses utilizando processamento de linguagem natural (NLP). A pesquisa compara o desempenho de NLP baseado em descoberta aberta (OD-NLP) com NLP baseado em dicionário de palavras (WD-NLP) para determinar qual abordagem é mais eficaz na extração de termos importantes relacionados a doenças em textos clínicos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é investigar se o uso de NLP baseado em descoberta aberta (OD-NLP), que não utiliza técnicas de dicionário, pode ser uma alternativa viável para identificar e expressar características distintivas de doenças em registros médicos japoneses. Os pesquisadores buscam comparar a eficácia do OD-NLP com o NLP baseado em dicionário de palavras (WD-NLP) na geração de tópicos em documentos clínicos e na correspondência desses tópicos com doenças específicas listadas na Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde, 10ª Revisão (CID-10).</p>	<p>Os resultados da pesquisa mostraram que, ao analisar documentos de 10.520 pacientes, tanto o OD-NLP quanto o WD-NLP segmentaram um grande número de entidades e palavras simultaneamente. Sem filtragem, ambos os métodos apresentaram baixos níveis de precisão e recall, sem diferença significativa na média harmônica da medida F. No entanto, os médicos observaram que o OD-NLP continha mais palavras significativas do que o WD-NLP. Quando os conjuntos de dados foram equilibrados usando TF-IDF, a medida F no OD-NLP foi maior em limiares mais baixos, indicando uma melhor performance. A análise dos conjuntos de dados próximos ao limiar máximo mostrou que o OD-NLP identificou mais doenças,</p>	<p>O estudo demonstrou que o uso de processamento de linguagem natural baseado em descoberta aberta (OD-NLP) pode ser uma abordagem eficaz para identificar e expressar características distintivas de doenças em registros médicos japoneses. Essa técnica pode ser aplicada em repositórios institucionais de saúde para melhorar a análise de documentos clínicos, facilitar a identificação de tópicos relevantes relacionados a doenças e auxiliar na recuperação de informações em ambientes clínicos. A implementação do OD-NLP pode contribuir para a construção de resumos de documentos e aprimorar a busca e recuperação de informações em registros médicos eletrônicos, beneficiando a</p>

				sugerindo que seus tópicos descreviam características das doenças de forma mais precisa. Mesmo quando a filtragem foi alterada para DMV, a superioridade do OD-NLP foi mantida.	prática clínica e a pesquisa em saúde.
2023	An Analytical Analysis of Text Stemming Methodologies in Information Retrieval and Natural Language Processing Systems	<p>O problema de pesquisa abordado no documento é a análise e avaliação das metodologias de stemming de texto em sistemas de recuperação de informação e processamento de linguagem natural. O estudo visa identificar as abordagens existentes, suas vantagens e desvantagens, bem como os desafios e questões que afetam o desempenho dos sistemas de stemming de texto. Além disso, o trabalho busca destacar lacunas na pesquisa atual e fornecer direções para futuros estudos na área de stemming de texto.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo do estudo é realizar uma análise abrangente e comparativa das diferentes abordagens de stemming de texto, incluindo stemmers baseados em conhecimento linguístico, estatísticos, baseados em corpus e híbridos. O estudo visa identificar a eficácia e precisão dessas abordagens em diferentes contextos, como em línguas altamente infletidas, e destacar os desafios e questões enfrentados por cada tipo de stemmer. Além disso, o trabalho busca fornecer insights sobre a avaliação de stemmers de texto e recomendações para pesquisas futuras, incluindo a utilização de algoritmos de aprendizado de máquina para aprimorar a extração de significado semântico de consultas de usuários.</p>	<p>Os resultados da pesquisa indicam que as técnicas de stemming baseadas em conhecimento linguístico são amplamente utilizadas em línguas altamente infletidas, como árabe, hindi e urdu, apresentando maior precisão. No entanto, essas abordagens enfrentam desafios como o reconhecimento de exceções e homógrafos. Por outro lado, as técnicas de stemming não supervisionadas lidam com questões como extração de significado semântico e ajuste de parâmetros automáticos. A seleção do método de avaliação de stemming de texto ainda é uma questão em aberto, com diferentes comportamentos observados em conjuntos de dados pequenos e grandes. Recomenda-se o uso de algoritmos de aprendizado de máquina para aprimorar a extração de significado semântico de consultas de usuários e partes de discurso. A escolha da técnica de stemming adequada é um processo complexo que requer</p>	<p>O estudo sobre metodologias de stemming de texto pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações e a análise de conteúdo. A utilização de stemmers pode ajudar na redução da variabilidade lexical, melhorando a precisão das buscas e facilitando a categorização e organização dos documentos. Além disso, a implementação de técnicas de stemming pode contribuir para a identificação de padrões e tendências nos dados armazenados nos repositórios, auxiliando na descoberta de informações relevantes e na otimização da experiência do usuário.</p>

				investigação detalhada devido à diversidade dos conjuntos de dados e bases de conhecimento.	
2023	An efficient long-text semantic retrieval approach via utilizing presentation learning on short-text	<p>O problema de pesquisa abordado neste estudo é a eficiência na recuperação de informações em textos longos, especialmente em cenários onde a interação com cada texto longo ou parágrafo é computacionalmente custosa e demorada. O objetivo é desenvolver um método de recuperação de informações em textos longos que seja mais eficiente, combinando a estrutura de duas torres e as vantagens do modelo BERT para representação contextual.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é propor um modelo eficiente de recuperação de informações em textos longos baseado no BERT (LTR-BERT) que melhore a velocidade de recuperação mantendo a maior parte do desempenho de recuperação de textos longos. O foco está em treinar o modelo LTR-BERT usando a relevância entre textos curtos, segmentar e armazenar off-line o texto longo, calcular apenas a codificação da consulta e os escores de correspondência durante a etapa de recuperação, e projetar uma estratégia de expansão de consulta para aprimorar a representação da consulta original. Além disso, o estudo visa garantir a precisão da recuperação e reduzir o tempo de resposta por meio de um mecanismo de interação sem parâmetros de treinamento.</p>	<p>Os resultados da pesquisa demonstraram que o modelo LTR-BERT proposto obteve um aumento de 2,74% nos valores de MRR@10 em comparação com o método de correspondência semântica focado na interação por BERT-CLS. Além disso, o número de documentos processados por milissegundo aumentou em 333 vezes. Isso indica uma melhoria significativa na eficiência e desempenho da recuperação de informações em textos longos com a abordagem proposta neste estudo.</p>	<p>O estudo propõe um modelo eficiente de recuperação de informações em textos longos, o LTR-BERT, que pode ser aplicado em repositórios institucionais para melhorar a eficiência e precisão na busca de documentos extensos. Ao utilizar o BERT para representação contextual e estratégias de interação e expansão de consulta, o LTR-BERT pode acelerar o processo de recuperação de informações sem comprometer a qualidade dos resultados. Isso pode beneficiar repositórios institucionais ao tornar a busca por documentos longos mais rápida e eficaz, melhorando a experiência do usuário e aumentando a produtividade na pesquisa de informações.</p>
2023	An Intelligent Conversational Agent for the Legal Domain	<p>O problema de pesquisa abordado no documento é a construção de um sistema de recuperação de informações baseado em deep learning usando um conjunto de pares de perguntas e respostas de resolução de conflitos legais. O</p>	<p>O objetivo do estudo é construir um sistema de recuperação de informações baseado em deep learning usando um conjunto de pares de perguntas e respostas de resolução de conflitos legais. Esse sistema visa servir como base para um agente de</p>	<p>O resultado da pesquisa demonstrou a eficácia dos modelos SBERT na construção de um sistema de recuperação de informações capaz de fornecer respostas úteis sobre resolução de conflitos legais a partir de consultas comuns. O</p>	<p>O estudo demonstrou a eficácia dos modelos SBERT na construção de um sistema de recuperação de informações para resolução de conflitos legais. Essa abordagem pode ser aplicada em repositórios institucionais para melhorar a</p>

		<p>objetivo é desenvolver um agente de conversação inteligente que possa fornecer recursos valiosos ao ser solicitado com consultas semelhantes. A pesquisa se concentra em associar cada pergunta com a resposta relacionada de forma apropriada, abordando o problema como uma recuperação de perguntas semelhantes em vez de uma recuperação de pergunta e resposta tradicional.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>	<p>conversação inteligente que pode fornecer recursos valiosos ao ser solicitado com consultas semelhantes. O foco está em associar cada pergunta com a resposta relacionada de forma apropriada, abordando o problema como uma recuperação de perguntas semelhantes em vez de uma recuperação de pergunta e resposta tradicional.</p>	<p>estudo destacou a importância de explorar o desempenho do sistema em uma gama mais ampla de tópicos legais e investigar maneiras de aprimorar suas capacidades de generalização, uma vez que a solução considerou apenas um conjunto específico de tópicos legais. Além disso, sugere a possibilidade de avaliação humana como uma métrica de avaliação mais abrangente. O trabalho também delineou futuras direções, vislumbrando uma solução mais abrangente para fornecer aconselhamento legal por meio de um chatbot generativo alimentado por um grande modelo de linguagem e um banco de dados de instrumentos legislativos e casos legais relevantes.</p>	<p>recuperação de informações legais relevantes. Ao utilizar modelos de linguagem avançados e técnicas de recuperação de informações baseadas em deep learning, os repositórios institucionais podem oferecer respostas mais precisas e úteis aos usuários que buscam orientação legal. Isso pode aprimorar a eficiência e a qualidade dos serviços prestados por esses repositórios, tornando a pesquisa e a navegação por informações legais mais eficazes e acessíveis.</p>
2023	An Intelligent Semi-Honest System for Secret Matching against Malicious Adversaries	<p>O problema de pesquisa abordado no artigo é a realização de correspondência segura de texto (text matching) em um ambiente onde participantes maliciosos podem estar presentes. O foco está em desenvolver um sistema inteligente semi-honesto que possa realizar essa correspondência de forma segura, resistindo a comportamentos maliciosos que possam ser cometidos por adversários maliciosos.</p>	<p>O objetivo do estudo é propor um Sistema Inteligente Semi-Honesto (ISHS) para realizar correspondência segura de texto contra adversários maliciosos. O sistema visa oferecer uma solução eficiente e segura para a correspondência de texto, utilizando métodos criptográficos avançados e protocolos de computação segura para proteger a privacidade dos dados e garantir a integridade do processo de correspondência. Além disso, o estudo busca</p>	<p>Os resultados da pesquisa incluem a proposição de um método de codificação mais eficiente para a criptografia ECC, o desenvolvimento de um protocolo de correspondência difusa de strings de texto baseado no modelo semi-honesto, a implementação de ferramentas criptográficas para resistir a ataques maliciosos, a demonstração da segurança do protocolo e a análise da eficiência por meio de simulações experimentais.</p>	<p>O estudo sobre o Sistema Inteligente Semi-Honesto para Correspondência Secreta contra Adversários Maliciosos pode ser aplicado em repositórios institucionais para garantir a segurança e privacidade dos dados armazenados. Através do protocolo desenvolvido, é possível realizar correspondência segura de texto, protegendo as informações contra possíveis ataques maliciosos. Isso pode ser especialmente útil em</p>

		<p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>analisar a eficiência computacional e a segurança do protocolo proposto, demonstrando sua viabilidade e relevância para aplicações práticas no campo da segurança cibernética.</p>	<p>Esses resultados contribuem significativamente para a área de segurança cibernética, oferecendo uma solução inovadora e eficaz para a correspondência segura de texto em ambientes com adversários maliciosos.</p>	<p>ambientes onde a integridade dos dados é crucial, como em repositórios de pesquisa, garantindo que as informações sensíveis estejam protegidas contra acessos não autorizados e preservando a confidencialidade dos dados armazenados.</p>
2023	An unsupervised perplexity-based method for boilerplate removal	<p>O problema de pesquisa abordado no estudo é a remoção de conteúdo indesejado e mal formado, como boilerplate, de dados extraídos da web, visando aprimorar a eficácia e eficiência de soluções de IA e busca.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>Os modelos de linguagem contribuem para a eficácia e eficiência da extração de conteúdo útil de dados extraídos da web, utilizando a perplexidade para avaliar a probabilidade de sentenças. Isso permite que o método identifique e remova conteúdo clichê, barulhento ou irrelevante, resultando em dados mais limpos e relevantes para tarefas posteriores, como recuperação de informações e classificação de texto. Os Modelos de Linguagem auxiliam na avaliação da qualidade do texto e permitem a extração de conteúdos bem formados, levando a um melhor desempenho em diversas tarefas relacionadas à web.</p>	<p>A pesquisa resultou em um método eficaz e eficiente baseado em perplexidade para a remoção de boilerplate de dados da web. Os pesquisadores demonstraram que o uso de Modelos de Linguagem pode ajudar a extrair conteúdo útil e bem formado, removendo partes indesejadas e ruidosas. Isso resulta em soluções de IA ou busca mais leves, que são mais eficazes e econômicas em termos de recursos.</p>	<p>O estudo propõe um método baseado em perplexidade para remover conteúdo indesejado de páginas da web, tornando as soluções de IA mais eficientes. Esse método pode ser aplicado em repositórios institucionais para extrair conteúdo relevante e limpo de documentos, melhorando a eficácia das buscas e classificações realizadas nesses repositórios.</p>
2023	Automatic extraction of inventive information out of patent texts in support of manufacturing design studies using Natural Languages Processing	<p>O problema de pesquisa abordado no documento é a extração automática de informações inventivas de textos de patentes para apoiar estudos de design de manufatura. A pesquisa visa superar os</p>	<p>O objetivo do estudo é apresentar uma metodologia para a extração automática de informações inventivas de textos de patentes, combinando a teoria de resolução de problemas inventivos (TRIZ)</p>	<p>Os resultados da pesquisa indicam que a ferramenta de extração conseguiu extrair 372 parâmetros dos textos de patentes analisados. Desses, 142 parâmetros correspondiam aos termos da lista de referência</p>	<p>O estudo sobre a extração automática de informações inventivas de textos de patentes usando NLP pode ser aplicado em repositórios institucionais para facilitar a análise e extração de conhecimento de grandes</p>

		<p>desafios de gerenciar grandes volumes de informações contidas em textos de patentes e extrair de forma eficiente conceitos inventivos relevantes para o processo de design na indústria de manufatura.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>com Processamento de Linguagem Natural (NLP). O estudo busca demonstrar como essa abordagem pode reduzir significativamente o tempo necessário para extrair informações úteis, aumentando em 36% a eficiência em comparação com a coleta de informações conduzida por especialistas. Além disso, um estudo de caso sobre a colheita de microplásticos do oceano é apresentado para ilustrar a aplicação da metodologia.</p>	<p>("gold list"), que serviu como base para a avaliação estatística dos resultados. A precisão, recall e F-measure foram calculados para diferentes cenários de aplicação da metodologia, incluindo etapas de pré-processamento, implementação do algoritmo Mallet e aplicação de clustering. O resultado mais favorável foi obtido quando a metodologia completa foi aplicada, resultando em uma precisão de 0,71, recall de 0,4 e F-measure de 0,51. Esses resultados demonstram a eficácia da abordagem proposta na extração de informações inventivas de textos de patentes.</p>	<p>volumes de documentos patentários. A metodologia proposta, que combina TRIZ e NLP, pode ser implementada em sistemas de gerenciamento de informações para identificar parâmetros relevantes, resolver problemas inventivos e acelerar o processo de inovação em diversas áreas da indústria. Essa abordagem automatizada pode ajudar a otimizar a busca por informações valiosas em textos de patentes, contribuindo para a tomada de decisões estratégicas e o desenvolvimento de novas soluções tecnológicas.</p>
2023	<p>Concept-Aware Video Captioning: Describing Videos With Effective Prior Information</p>	<p>O problema de pesquisa abordado neste artigo é a melhoria da geração de legendas de vídeos por meio da incorporação de informações prévias eficazes e da detecção precisa de conceitos no contexto de legendagem de vídeos. A proposta é desenvolver o framework CARE (Concept-Aware Video Captioning) para aprimorar a qualidade e a precisão das legendas geradas para vídeos, utilizando conceitos identificados de forma precisa e orientando a geração de legendas com base em informações globais e locais semânticas.</p>	<p>O objetivo deste estudo é propor e desenvolver o framework Concept-Aware Video Captioning (CARE) para facilitar a geração de legendas de vídeos de forma mais plausível e precisa. O CARE busca detectar conceitos de forma precisa por meio da detecção de conceitos orientada por multimodalidade e fornecer informações prévias adequadas para a geração de legendas por meio de orientação semântica global-local. Contudo, o objetivo é melhorar a qualidade e a precisão das legendas geradas para vídeos, aproveitando conceitos identificados de forma precisa e orientando o processo de</p>	<p>Os resultados da pesquisa indicam que o framework Concept-Aware Video Captioning (CARE) proposto neste estudo demonstrou melhorias significativas na geração de legendas de vídeos em comparação com baselines fortes e métodos de ponta existentes. Os experimentos extensivos realizados em conjuntos de dados como MSVD, MSR-VTT e VATEX mostraram que o CARE se adapta bem a várias redes codificador-decodificador e supera os métodos tradicionais de legendagem de vídeos. Além disso, o estudo destaca que o CARE foi capaz de detectar</p>	<p>O estudo sobre Concept-Aware Video Captioning (CARE) pode ser aplicado em repositórios institucionais para melhorar a descrição de vídeos e torná-los mais acessíveis e compreensíveis para os usuários. Ao utilizar o framework CARE, os repositórios podem gerar legendas de vídeos de forma mais precisa e plausível, facilitando a busca e o entendimento do conteúdo audiovisual. Isso pode melhorar a experiência do usuário, tornar o conteúdo mais inclusivo e aumentar a utilidade e relevância dos vídeos armazenados nos repositórios institucionais.</p>

		<p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Relevance Feedback, expansão da consulta e serviços de recomendação</p>	<p>geração de legendas com base em informações semânticas globais e locais.</p>	<p>conceitos de forma precisa por meio da detecção de conceitos orientada por multimodalidade e oferecer informações prévias adequadas para a geração de legendas por meio de orientação semântica global-local.</p>	
2023	Consumer Document Analytical Accelerator Hardware	<p>O problema de pesquisa abordado no documento é a necessidade de desenvolver um acelerador de hardware para análise de documentos do consumidor que possa lidar com tarefas baseadas em processamento de linguagem natural, como recuperação de informações e sumarização de texto, de forma eficiente e segura. O foco está na implementação de um sistema que utilize a equação TF-IDF (Term Frequency-Inverse Document Frequency) por meio de estruturas analógicas de memristores em cruzamento, visando melhorar a velocidade e a confidencialidade das tarefas de processamento de linguagem natural em dispositivos de borda.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo do estudo é desenvolver um acelerador de hardware para análise de documentos do consumidor que utilize a implementação memristiva do TF-IDF para melhorar a eficiência e a segurança das tarefas de processamento de linguagem natural, como recuperação de informações e sumarização de texto. O sistema proposto visa lidar com as limitações de recursos e conectividade dos dispositivos de processamento próximos à borda, reduzindo a latência durante o processamento em nuvem e os custos computacionais. Além disso, o objetivo é comparar o desempenho do acelerador de hardware com conjuntos de dados de referência padrão e abordar questões de variabilidade e relacionadas a dispositivos na análise.</p>	<p>O resultado da pesquisa foi a implementação bem-sucedida de um acelerador de hardware para análise de documentos do consumidor, utilizando a equação TF-IDF com memristores em cruzamento. O sistema desenvolvido foi capaz de realizar tarefas de processamento de linguagem natural, como recuperação de informações e sumarização de texto, de forma eficiente e segura. A arquitetura em camadas permitiu a inclusão de múltiplos documentos com mais de dez frases, armazenando cada palavra em modo de coluna única. A utilização de estruturas analógicas de memristores, comparadores e circuitos logarítmicos permitiu simular a operação TF-IDF. O estudo concluiu que o circuito proposto pode ser integrado a dispositivos embarcados na borda para realizar diversas tarefas de processamento de linguagem natural em tempo real, sem</p>	<p>O estudo sobre o acelerador de hardware para análise de documentos do consumidor, utilizando a implementação memristiva do TF-IDF, pode ser aplicado em repositórios institucionais para aprimorar a eficiência e segurança das tarefas de processamento de linguagem natural. Ao integrar esse sistema em repositórios institucionais, é possível melhorar a recuperação de informações, a sumarização de texto e outras análises de documentos de forma mais rápida e confiável. Isso pode beneficiar a organização, permitindo uma análise mais eficaz e ágil dos documentos armazenados, contribuindo para uma melhor gestão da informação e tomada de decisões embasadas em dados.</p>

				comprometer a confidencialidade ou a redução de latência. Além disso, o trabalho destacou a importância de técnicas de extração de hardware adequadas e modelos de otimização para lidar com grandes volumes de dados e melhorar o desempenho do sistema.	
2023	Creating CREATE queries with multi-task deep neural networks	<p>O problema de pesquisa abordado neste estudo é a criação automática de modelos de banco de dados a partir de expressões em linguagem natural. Os autores buscam desenvolver um sistema que seja capaz de gerar consultas CREATE a partir de descrições em linguagem natural, enfrentando desafios como a extração automática de informações de modelos conceituais representados em texto, a utilização eficiente de modelos de linguagem pré-treinados e a coleta de dados necessária para treinar modelos de aprendizado profundo.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	O objetivo deste estudo é propor um framework geral para a geração automatizada de consultas CREATE a partir de expressões em linguagem natural. Além disso, busca-se desenvolver modelos de referência para avaliar a eficácia do modelo principal proposto e verificar a utilidade de modelos de linguagem pré-treinados na geração não apenas de consultas CREATE, mas também de consultas SELECT.	Com base nos resultados da pesquisa, foi constatado que o problema de criação de tabelas é solucionável quando a definição do problema é corretamente estabelecida. O estudo demonstrou que ao restringir o escopo de saída de forma significativa, o framework proposto pode criar modelos de banco de dados fundamentais. No entanto, a solução apresentada é limitada pelos dados coletados, especialmente devido à baixa quantidade de tipos de dados e restrições disponíveis para treinar o modelo. Como direção futura, sugere-se a proposição de uma solução mais sofisticada que capture diferentes tipos de restrições de banco de dados e seja capaz de gerar modelos de banco de dados mais avançados com um conjunto de dados mais extenso.	O estudo propõe um framework que pode ser aplicado em repositórios institucionais para automatizar a criação de consultas CREATE a partir de descrições em linguagem natural. Ao utilizar modelos de linguagem pré-treinados e técnicas de aprendizado profundo, é possível gerar consultas de forma eficiente e precisa, facilitando a criação e manutenção de modelos de banco de dados em repositórios institucionais de forma automatizada e escalável.

2023	Crop Information Retrieval Framework Based on LDW-Ontology and SNM-BERT Techniques	<p>O problema de pesquisa abordado neste artigo é a dificuldade em recuperar informações precisas e relevantes sobre agricultura na internet, devido a desafios como baixa taxa de relevância, detecção falsa de recursos de recuperação, formato de dados não estruturados, dados multivariados, ortografia irrelevante, e tempo de computação elevado. A pesquisa visa desenvolver um sistema inteligente de Recuperação de Informações (IR) para superar essas complicações, utilizando uma estrutura baseada em Ontologia LDW (Levenshtein Distance Weight-centric) e técnicas SNM-BERT (Sutskever Nesterov Momentum-centred Bidirectional Encoder Representation from Transformer).</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é apresentar um framework de recuperação de informações sobre culturas agrícolas de forma semântica, permitindo a rápida busca de dados relevantes com alta precisão. O foco é superar os desafios existentes na recuperação de informações agrícolas, como baixa taxa de relevância, formato de dados não estruturados, baixa taxa de similaridade, ortografia irrelevante, dados multivariados e tempo de computação elevado. O objetivo é desenvolver um sistema de IR baseado em Ontologia LDW e técnicas SNM-BERT para lidar com esses problemas e melhorar a eficiência da recuperação de informações agrícolas.</p>	<p>Os resultados da pesquisa mostraram que o framework proposto alcançou uma precisão de 94% para consultas simples e 92% para consultas complexas, superando as técnicas existentes. A avaliação experimental comparativa demonstrou que o SNM-BERT proposto obteve melhores resultados em termos de precisão, recall, F-score, acurácia, informações retornadas versus efetivas, resultados recuperados e tempo de recuperação de consulta em comparação com técnicas como LSTM, BiLSTM, BERT e AN. Os valores mais altos dessas métricas indicam uma técnica aprimorada para lidar com a recuperação de informações agrícolas.</p>	<p>O estudo propõe um framework de recuperação de informações agrícolas baseado em Ontologia LDW e técnicas SNM-BERT para melhorar a precisão e eficiência da busca de dados relevantes. Essa abordagem pode ser aplicada em repositórios institucionais para facilitar a recuperação de informações específicas sobre culturas agrícolas, permitindo aos usuários acessar rapidamente dados relevantes com alta precisão. Isso pode beneficiar instituições agrícolas, pesquisadores e profissionais do setor ao tornar mais eficiente o acesso e a análise de informações agrícolas essenciais para tomada de decisões e pesquisa.</p>
2023	Deep internally connected transformer hashing for image retrieval	<p>O problema de pesquisa abordado no artigo é a melhoria da precisão da recuperação de imagens por meio do desenvolvimento de um novo método de hashing profundo</p>	<p>O objetivo do estudo apresentado no artigo é propor e desenvolver o método DICTH (Deep internally connected transformer hashing) para a recuperação de imagens.</p>	<p>Os resultados da pesquisa demonstraram que o método DICTH proposto apresentou melhorias significativas na precisão da recuperação de imagens em comparação com</p>	<p>O estudo sobre o método DICTH para recuperação de imagens pode ser aplicado em repositórios institucionais para aprimorar a eficiência e precisão da busca e recuperação de</p>

		<p>chamado DICTH. O objetivo é explorar a informação de dependência global entre pares de Consulta-Chave (Query-Key) para aprimorar a representação de características das imagens e, assim, a eficácia do processo de recuperação de imagens.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>Especificamente, o objetivo é melhorar a precisão da recuperação de imagens ao combinar um bloco Transformer internamente conectado com a arquitetura ResNet18, visando estabelecer uma dependência de longa distância no espaço de características das imagens. Além disso, o estudo busca utilizar uma função de perda aprimorada, denominada T-cross-entropy loss, para promover a aprendizagem de códigos hash mais distintos entre classes, especialmente em conjuntos de dados grandes com informações de rótulos complexas.</p>	<p>outros métodos de hashing profundo. Por meio de experimentos realizados em conjuntos de dados como CIFAR10, NUS-WIDE e MS-COCO, o DICTH mostrou desempenho superior, destacando-se pela capacidade de explorar semânticas de alta dimensão das imagens, injetar representações de características refinadas com autoatenção multi-cabeça e mapear códigos hash mais precisos e reconhecíveis para aprimorar o desempenho da recuperação de imagens. Além disso, a introdução do módulo Transformer ICT e da perda T-cross-entropy contribuiu significativamente para a eficácia do DICTH, melhorando a capacidade de distinguir entre classes de códigos hash e estabelecer relações de dependência global nas características das imagens.</p>	<p>imagens. Ao incorporar um bloco Transformer internamente conectado e uma função de perda aprimorada, o DICTH melhora a representação de características das imagens, permitindo uma recuperação mais precisa e eficaz. Essa abordagem pode beneficiar repositórios institucionais ao facilitar a busca e recuperação de imagens com maior precisão e rapidez, melhorando a experiência do usuário e a eficácia na localização de conteúdo visual relevante.</p>
2023	Defining Semantically Close Words of Kazakh Language with Distributed System Apache Spark	<p>O problema de pesquisa abordado no estudo é a definição de palavras semanticamente próximas na língua cazaque, visando melhorar a precisão e eficiência de sistemas de recuperação de informações. A pesquisa busca explorar como a utilização de modelos de incorporação de palavras e algoritmos distribuídos, como o Apache</p>	<p>O objetivo do estudo é determinar palavras semanticamente próximas na língua cazaque, utilizando métodos de similaridade semântica para aprimorar o desempenho em tarefas de recuperação de informações. A pesquisa visa representar as palavras em forma de vetores, calcular a similaridade semântica entre termos e</p>	<p>Os resultados da pesquisa incluem o desenvolvimento de um novo conjunto de dados de pares de palavras semanticamente próximas na língua cazaque, a proposição e implementação de um método para encontrar palavras semanticamente similares, a avaliação do desempenho em termos de complexidade temporal, a melhoria do tempo</p>	<p>O estudo sobre a definição de palavras semanticamente próximas na língua cazaque, utilizando métodos de similaridade semântica e algoritmos distribuídos como o Apache Spark, pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações. Ao implementar o método proposto, os repositórios podem aprimorar</p>

		<p>Spark, pode contribuir para identificar palavras semanticamente relacionadas no idioma cazaque.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>desenvolver um método distribuído que execute tarefas de comparação em paralelo, utilizando o Apache Spark, a fim de reduzir o tempo de cálculo necessário para identificar pares de palavras semanticamente próximas.</p>	<p>de computação para muitos pares de palavras por meio de uma abordagem paralelizada e a proposição de um pipeline de processamento para encontrar documentos semanticamente similares em idiomas com recursos linguísticos limitados. A pesquisa demonstra a aplicabilidade do método proposto na língua cazaque e destaca a importância da representação vetorial e da comparação distribuída de palavras para tarefas de processamento de linguagem natural.</p>	<p>a precisão na busca por documentos relacionados, facilitando a identificação de conteúdo relevante e aprimorando a experiência do usuário ao encontrar informações específicas dentro do repositório. Além disso, a abordagem distribuída pode acelerar o processo de comparação de palavras, tornando a busca mais eficiente e permitindo uma análise mais abrangente do conteúdo disponível nos repositórios institucionais.</p>
2023	End-to-End Pre-Training With Hierarchical Matching and Momentum Contrast for Text-Video Retrieval	<p>O problema de pesquisa abordado neste trabalho é a melhoria do pré-treinamento de vídeo-linguagem e da recuperação de texto-vídeo por meio do uso de correspondência hierárquica e contraste de momento. A pesquisa visa explorar informações semânticas hierárquicas em vídeos para aprimorar a compreensão semântica e a recuperação de informações entre vídeos e textos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste trabalho é apresentar uma rede de pré-treinamento de ponta a ponta com Correspondência Hierárquica e Contraste de Momento, denominada HMMC. O principal objetivo é explorar as informações semânticas hierárquicas em vídeos por meio de correspondência semântica em vários níveis entre vídeos e textos. Além disso, o trabalho visa incorporar o contraste de momento para melhorar as representações multimodais e contribuir para a generalização das representações em tarefas de recuperação de texto-vídeo.</p>	<p>Os resultados da pesquisa demonstraram a eficácia do modelo HMMC para tarefas de recuperação de vídeo, tanto em cenários de zero-shot quanto em ajustes finos, em inglês e chinês. O modelo HMMC, com sua abordagem de pré-treinamento de ponta a ponta e a combinação de quatro tarefas proxy (VTM, FTM, FAM e MLM), mostrou vantagens significativas na recuperação de texto-vídeo. Além disso, a introdução do mecanismo de atualização de momento permitiu incorporar mais amostras negativas para aprendizado contrastivo, contribuindo para a generalização das representações. Os resultados experimentais foram validados</p>	<p>O estudo sobre pré-treinamento de vídeo-linguagem com Correspondência Hierárquica e Contraste de Momento pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações multimodais, como vídeos e textos. Ao utilizar a abordagem de correspondência hierárquica para entender as conexões semânticas entre vídeos e textos, juntamente com o contraste de momento para aprimorar as representações multimodais, os repositórios institucionais podem aprimorar a eficiência e a qualidade da recuperação de conteúdo multimodal, facilitando a busca e o acesso a informações</p>

				em dois conjuntos de dados de referência para recuperação de texto-vídeo, demonstrando a eficácia e a vantagem do modelo proposto.	relevantes de forma mais precisa e abrangente.
2023	Ethio-Semitic language identification using convolutional neural networks with data augmentation	<p>O problema de pesquisa abordado neste estudo é a identificação de línguas etiosemitas, que são consideradas línguas sub-representadas e sub-recursos no contexto de tecnologias de processamento de linguagem falada. A falta de recursos linguísticos disponíveis para essas línguas dificulta o desenvolvimento de tecnologias de reconhecimento de fala e assistentes de voz que possam atender às necessidades dessas comunidades linguísticas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é desenvolver um modelo de Identificação de Linguagem para línguas etiosemitas, com foco específico nas línguas Amárico, Geez, Guragigna e Tigrigna. O modelo utiliza Redes Neurais Convolucionais (CNN) em conjunto com características acústicas como coeficientes cepstrais de frequência melódica (MFCC) e espectrograma melódico para identificar com precisão as línguas alvo. O objetivo é alcançar uma alta taxa de acurácia na identificação das línguas selecionadas, contribuindo assim para o desenvolvimento de tecnologias de processamento de linguagem para essas comunidades linguísticas sub-representadas.</p>	<p>Os resultados da pesquisa indicam que o modelo de Identificação de Linguagem desenvolvido para línguas etiosemitas, utilizando Redes Neurais Convolucionais (CNN) em conjunto com características acústicas como MFCC e espectrograma melódico, obteve um desempenho significativamente alto. O modelo alcançou uma precisão média de 97% para testes, 97.4% para validação e 99.5% para treinamento, demonstrando a eficácia da abordagem proposta. Além disso, a utilização de dados aumentados e características híbridas mostrou-se mais eficaz do que o uso isolado de MFCC ou espectrograma melódico, destacando a importância desses recursos combinados para a identificação de línguas.</p>	<p>O estudo sobre identificação de línguas etiosemitas utilizando Redes Neurais Convolucionais (CNN) e características acústicas pode ser aplicado em repositórios institucionais para aprimorar a indexação e recuperação de conteúdo multilíngue. Ao implementar o modelo de Identificação de Linguagem desenvolvido neste estudo, os repositórios institucionais podem automatizar a identificação de línguas etiosemitas em documentos de áudio, facilitando a organização e acesso a informações em diferentes idiomas. Isso pode melhorar a usabilidade e eficiência dos repositórios, especialmente para comunidades linguísticas menos representadas.</p>
2023	Evaluation of an Arabic Chatbot Based on Extractive Question-Answering Transfer Learning and Language Transformers	<p>O problema de pesquisa abordado neste estudo é a avaliação de um chatbot em árabe com base em aprendizado de transferência de perguntas e respostas extrativas e transformadores de linguagem. O objetivo é melhorar o</p>	<p>O objetivo deste estudo é desenvolver um chatbot em árabe baseado em perguntas e respostas extrativas que possa fornecer respostas precisas e eficientes a perguntas feitas em árabe, de forma semelhante a como um especialista humano</p>	<p>Os resultados da pesquisa demonstraram que a combinação do poder da arquitetura de transformadores com chatbots extrativos pode fornecer respostas mais precisas e contextualmente relevantes a perguntas em</p>	<p>O estudo sobre a avaliação de um chatbot em árabe com base em aprendizado de transferência de perguntas e respostas extrativas e transformadores de linguagem pode ser aplicado em repositórios institucionais para melhorar a interação com</p>

		<p>desempenho dos diálogos conversacionais em chatbots árabes, utilizando técnicas avançadas de processamento de linguagem natural, como transfer learning e transformers, para responder automaticamente a perguntas com base em um corpus textual fornecido.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>forneceria respostas. Além disso, o estudo visa avaliar e melhorar o desempenho dos chatbots em árabe por meio do uso de transfer learning e transformers de linguagem, explorando modelos como AraBERT, CAMeLBERT, AraElectra-SQuAD e AraElectra (Gerador/Discriminador).</p>	<p>árabe. Especificamente, o modelo AraElectra-SQuAD apresentou um desempenho notável, alta confiança e robustez, destacando seu potencial para aplicações práticas em tarefas de processamento de linguagem natural para chatbots em árabe. Os experimentos realizados mostraram que esse modelo superou consistentemente os outros modelos avaliados, com uma pontuação média de confiança de 0,6422 e uma pontuação média de similaridade de 0,9773 no primeiro conjunto de dados, e uma pontuação média de confiança de 0,6658 e uma pontuação média de similaridade de 0,9660 no segundo conjunto de dados.</p>	<p>usuários que falam árabe. Ao implementar modelos avançados de processamento de linguagem natural, como AraElectra-SQuAD, os repositórios institucionais podem oferecer respostas mais precisas e relevantes às perguntas dos usuários em árabe, melhorando a experiência do usuário e facilitando o acesso à informação. Essa abordagem pode aprimorar a eficiência na busca e recuperação de informações em repositórios institucionais, tornando a interação mais natural e eficaz para os usuários de língua árabe.</p>
2023	Evaluation of Language Models on Romanian XQuAD and RoITD datasets	<p>O problema de pesquisa abordado neste estudo é a avaliação do desempenho de diferentes modelos de linguagem em conjuntos de dados em romeno, especificamente nos conjuntos de dados XQuAD e RoITD. O estudo visa comparar o desempenho de vários modelos de NLP em tarefas de question answering em romeno e discutir as implicações desses resultados para o desenvolvimento de modelos de</p>	<p>O objetivo deste estudo é avaliar o desempenho de oito modelos de processamento de linguagem natural (NLP) em dois conjuntos de dados em romeno, XQuAD e RoITD. Os modelos são avaliados em tarefas de question answering para determinar quais modelos se saem melhor nessas tarefas específicas em romeno. Além disso, o estudo busca fornecer insights valiosos sobre o desempenho desses modelos em dados de linguagem romena e oferecer orientações para pesquisas futuras nessa área.</p>	<p>Os resultados da pesquisa indicam que os modelos bert-base-multilingual-cased e bert-base-multilingual-uncased apresentaram o melhor desempenho nas tarefas de question answering nos conjuntos de dados XQuAD e RoITD em romeno. Por outro lado, os modelos RoBERT-small e DistilBERT tiveram um desempenho inferior nessas tarefas. O estudo também destaca que o modelo RoGPT2 teve um desempenho decente, mas não superou o estado-da-</p>	<p>O estudo sobre a avaliação de modelos de linguagem em conjuntos de dados em romeno pode ser aplicado em repositórios institucionais para melhorar a eficiência e precisão de sistemas de processamento de linguagem natural em romeno. Ao utilizar os modelos de melhor desempenho identificados no estudo, as instituições podem aprimorar a capacidade de responder a perguntas em romeno, facilitando a busca e recuperação de informações em</p>

		<p>processamento de linguagem romena no futuro.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>		<p>arte. Esses resultados são consistentes com a literatura existente sobre modelos de processamento de linguagem natural e as características específicas da língua romena. Recomenda-se que trabalhos futuros busquem melhorar o desempenho desses modelos por meio de ajustes em conjuntos de dados maiores ou utilizando técnicas como aprendizado por transferência.</p>	<p>documentos nesse idioma. Isso pode beneficiar pesquisadores, estudantes e profissionais que trabalham com conteúdo em romeno, tornando a interação com repositórios institucionais mais eficaz e produtiva.</p>
2023	Evaluation of semantic relations impact in query expansion-based retrieval systems	<p>O problema de pesquisa abordado no documento é a avaliação do impacto das relações semânticas na expansão de consultas em sistemas de recuperação de informação baseados em consultas. O estudo visa quantificar o efeito de diferentes relações semânticas na classificação de consultas e determinar a melhor combinação dessas relações para aprimorar a precisão e eficácia dos sistemas de recuperação de informação.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo do estudo é propor, ajustar e avaliar um procedimento automático para gerar recursos semânticos que possam ser utilizados por sistemas de recuperação baseados em expansão de consultas. O foco da pesquisa é identificar a combinação mais eficaz dos recursos gerados e avaliar o impacto de cada combinação na associação de consultas em linguagem natural a uma classe pré-definida.</p>	<p>O resultado da pesquisa foi a identificação da combinação mais eficaz das diferentes relações semânticas para classificar uma pergunta como um rótulo de taxonomia. Os recursos semânticos gerados foram combinados de várias maneiras, resultando em diferentes configurações. Foram consideradas todas as combinações possíveis sem repetição, resultando em um número específico de configurações com base nas 11 relações semânticas presentes nos repositórios semânticos utilizados. A avaliação mostrou a importância de escolher a combinação correta de relações semânticas para melhorar a precisão e eficácia dos sistemas de recuperação de informação baseados em consultas.</p>	<p>O estudo pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações por meio da expansão de consultas com base em relações semânticas. Ao utilizar recursos semânticos gerados automaticamente e combinados de forma eficaz, os repositórios institucionais podem melhorar a classificação e associação de consultas em linguagem natural a classes específicas da taxonomia, resultando em uma recuperação mais precisa e eficiente de informações relevantes para os usuários.</p>

2023	Event Extraction With Dynamic Prefix Tuning and Relevance Retrieval	<p>O problema de pesquisa abordado neste artigo é a extração de eventos em texto sem a necessidade de anotações manuais de entidades, o que é mais prático e relevante para aplicações do mundo real. O foco está em desenvolver um método eficaz que possa extrair eventos, incluindo gatilhos e argumentos, de forma precisa e eficiente, sem depender de anotações detalhadas de entidades.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é propor um método de extração de eventos chamado GREE, que utiliza uma abordagem generativa baseada em modelos de template com ajuste dinâmico de prefixos e um framework de recuperação de relevância. O principal objetivo é melhorar a precisão e eficácia da extração de eventos, integrando informações contextuais, prefixos específicos de tipo e um mecanismo de recuperação de relevância. O estudo visa superar desafios encontrados em métodos tradicionais de extração de eventos e demonstrar a eficácia do modelo proposto em comparação com abordagens existentes.</p>	<p>Os resultados da pesquisa demonstram que o modelo proposto, GREE, alcança resultados competitivos com o estado-da-arte em modelos baseados em classificação, como o O NEIE, na tarefa de extração de eventos nos conjuntos de dados ACE 2005 e ERE. Além disso, o modelo GREE supera o desempenho de outros métodos na tarefa de extração de eventos em ERE. A pesquisa também mostra que o modelo GREE é eficaz na adaptação para extrair novos tipos de eventos de forma eficaz, demonstrando sua capacidade de aprendizado e generalização.</p>	<p>O estudo sobre extração de eventos com ajuste dinâmico de prefixos e recuperação de relevância pode ser aplicado em repositórios institucionais para melhorar a extração de informações relevantes de documentos textuais. Ao utilizar o modelo GREE, é possível aprimorar a precisão na identificação de eventos, gatilhos e argumentos, facilitando a organização e recuperação de dados em repositórios. Isso pode contribuir para uma melhor indexação, busca e análise de informações em documentos acadêmicos, relatórios e outros conteúdos textuais armazenados em repositórios institucionais.</p>
2023	FindVehicle and VehicleFinder: a NER dataset for natural language-based vehicle retrieval and a keyword-based cross-modal vehicle retrieval system	<p>A pesquisa visa superar as limitações dos métodos existentes, que geralmente utilizam técnicas de análise de dependência ou rotulagem de papéis semânticos para encontrar palavras-chave relacionadas aos atributos dos veículos, resultando em uma quantidade significativa de trabalho de pré-processamento e pós-processamento, além de possíveis erros na extração de palavras-chave em consultas de linguagem natural complexas, o problema de pesquisa central</p>	<p>O principal objetivo deste trabalho é melhorar a eficiência e a precisão dos sistemas de recuperação de veículos, tanto por meio de consultas de linguagem natural quanto por meio de consultas baseadas em palavras-chave, no contexto do domínio do tráfego. o objetivo principal é simplificar e aprimorar os sistemas de recuperação de veículos no contexto do tráfego, utilizando técnicas de NER e conjuntos de dados específicos para melhorar a precisão,</p>	<p>O estudo apresentou o conjunto de dados FindVehicle, o primeiro conjunto NER no domínio do tráfego, com informações detalhadas sobre veículos. Foi desenvolvido o sistema VehicleFinder para recuperação de veículos por comandos de texto, alcançando alta precisão e recall. Os modelos baseados em Transformer superaram os baseados em RNN, com o modelo FLERT obtendo a melhor pontuação. O FindVehicle enfrentou desafios específicos, mas mostrou</p>	<p>O estudo sobre o conjunto de dados FindVehicle e o sistema VehicleFinder pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações relacionadas a veículos em contextos de tráfego. A utilização do FindVehicle pode melhorar a precisão da identificação de entidades veiculares em textos descritivos, enquanto o VehicleFinder pode facilitar a recuperação de veículos por meio de consultas de linguagem natural. Essas ferramentas</p>

		<p>abordado neste documento é a simplificação e aprimoramento dos sistemas de recuperação de veículos por meio da aplicação de técnicas de NER e da construção de conjuntos de dados específicos no domínio do tráfego.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>eficiência e usabilidade desses sistemas.</p>	<p>eficácia na melhoria dos sistemas de recuperação de veículos.</p>	<p>podem otimizar a busca e organização de dados sobre veículos em repositórios institucionais, contribuindo para uma gestão mais eficiente e precisa das informações relacionadas ao tráfego.</p>
2023	Fuzzy Local Information C - Means based clustering and Fractional Dwarf Mongoose optimization enabled deep learning for relevant document retrieval	<p>O problema de pesquisa abordado neste estudo é a necessidade de aprimorar as técnicas de recuperação de documentos relevantes por meio da aplicação de métodos inovadores que considerem a semântica, o contexto e a interdependência das palavras nos documentos. As abordagens tradicionais de recuperação de documentos muitas vezes são limitadas em sua capacidade de compreender a complexidade dos termos e relações entre eles, o que pode resultar em resultados de busca menos precisos e relevantes. Assim, o estudo propõe um modelo que utiliza clustering baseado em FLICM e otimização DMOF em conjunto com deep learning para</p>	<p>O objetivo deste estudo é desenvolver um modelo inovador para a recuperação de documentos relevantes, que seja capaz de melhorar a precisão e a relevância dos resultados de busca. Para alcançar esse objetivo, os pesquisadores propõem a utilização de técnicas avançadas, como FLICM clustering, otimização DMOF e deep learning, para aprimorar a compreensão da semântica e do contexto dos documentos e consultas. O modelo visa superar as limitações das abordagens tradicionais de recuperação de documentos, oferecendo uma solução mais eficaz e precisa para a busca de informações relevantes.</p>	<p>O resultado da pesquisa foi a criação de um modelo de recuperação de documentos relevantes baseado em DMOF-Deep CNN, que se mostrou superior a outros modelos existentes. Este modelo alcançou uma precisão máxima de 0.854, recall de 0.913 e F1-score de 0.882, demonstrando sua eficácia na recuperação de documentos relevantes em comparação com abordagens convencionais. A integração de técnicas como FLICM clustering, otimização DMOF e deep learning permitiu melhorar significativamente a compreensão da relevância dos documentos em relação às consultas dos usuários, resultando em um desempenho</p>	<p>O estudo propõe um modelo inovador de recuperação de documentos relevantes utilizando técnicas avançadas como FLICM clustering, otimização DMOF e deep learning. Essa abordagem pode ser aplicada em repositórios institucionais para aprimorar a busca e recuperação de informações relevantes. Ao integrar essas técnicas, o modelo pode melhorar a precisão e relevância dos resultados de busca, facilitando a localização de documentos específicos e aumentando a eficiência na recuperação de informações em repositórios institucionais.</p>

		<p>aprimorar a recuperação de documentos relevantes.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>		<p>aprimorado no processo de recuperação de informações.</p>	
2023	<p>Heterogeneous data-based information retrieval using a fine-tuned pre-trained BERT language model</p>	<p>O problema de pesquisa abordado no documento é a recuperação de informações a partir de documentos escritos, especialmente lidando com a diversidade e incerteza das linguagens reais. A aplicação da recuperação de informações é desafiadora devido a consultas vagas formuladas com base em linguagens naturais, resultando em conteúdos recuperados que não correspondem às informações necessárias. Isso ocorre devido a problemas como desajuste de palavras, dificuldade em encontrar palavras-chave apropriadas em documentos extensos e distribuição de informações importantes em locais diversos nos documentos. Além disso, as relações semânticas ambíguas e complexas entre as frases tornam a extração precisa de informações necessárias por meio de consultas de usuários mais difícil.</p>	<p>O objetivo do estudo apresentado no documento é introduzir uma técnica inovadora de recuperação de informações baseada em dois modelos BERT diferentes. O primeiro modelo, chamado de BERT sintático, recebe a consulta e o documento como entrada e gera a forma incorporada tanto do documento quanto da consulta. A inclusão da camada de atenção sintática com as camadas de autoatenção mascarada e atenção tópica ajuda a identificar a relação entre os tokens, incorporar de forma mais eficaz e escolher o parágrafo mais relevante do documento em relação à consulta.</p>	<p>Os resultados da pesquisa indicam que a técnica proposta de recuperação de informações, que combina o BERT sintático e o BERT ajustado, obteve um desempenho superior em comparação com métodos convencionais. A análise baseada em várias medidas de avaliação demonstrou que a técnica de recuperação de informações proposta superou os métodos convencionais. A utilização do modelo BERT sintático para extrair o parágrafo mais informativo com base na consulta escolhe todos os parágrafos relevantes para a consulta. A incorporação de autoatenção e máscara tópica ajuda a extrair os melhores parágrafos estimando a relação entre eles. Além disso, a otimização da função de perda usando o algoritmo AIAO proposto melhorou a precisão do processo de recuperação de informações. O algoritmo AIAO</p>	<p>O estudo sobre recuperação de informações usando modelos BERT pode ser aplicado em repositórios institucionais para melhorar a eficiência e precisão na busca de informações em documentos escritos. Ao utilizar o BERT sintático e o BERT ajustado, é possível aprimorar a extração de informações relevantes com base em consultas dos usuários. Isso pode beneficiar repositórios institucionais ao tornar a recuperação de documentos mais precisa, reduzindo o tempo gasto na busca por informações específicas e melhorando a experiência do usuário ao acessar conteúdos relevantes de forma mais eficiente.</p>

		<p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>		<p>proposto possui uma taxa de convergência mais rápida, o que ajuda a minimizar a complexidade temporal do modelo de recuperação de informações proposto. O equilíbrio eficiente entre diversificação e intensificação reduziu a captura de ótimos locais e obteve a melhor solução global em menos tempo. Assim, o método proposto alcançou um resultado superior em comparação com os métodos convencionais.</p>	
2023	IKAR: An Interdisciplinary Knowledge-Based Automatic Retrieval Method from Chinese Electronic Medical Record	<p>O problema de pesquisa abordado neste documento é a dificuldade de processar e extrair informações importantes de relatórios de ultrassom em chinês na área de obstetrícia e ginecologia. Os desafios incluem a falta de estrutura padrão nos relatórios, a complexidade da língua chinesa, a ausência de conjuntos de dados rotulados em chinês disponíveis publicamente e a necessidade de identificar entidades nomeadas em textos sem espaços entre as palavras. A pesquisa visa desenvolver um método de recuperação automática de informações baseado em conhecimento interdisciplinar para lidar com esses desafios e extrair informações relevantes dos relatórios de ultrassom.</p>	<p>O objetivo deste estudo é propor e verificar experimentalmente um método de recuperação automática de informações baseado em conhecimento interdisciplinar para relatórios de ultrassom em chinês na área de obstetrícia e ginecologia. O método IKAR visa melhorar a eficácia e eficiência da extração de diagnósticos de ultrassom a partir de descrições ultrassonográficas, com o objetivo de alcançar uma precisão, recall e F-score em torno de 90%. Além disso, o estudo busca abordar os desafios específicos encontrados na tarefa de diagnóstico em chinês e melhorar o desempenho da inferência clínica.</p>	<p>O resultado da pesquisa demonstrou que o método IKAR, proposto para a recuperação automática de informações em relatórios de ultrassom em chinês na área de obstetrícia e ginecologia, obteve um desempenho significativamente bom. O método alcançou uma precisão de 89,38%, recall de 91,09% e F-score de 90,23% ao ser aplicado a um conjunto de dados hospitalares para verificação experimental. Comparado a modelos tradicionais como RNN e LSTM, o método IKAR apresentou uma melhoria geral de 1% a 8% em precisão e 1% a 3% em F-score. A análise detalhada do modelo revelou que 90% dos componentes dos relatórios de ultrassom alcançaram um F-score de 80% ou mais, e 50%</p>	<p>O estudo do método IKAR para recuperação automática de informações em relatórios de ultrassom em chinês na área de obstetrícia e ginecologia pode ser aplicado em repositórios institucionais de saúde para melhorar a recuperação de dados clínicos, padronizar relatórios, apoiar a tomada de decisão clínica e facilitar a pesquisa e análise de dados. A implementação do IKAR nos repositórios pode otimizar a interpretação de diagnósticos, permitir buscas mais eficientes e contribuir para estudos clínicos mais abrangentes e baseados em evidências.</p>

		<p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>		<p>dos componentes atingiram 90% ou mais. Esses resultados indicam que o método IKAR é eficaz na extração de diagnósticos de ultrassom em chinês e pode ser uma ferramenta útil para auxiliar no diagnóstico clínico final em ginecologia e obstetrícia.</p>	
2023	Integrating legal event and context information for Chinese similar case analysis	<p>O problema de pesquisa abordado neste artigo é a necessidade de aprimorar a análise de casos semelhantes no contexto jurídico chinês, especificamente através da integração de informações sobre eventos legais e contexto. A pesquisa visa melhorar a precisão e interpretabilidade da análise de casos semelhantes por meio do desenvolvimento do Legal Event-Context Model (LECM).</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é avaliar a capacidade do modelo Legal Event-Context (LECM) para análise de casos semelhantes (SCA) em dois tipos específicos de tarefas: Similar Case Matching (SCM) e Similar Case Retrieval (SCR). O SCM visa medir a similaridade entre documentos legais e selecionar o caso mais semelhante ao caso-alvo, enquanto o SCR tem como objetivo recuperar casos semelhantes a partir de uma base de dados. O LECM foi desenvolvido para integrar informações sobre eventos legais e contexto, a fim de aprimorar a análise de casos semelhantes no contexto jurídico chinês.</p>	<p>Os resultados da pesquisa indicaram que o Legal Event-Context Model (LECM) obteve melhorias significativas em termos de desempenho na análise de casos semelhantes (SCA) em comparação com modelos de referência. Em média, o LECM demonstrou um aumento de cerca de 13% na precisão média e 11% na acurácia para as tarefas de Similar Case Retrieval (SCR) e Similar Case Matching (SCM), respectivamente. Esses resultados destacam a eficácia do LECM na utilização do conhecimento de eventos e contexto para aprimorar o desempenho da análise de casos semelhantes e seu potencial aplicação em diversas tarefas de análise de documentos legais.</p>	<p>O estudo sobre o Legal Event-Context Model (LECM) para análise de casos semelhantes em documentos legais pode ser aplicado em repositórios institucionais para aprimorar a recuperação e análise de informações jurídicas. Ao integrar informações sobre eventos legais e contexto, o LECM pode melhorar a precisão e interpretabilidade da análise de casos semelhantes, auxiliando na identificação de documentos relevantes e na compreensão de relações jurídicas complexas. Essa abordagem pode ser útil para aprimorar a eficiência e eficácia da pesquisa e recuperação de informações em repositórios institucionais, beneficiando profissionais do direito e pesquisadores na análise de documentos legais.</p>
2023	Introduction to Mathematical Language Processing: Informal Proofs, Word	<p>O problema de pesquisa abordado neste documento é a automação da descoberta em matemática e ciência, que</p>	<p>O objetivo é desenvolver modelos de processamento de linguagem matemática capazes de lidar com relações entre</p>	<p>O resultado da pesquisa é a análise de métodos de processamento de linguagem matemática em cinco subáreas</p>	<p>O estudo sobre processamento de linguagem matemática pode ser aplicado em repositórios institucionais para melhorar a</p>

	<p>Problems, and Supporting Tasks</p>	<p>requer métodos sofisticados de extração de informações e raciocínio abstrato. O objetivo é desenvolver modelos capazes de processar de forma convincente as relações entre elementos matemáticos e linguagem natural, a fim de produzir soluções para problemas de valor no mundo real.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>	<p>elementos matemáticos e linguagem natural, permitindo a produção de soluções para problemas do mundo real de forma automatizada e explicável. Esses modelos devem ser capazes de raciocinar sobre elementos matemáticos, como equações, expressões e variáveis, e também sobre texto em linguagem natural, realizando tarefas como seleção de premissas, extração de definições de identificadores e prova informal de teoremas.</p>	<p>estratégicas: extração de definições de identificadores, recuperação de fórmulas, seleção de premissas em linguagem natural, resolução de problemas matemáticos em palavras e prova informal de teoremas. O estudo destaca metodologias predominantes, limitações existentes, tendências gerais e direções promissoras para pesquisas futuras.</p>	<p>extração de informações e o raciocínio abstrato em documentos acadêmicos. Isso pode facilitar a identificação de definições de identificadores, recuperação de fórmulas relevantes, seleção de premissas importantes e resolução de problemas matemáticos em palavras. Essas aplicações podem aprimorar a organização e a acessibilidade de conteúdo matemático em repositórios institucionais, tornando o conhecimento mais acessível e útil para pesquisadores e estudantes.</p>
2023	<p>Learned Text Representation for Amharic Information Retrieval and Natural Language Processing</p>	<p>O problema de pesquisa abordado neste estudo é a falta de modelos de linguagem pré-treinados e acessíveis publicamente para a língua amárica, que é uma língua sub-representada. Isso resulta na falta de investigação sobre o impacto da representação de texto aprendida em tarefas de Processamento de Linguagem Natural (NLP) para o amárico. O estudo visa explorar o impacto de modelos de embedding de palavras (word2vec, GloVe, fastText) e modelos BERT no texto amárico, bem como investigar a eficácia desses</p>	<p>O principal objetivo deste estudo é abordar a lacuna existente em relação à falta de modelos de linguagem pré-treinados e acessíveis publicamente para a língua amárica, que é uma língua sub-representada. Para atender a essa necessidade, os pesquisadores propuseram a construção de modelos de linguagem pré-treinados específicos para o amárico e disponibilizá-los para a comunidade de pesquisa. Além disso, eles planejaram realizar o ajuste fino desses modelos pré-treinados para tarefas de Processamento de Linguagem Natural (NLP) e Recuperação de</p>	<p>Os resultados da pesquisa indicaram que a representação de texto aprendida usando uma abordagem baseada em palavras superou as abordagens baseadas em raízes e em palavras de superfície em todas as tarefas consideradas no estudo. Modelos como BERT e fastText apresentaram resultados promissores. Além disso, os pesquisadores construíram modelos pré-treinados específicos para o amárico e os ajustaram para tarefas de classificação de texto, obtendo melhorias nas métricas de desempenho ao longo das épocas de treinamento. os</p>	<p>O estudo sobre representação de texto aprendida para o amárico pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações e o processamento de documentos nessa língua sub-representada. A utilização dos modelos pré-treinados de embedding de palavras e BERT específicos para o amárico pode melhorar a precisão na busca e classificação de documentos, facilitando o acesso a informações relevantes. Além disso, a expansão de consulta baseada em palavras pode enriquecer as consultas realizadas nos repositórios,</p>

		<p>modelos em tarefas de IR e NLP para o amárico.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>Informação (IR) relacionadas ao amárico. os objetivos deste estudo visam preencher a lacuna de recursos de linguagem para o amárico, investigar o impacto das representações de texto aprendidas em tarefas de NLP e IR para o amárico e fornecer uma base sólida para futuras pesquisas nessa área.</p>	<p>resultados da pesquisa mostraram a viabilidade e a eficácia das representações de texto aprendidas para o amárico, bem como a importância de disponibilizar recursos e modelos pré-treinados para impulsionar o desenvolvimento de aplicações de Processamento de Linguagem Natural e Recuperação de Informação para essa língua sub-representada.</p>	<p>aumentando a eficiência na recuperação de documentos.</p>
2023	<p>Learning bilingual word embedding for automatic text summarization in low resource language</p>	<p>O problema de pesquisa abordado neste estudo é a melhoria da capacidade de resumo de texto automático em idiomas de baixo recurso, como o indonésio, por meio do aprendizado de incorporação de palavras bilíngues. Os métodos VecMap e BiVec são utilizados para criar incorporações de palavras bilíngues entre o indonésio e o inglês, visando superar as limitações de recursos linguísticos nesses idiomas e melhorar a qualidade do resumo automático de texto.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é investigar e comparar dois métodos de aprendizado de incorporação de palavras bilíngues, VecMap e BiVec, para os idiomas indonésio e inglês. Esses métodos são avaliados quanto à sua eficácia na indução de léxico bilíngue e na tarefa de resumo de texto, com o intuito de melhorar a capacidade de resumo automático em idiomas de baixo recurso. Além disso, o estudo busca explorar a aplicação de transferência de aprendizado entre idiomas para superar as disparidades de recursos linguísticos entre idiomas de alto e baixo recurso.</p>	<p>Os resultados da pesquisa indicam que o método BiVec superou o método VecMap em termos de desempenho na indução de léxico bilíngue e na tarefa de resumo de texto. Ao ser treinado diretamente no corpus paralelo, o BiVec demonstrou melhores resultados, especialmente quando combinado com a recuperação CSLS. Esses resultados sugerem que o método BiVec, com sua abordagem conjunta, é mais eficaz na geração de incorporações de palavras bilíngues e na melhoria da capacidade de resumo automático de texto em comparação com o método VecMap. Além disso, foi observado que tanto o VecMap quanto o BiVec ainda enfrentam problemas de "hubness" nas incorporações de palavras bilíngues entre o inglês e o</p>	<p>O estudo sobre aprendizado de incorporação de palavras bilíngues para resumo automático de texto em idiomas de baixo recurso, como o indonésio, pode ser aplicado em repositórios institucionais para melhorar a capacidade de resumo automático em diferentes idiomas. A utilização dos métodos VecMap e BiVec pode ajudar a aprimorar a qualidade dos resumos gerados automaticamente, tornando as informações mais acessíveis e compreensíveis para os usuários dos repositórios. Além disso, a aplicação de transferência de aprendizado entre idiomas pode facilitar a tradução e resumo de conteúdos em idiomas menos comuns, ampliando o alcance e a utilidade dos repositórios institucionais.</p>

				indonésio, e a utilização do método CSLS pode ajudar a mitigar esse problema.	
2023	Linguistically aware evaluation of coreference resolution from the perspective of higher-level applications	<p>O problema de pesquisa abordado neste artigo é a avaliação de resolução de correferência a partir de uma perspectiva linguística e de aplicações de nível superior. O estudo busca melhorar a avaliação da resolução de correferência, levando em consideração não apenas métricas tradicionais, como precisão e revocação, mas também aspectos linguísticos e semânticos mais complexos. A pesquisa visa aprimorar a compreensão e a avaliação da resolução de correferência para impulsionar o desenvolvimento de aplicações de processamento de linguagem natural mais avançadas e precisas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é apresentar uma nova estratégia de avaliação da resolução de correferência que leve em consideração informações linguísticas e semânticas, visando superar as limitações das métricas existentes. A proposta é desenvolver um modelo de avaliação que seja facilmente expansível e capaz de lidar com novos tipos de erros que possam surgir em diferentes aplicações de nível superior. Além disso, o estudo busca demonstrar a importância da resolução de correferência para a qualidade de serviços de busca de informações implementados por aplicações de processamento de linguagem natural.</p>	<p>O resultado da pesquisa foi a proposição de uma nova estratégia de avaliação da resolução de correferência que se concentra em informações linguísticas e semânticas, visando abordar algumas das limitações das métricas existentes. A avaliação foi realizada no contexto do desenvolvimento de capacidades de resolução de correferência para a língua lituana, utilizando recursos linguísticos específicos desse idioma. A estratégia de avaliação proposta não é dependente de idioma, o que a torna aplicável em diferentes contextos linguísticos.</p>	<p>O estudo sobre a avaliação linguística da resolução de correferência pode ser aplicado em repositórios institucionais para aprimorar a qualidade e precisão da busca de informações. Ao utilizar a nova estratégia de avaliação proposta, os repositórios podem melhorar a resolução de correferência em documentos e metadados, facilitando a recuperação de informações relevantes para os usuários. Isso contribui para uma melhor organização e acessibilidade dos conteúdos nos repositórios, tornando a busca mais eficiente e precisa para os usuários em busca de informações específicas.</p>
2023	Logic-Based Inference With Phrase Abduction Using Vision-and-Language Models	<p>O problema de pesquisa abordado neste estudo é a falta de conhecimento frasal necessário para resolver problemas de Reconhecimento</p>	<p>O objetivo deste estudo é melhorar os sistemas de inferência lógica ao abordar a falta de conhecimento frasal necessário para resolver</p>	<p>Os resultados da pesquisa indicam que o sistema de inferência lógica proposto, que utiliza conhecimento frasal e compara representações visuais</p>	<p>O estudo sobre Inferência Lógica com Abdução de Frases Usando Modelos de Visão e Linguagem pode ser aplicado em repositórios institucionais</p>

		<p>de Entendimento Textual (RTE). A pesquisa investiga como as representações visuais de frases podem ser utilizadas para capturar relações de implicação entre textos e melhorar o desempenho de sistemas de inferência lógica em tarefas de RTE.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>	<p>problemas de Reconhecimento de Entendimento Textual (RTE). O estudo propõe comparar as representações visuais de frases para determinar se podem capturar relações de implicação entre as frases e, assim, aprimorar o desempenho dos sistemas de inferência lógica em tarefas de RTE.</p>	<p>de frases, melhorou a precisão em comparação com um sistema de inferência lógica anterior. Além disso, as funções de avaliação assimétricas usando modelos de visão e linguagem foram eficazes na captura de relações de implicação de pares de palavras. As representações de frases geradas pelos modelos de visão e linguagem DALL-E e CLIP foram capazes de abordar a falta de conhecimento frasal em tarefas de RTE, demonstrando a viabilidade e eficácia dessa abordagem.</p>	<p>para aprimorar a compreensão e a análise de textos complexos. A utilização de representações visuais de frases e a injeção de conhecimento frasal em sistemas de inferência lógica podem melhorar a precisão na identificação de relações de implicação entre textos. Isso pode beneficiar repositórios institucionais ao facilitar a recuperação de informações relevantes e a análise de conteúdo de forma mais eficaz e precisa.</p>
2023	Many Hands Make Light Work: Transferring Knowledge From Auxiliary Tasks for Video-Text Retrieval	<p>O problema de pesquisa abordado neste artigo é a recuperação de informações em vídeos por meio de descrições em linguagem natural, ou vice-versa, o que é conhecido como recuperação de vídeo-texto. Este é um desafio significativo devido à grande quantidade de vídeos produzidos diariamente e à necessidade de encontrar maneiras eficazes de buscar e relacionar informações em diferentes modalidades de dados, como texto e vídeo.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade</p>	<p>O objetivo deste estudo é propor um método inovador baseado em redes neurais gráficas (GNN) para a recuperação de vídeo-texto, que progressivamente explora conhecimento útil de tarefas auxiliares de forma geral para específica, a fim de auxiliar o processo principal de recuperação de informações. O modelo proposto visa aprender com tarefas auxiliares em um framework de GNN, transferindo conhecimento útil de outras tarefas relevantes de recuperação de informações de forma estruturada e eficaz.</p>	<p>Os resultados da pesquisa demonstraram que o modelo proposto superou os métodos do estado-da-arte em quatro conjuntos de dados de referência para recuperação de vídeo-texto. O uso de redes neurais gráficas para extrair similaridades estruturais entre amostras de diferentes modalidades e transferir conhecimento útil de tarefas auxiliares em uma abordagem geral para específica mostrou-se eficaz na melhoria do desempenho de recuperação de informações em vídeos com base em texto e vice-versa. Além disso, o modelo proposto apresentou uma complexidade e eficiência favoráveis em termos</p>	<p>O estudo propõe um método inovador baseado em redes neurais gráficas para melhorar a recuperação de informações em vídeos com base em texto e vice-versa. Esse modelo pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de conteúdo multimodal, permitindo uma busca mais eficaz e precisa de informações em diferentes formatos, como vídeos e textos. Ao utilizar a transferência de conhecimento de tarefas auxiliares de forma geral para específica, o modelo pode melhorar a relevância e a eficiência da recuperação de informações, beneficiando a organização e acesso aos</p>

		percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA		de número de parâmetros, requisitos de memória e tempo de inferência em comparação com outros métodos do estado-da-arte.	recursos nos repositórios institucionais.
2023	MatchACNN: A Multi-Granularity Deep Matching Model	<p>O problema de pesquisa abordado no artigo "MatchACNN: A Multi-Granularity Deep Matching Model" é a melhoria do desempenho em tarefas de recuperação de informações por meio do desenvolvimento de um modelo de correspondência de texto mais eficaz e preciso. O foco principal é aprimorar a capacidade de correspondência de texto em diferentes granularidades para lidar com consultas de pesquisa e documentos de maneira mais eficiente.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo do estudo apresentado no artigo "MatchACNN: A Multi-Granularity Deep Matching Model" é propor um modelo de correspondência de texto de várias granularidades (MatchACNN) que aborde a falta de correspondência de várias granularidades nos modelos de correspondência de profundidade existentes. O modelo proposto visa extrair características de diferentes dimensões, tratando a correspondência de texto como reconhecimento de imagem e empregando uma rede neural convolucional bidimensional e um mecanismo de atenção no reconhecimento de imagem. Além disso, o estudo busca demonstrar experimentalmente a superioridade do modelo proposto em relação a outros métodos de aprendizado profundo existentes.</p>	<p>Os resultados da pesquisa demonstraram que o modelo MatchACNN superou significativamente os métodos de aprendizado profundo existentes em tarefas de correspondência de texto. Os experimentos realizados nos conjuntos de dados Wiki QA Corpus, NFCorpus e TREC QA mostraram que o MatchACNN obteve um desempenho superior em comparação com os métodos de aprendizado profundo tradicionais. Além disso, o modelo proposto conseguiu lidar com diferentes granularidades de texto de forma mais eficaz, melhorando a precisão da correspondência e superando os modelos existentes em termos de velocidade de execução.</p>	<p>O estudo sobre o modelo MatchACNN para correspondência de texto em diferentes granularidades pode ser aplicado em repositórios institucionais para melhorar a eficiência e precisão da recuperação de informações. Ao utilizar esse modelo, os repositórios podem aprimorar a correspondência entre consultas de usuários e documentos armazenados, resultando em uma melhor experiência de busca. Além disso, a capacidade do MatchACNN de extrair características de diferentes dimensões pode ajudar a identificar informações relevantes em documentos de forma mais eficaz, facilitando a recuperação de conteúdo específico nos repositórios institucionais.</p>
2023	MIRACL: A Multilingual Retrieval Dataset Covering 18 Diverse Languages	<p>O problema de pesquisa abordado no artigo sobre o conjunto de dados MIRACL é a melhoria do acesso à informação para populações diversas em todo o mundo, por</p>	<p>O objetivo do conjunto de dados MIRACL é fornecer um recurso de alta qualidade para treinamento e avaliação de modelos de recuperação de informações em diversos</p>	<p>O resultado da pesquisa resultou na criação do conjunto de dados MIRACL, um recurso multilíngue para recuperação de informações em 18 idiomas diferentes. O conjunto de dados</p>	<p>O estudo sobre o conjunto de dados MIRACL pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações em diferentes idiomas. Ao utilizar o</p>

		<p>meio de um conjunto de dados de recuperação multilíngue que cobre 18 idiomas diversos. O foco está em fornecer suporte para tarefas de recuperação monolíngue, com o objetivo de aprimorar as capacidades de acesso à informação para uma ampla gama de usuários.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>idiomas. Ele visa acelerar o progresso na pesquisa de recuperação multilíngue, oferecendo um conjunto de dados abrangente e robustamente anotado que pode servir como base para o desenvolvimento e avaliação de modelos de recuperação em diferentes idiomas. Além disso, o MIRACL busca responder a questões científicas interessantes sobre transferência interlingual no contexto da recuperação multilíngue.</p>	<p>contém mais de 726 mil julgamentos de relevância de alta qualidade para 78 mil consultas sobre artigos da Wikipedia em vários idiomas. Esses dados foram anotados por falantes nativos contratados pela equipe de pesquisa. O MIRACL abrange idiomas que são tanto tipologicamente próximos quanto distantes, pertencentes a 10 famílias de idiomas e 13 subfamílias, e está disponível para apoiar pesquisas futuras na área de recuperação de informações multilíngue.</p>	<p>MIRACL para treinar e avaliar modelos de recuperação de informações, os repositórios institucionais podem aprimorar a precisão e eficácia de suas ferramentas de busca, permitindo que usuários encontrem informações relevantes em uma variedade de idiomas. Isso pode beneficiar pesquisadores, estudantes e profissionais que acessam e contribuem para repositórios institucionais, facilitando a descoberta de conteúdo relevante em um contexto multilíngue.</p>
2023	Modified Seagull Optimization With Deep Learning for Affect Classification in Arabic Tweets	<p>O problema de pesquisa abordado neste estudo é a classificação de afetos ou emoções em tweets em árabe. A detecção e categorização de emoções em textos em árabe representam um desafio devido à complexidade da linguagem e à necessidade de desenvolver métodos precisos de mineração de emoções. O estudo visa desenvolver uma técnica chamada Modified Seagull Optimization with Deep Learning based Affect Classification on Arabic Tweets (MSGODL-ACAT) para lidar com esse problema.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade</p>	<p>O objetivo deste estudo é desenvolver e avaliar a técnica MSGODL-ACAT para a classificação automatizada de afetos em tweets em árabe. A técnica proposta visa reconhecer e categorizar os efeitos ou emoções presentes nos tweets em árabe. Para atingir esse objetivo, a técnica utiliza a otimização do algoritmo Seagull (MSGO) para ajuste dos hiperparâmetros do modelo de rede neural profunda (DBN), melhorando assim os resultados de classificação.</p>	<p>Os resultados da pesquisa indicam que a abordagem MSGODL-ACAT demonstrou um desempenho eficaz em relação a outras técnicas atuais de classificação de afetos em tweets em árabe. A avaliação experimental da técnica utilizando bancos de dados de tweets em árabe mostrou resultados positivos, destacando a eficácia do algoritmo MSGODL-ACAT. Além disso, a técnica obteve resultados promissores em termos de precisão, recall e F-score na categorização de diferentes tipos de emoções presentes nos tweets em árabe.</p>	<p>O estudo sobre a classificação de afetos em tweets em árabe utilizando a técnica MSGODL-ACAT pode ser aplicado em repositórios institucionais para aprimorar a análise de sentimentos em textos em árabe. Ao implementar essa técnica em repositórios, é possível automatizar a categorização de emoções em conteúdos em árabe, facilitando a organização e recuperação de informações relevantes. Isso contribui para uma melhor compreensão do sentimento expresso em textos em árabe, beneficiando áreas como pesquisa acadêmica, análise de dados e tomada de decisões baseada em sentimentos.</p>

		percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA			
2023	Multi-level knowledge-driven feature representation and triplet loss optimization network for image–text retrieval	<p>O problema de pesquisa abordado neste estudo é a otimização da recuperação de informações entre imagens e textos por meio do desenvolvimento de um novo modelo, o Multi-level Knowledge-driven feature representation and Triplet Loss Optimization Network (MKTLON). O objetivo é melhorar a representação de características visuais e textuais, incorporando conhecimento prévio e adotando uma estratégia de margem adaptativa na perda de tripletos, a fim de aprimorar a precisão da recuperação de informações entre imagens e textos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>	<p>O estudo aborda a otimização da recuperação de informações entre imagens e textos por meio do desenvolvimento do Multi-level Knowledge-driven feature representation and Triplet Loss Optimization Network (MKTLON). O objetivo é aprimorar a representação de características visuais e textuais, incorporando conhecimento prévio e adotando uma estratégia de margem adaptativa na perda de tripletos para melhorar a precisão da recuperação de informações entre imagens e textos. O MKTLON utiliza o mecanismo de autoatenção aprimorado (ERSA) para melhorar as capacidades de representação de características de imagens e textos, introduzindo conhecimento prévio. Além disso, o modelo proposto é avaliado em duas bases de dados públicas, Flickr30K e MSCOCO, demonstrando melhorias significativas em relação a abordagens avançadas existentes.</p>	<p>Os resultados da pesquisa demonstraram que o modelo proposto, o Multi-level Knowledge-driven feature representation and Triplet Loss Optimization Network (MKTLON), superou as abordagens avançadas existentes em termos de métricas de avaliação, como o <i>rSum</i>. Em experimentos realizados nas bases de dados públicas Flickr30K e MSCOCO, o MKTLON obteve melhorias significativas de 5.7% e 4.3% em comparação com o estado-da-arte. Esses resultados destacam a eficácia do uso de conhecimento prévio e da estratégia de margem adaptativa na otimização da recuperação de informações entre imagens e textos. Além disso, a análise qualitativa dos resultados também confirmou a eficácia do modelo proposto, mesmo em casos em que os exemplos recuperados incorretamente apresentavam alta similaridade com a consulta original.</p>	<p>O estudo propõe um modelo inovador, o Multi-level Knowledge-driven feature representation and Triplet Loss Optimization Network (MKTLON), que melhora a recuperação de informações entre imagens e textos. Esse modelo pode ser aplicado em repositórios institucionais para aprimorar a precisão da busca e recuperação de conteúdo multimodal, como imagens e textos. Ao incorporar conhecimento prévio e adotar uma estratégia de margem adaptativa na otimização da rede, o MKTLON pode melhorar a representação de características visuais e textuais, resultando em resultados de busca mais relevantes e precisos para os usuários dos repositórios institucionais.</p>
2023	Multi-Modal Transformer With Global-Local	<p>O problema de pesquisa abordado neste estudo é a recuperação de imagens com</p>	<p>O objetivo deste estudo é desenvolver um framework chamado ComqueryFormer que</p>	<p>Os resultados da pesquisa demonstraram que o framework ComqueryFormer superou os</p>	<p>O estudo sobre o framework ComqueryFormer, que utiliza um mecanismo de alinhamento</p>

	Alignment for Composed Query Image Retrieval	<p>base em consultas compostas, que consistem em uma imagem de referência e um texto de modificação. O desafio principal é desenvolver um sistema eficaz que possa compreender e processar adequadamente essas consultas compostas, permitindo a recuperação precisa de imagens relevantes que atendam às intenções de busca dos usuários.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>utiliza um mecanismo eficaz de alinhamento global-local para melhorar a recuperação de imagens com base em consultas compostas. O objetivo é aprimorar a interação entre informações visuais e textuais, permitindo uma representação conjunta mais precisa de imagens e texto para facilitar a recuperação de imagens relevantes em resposta a consultas compostas. Além disso, o estudo visa superar as limitações de métodos tradicionais e alcançar resultados superiores em termos de precisão e eficácia na recuperação de imagens.</p>	<p>resultados de estado-da-arte em três conjuntos de dados públicos: FashionIQ, CIRR e Fashion200K. O mecanismo global-local de alinhamento proposto mostrou-se eficaz na redução da distância entre a consulta composta e a imagem-alvo, resultando em melhorias significativas na precisão da recuperação de imagens. Além disso, o ComqueryFormer foi capaz de lidar com consultas compostas complexas e diversificadas, obtendo melhorias substanciais em comparação com métodos tradicionais, especialmente em cenários desafiadores onde as consultas envolvem mudanças detalhadas e sutis entre modalidades de imagem e texto.</p>	<p>global-local para recuperação de imagens com base em consultas compostas, pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de conteúdo visual com base em consultas que combinam imagens de referência e texto descritivo. Essa abordagem pode melhorar a precisão e relevância dos resultados de pesquisa, permitindo aos usuários encontrar imagens específicas com maior facilidade e eficiência. Além disso, a integração do ComqueryFormer em repositórios institucionais pode facilitar a descoberta de conteúdo visual relevante para pesquisas acadêmicas, apresentações e outros fins educacionais e profissionais.</p>
2023	Natural Language Processing with Optimal Deep Learning-Enabled Intelligent Image Captioning System	<p>O problema de pesquisa abordado no documento é a geração automatizada de legendas para imagens, conhecida como image captioning. Especificamente, o estudo se concentra em desenvolver um sistema de geração de legendas de imagens inteligente, habilitado por aprendizado profundo, que seja otimizado para fornecer descrições precisas e significativas para imagens. O objetivo é aprimorar a capacidade do sistema de</p>	<p>O objetivo do estudo é desenvolver um sistema de Geração de Legendas de Imagens Inteligente Habilitado por Aprendizado Profundo, que seja capaz de produzir descrições precisas e significativas para imagens de forma automatizada. Esse sistema, denominado NLPODL-IICS, visa aprimorar a capacidade de compreensão do conteúdo visual das imagens e gerar legendas de alta qualidade por meio da combinação de técnicas avançadas de</p>	<p>O estudo resultou no desenvolvimento de um modelo chamado Natural Language Processing with Optimal Deep Learning-Enabled Intelligent Image Captioning System (NLPODL-IICS), que utiliza técnicas avançadas de processamento de linguagem natural e redes neurais profundas para gerar descrições precisas e significativas para imagens de forma automatizada. Esse modelo foi validado experimentalmente com sucesso em dois conjuntos de</p>	<p>O estudo sobre o Natural Language Processing with Optimal Deep Learning-Enabled Intelligent Image Captioning System apresenta um modelo avançado para geração automatizada de legendas de imagens. Esse sistema, NLPODL-IICS, combina técnicas de processamento de linguagem natural e redes neurais profundas para produzir descrições precisas e significativas para imagens. A aplicação desse modelo em repositórios institucionais pode</p>

		<p>compreender o conteúdo visual das imagens e gerar legendas de alta qualidade de forma automatizada, utilizando técnicas avançadas de processamento de linguagem natural e redes neurais profundas.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>processamento de linguagem natural e redes neurais profundas. O foco principal é melhorar a precisão e a qualidade das legendas geradas para imagens, contribuindo para avanços na área de image captioning.</p>	<p>dados de referência, demonstrando um desempenho superior em comparação com outros modelos existentes. A pesquisa contribui para avanços na área de image captioning, melhorando a qualidade e a precisão das legendas geradas para imagens.</p>	<p>melhorar a indexação e recuperação de informações visuais, tornando o conteúdo mais acessível e compreensível para os usuários. Além disso, a geração automatizada de legendas pode facilitar a organização e categorização de imagens em repositórios, contribuindo para uma melhor experiência de busca e navegação.</p>
2023	NCC: Neural concept compression for multilingual document recommendation	<p>O problema de pesquisa abordado no artigo "NCC: Neural concept compression for multilingual document recommendation" é a melhoria da qualidade de sistemas de recomendação de documentos multilíngues e de recuperação de informações, criando uma representação independente de idioma. O foco principal é utilizar mapeamentos para alinhar espaços de representação monolíngue, usando compressão de conceitos, a fim de criar representações interlinguais.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Informação</p>	<p>O objetivo do artigo "NCC: Neural concept compression for multilingual document recommendation" é propor um método inovador para gerar representações de documentos interlinguais usando compressão de conceitos de redes neurais. O principal intuito é melhorar a qualidade de sistemas de recomendação de documentos multilíngues e de recuperação de informações, criando representações independentes de idioma e superando abordagens tradicionais de recuperação e recomendação cruzada de idiomas.</p>	<p>Os resultados da pesquisa demonstraram que o método proposto, chamado Neural Concept Compression (NCC), superou abordagens tradicionais de recuperação e recomendação cruzada de idiomas em experimentos realizados nos conjuntos de dados JRC-Acquis e EU Bookshop. Os experimentos mostraram que o NCC obteve melhores taxas de recuperação de documentos e classificação recíproca em comparação com outras abordagens, como LSI, LCA, LCC e NCA, tanto no conjunto de dados JRC-Acquis quanto no EU Bookshop.</p>	<p>O estudo sobre Neural Concept Compression (NCC) para recomendação de documentos multilíngues pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações e recomendações de documentos em diferentes idiomas. Ao utilizar a compressão de conceitos neurais para criar representações interlinguais, os repositórios institucionais podem oferecer aos usuários uma experiência de busca mais eficaz e personalizada, independentemente do idioma em que os documentos estão disponíveis. Isso pode melhorar a acessibilidade e a relevância dos documentos disponíveis nos repositórios, facilitando a descoberta de informações importantes por parte dos usuários.</p>

2023	NLP-based approach for automated safety requirements information retrieval from project documents	<p>O problema de pesquisa abordado neste estudo é a necessidade de desenvolver uma abordagem eficaz para recuperar requisitos de segurança de forma automatizada a partir de documentos de projeto. A diversidade e irregularidade dos requisitos de segurança apresentam desafios na recuperação de informações, e a precisão na identificação desses requisitos é crucial para o sucesso da prática de gestão de segurança. Portanto, o problema central deste estudo é propor uma abordagem que apoie a recuperação de requisitos de segurança a partir de documentos de projeto, visando aprimorar a eficácia das decisões de gestão de segurança.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é desenvolver e implementar um framework baseado em processamento de linguagem natural (NLP) para a mineração automatizada de requisitos de segurança especificados por gerentes de projeto em documentos de projeto. O foco principal é melhorar a capacidade de identificar e recuperar com precisão os requisitos de segurança relevantes, a fim de apoiar a tomada de decisões em gestão de segurança durante a fase de construção. Além disso, o estudo visa demonstrar a eficácia do framework proposto por meio de experimentos práticos e validar sua utilidade na prática de gestão de segurança em projetos de construção.</p>	<p>Os resultados da pesquisa demonstraram a eficácia do framework proposto, o RRDA (Requirement Retrieval and Document Association), na recuperação automatizada de requisitos de segurança a partir de documentos de projeto. A abordagem baseada em processamento de linguagem natural (NLP) permitiu a identificação precisa e eficiente dos requisitos de segurança especificados pelos gerentes de projeto, superando desafios como a diversidade e irregularidade na expressão desses requisitos em linguagem natural. Além disso, os experimentos práticos realizados para validar o framework mostraram que a associação de documentos relevantes aos requisitos de segurança foi bem-sucedida, contribuindo para a tomada de decisões mais informadas em gestão de segurança durante a fase de construção. A aplicação do RRDA proporcionou uma abordagem inovadora e eficaz para a recuperação de informações críticas de segurança, auxiliando na melhoria do desempenho e na correção de desvios de execução em práticas de segurança na construção.</p>	<p>O estudo apresenta um framework baseado em processamento de linguagem natural (NLP) para a recuperação automatizada de requisitos de segurança de documentos de projeto. Essa abordagem pode ser aplicada em repositórios institucionais para facilitar a identificação e recuperação eficiente de informações críticas de segurança especificadas por gerentes de projeto. Ao implementar o framework RRDA (Requirement Retrieval and Document Association), os repositórios institucionais podem aprimorar a gestão de segurança, tomar decisões mais informadas e corrigir desvios de execução em práticas de segurança na construção.</p>
------	---	--	---	--	--

2023	OIE4PA: open information extraction for the public administration	<p>O problema de pesquisa abordado no artigo "OIE4PA: open information extraction for the public administration" é a extração de informações e relações relevantes de documentos de licitações no setor de Administração Pública. O foco está em desenvolver uma ferramenta de extração de informações aberta e multilíngue para lidar com a vasta quantidade de documentos não estruturados e digitalizados presentes no setor público. A equipe de pesquisa busca melhorar a eficiência na extração e classificação de relações para facilitar a análise e o uso dessas informações nos processos de licitação e gestão pública.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O principal objetivo do trabalho é preencher as lacunas existentes na disponibilidade de soluções para lidar com domínios restritos, como o setor público, e idiomas diferentes do inglês. Além disso, a pesquisa visa explorar o impacto de um sistema de Extração de Informações Abertas (Open IE) em um domínio específico, como o setor de compras públicas. O foco está em desenvolver e avaliar uma ferramenta de Open IE (OIE4PA) para extrair e classificar relações de documentos de licitações, visando melhorar a eficiência na recuperação de informações relevantes para profissionais da Administração Pública.</p>	<p>O estudo avaliou o desempenho do OIE4PA, uma ferramenta de Extração de Informações Abertas para licitações, encontrando resultados promissores em testes controlados e práticos. A metodologia combina algoritmos tradicionais de classificação com aprendizado automático para lidar com a escassez de dados. O estudo forneceu um conjunto de dados anotados manualmente de licitações italianas, embora destaque a necessidade de aumentar seu tamanho para implementar técnicas de aprendizado profundo. Um desafio identificado é a falta de conhecimento prévio sobre as relações a serem extraídas no contexto das licitações, destacando a importância do feedback dos usuários para melhorar o sistema de busca.</p>	<p>O estudo sobre OIE4PA, uma ferramenta de Extração de Informações Abertas para licitações na Administração Pública, pode ser aplicado em repositórios institucionais para melhorar a extração e classificação de informações relevantes de documentos. Ao utilizar técnicas de Open IE e aprendizado supervisionado com autoaprendizagem, é possível automatizar a identificação de relações em textos não estruturados, facilitando a organização e recuperação de dados em repositórios. Isso pode aumentar a eficiência na análise de documentos, auxiliando na tomada de decisões e na gestão de informações em ambientes institucionais.</p>
2023	Ontology Learning Applications of Knowledge Base Construction for Microelectronic Systems Information	<p>O problema de pesquisa abordado neste artigo é a construção de bases de conhecimento para sistemas microeletrônicos, utilizando técnicas de aprendizado de ontologia. O foco está em como a inteligência artificial pode ser aplicada para extrair e organizar informações relevantes nesse domínio, visando facilitar a</p>	<p>O objetivo deste estudo é explorar a aplicação de técnicas de aprendizado de ontologia na construção de bases de conhecimento para informações de sistemas microeletrônicos. O objetivo principal é automatizar o processo de construção de bases de conhecimento a partir de artigos especializados, manuais e textos relacionados</p>	<p>Os resultados apresentados neste estudo demonstram a viabilidade e eficácia do sistema proposto na construção de bases de conhecimento a partir de textos do domínio microeletrônico. A abordagem de utilizar uma rede neural LSTM bidirecional para reconhecimento de entidades nomeadas e a aplicação de uma</p>	<p>O estudo sobre a construção de bases de conhecimento para sistemas microeletrônicos utilizando aprendizado de ontologia pode ser aplicado em repositórios institucionais para melhorar a organização e recuperação de informações técnicas. A utilização de técnicas de reconhecimento de entidades nomeadas e a criação de</p>

		<p>compreensão e o uso dessas informações de forma eficiente e precisa.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>aos domínios automotivo e microeletrônico, utilizando uma rede neural LSTM bidirecional. O estudo visa extrair informações essenciais e classificar entidades nomeadas para criar bases de conhecimento estruturadas e compreensíveis.</p>	<p>ontologia básica GENIAL! foram eficazes na geração de bases de conhecimento a partir de documentos eletrônicos. A integração de processos de extração de conhecimento e representação em forma de grafos proporcionou uma compreensão técnica e abrangente dos textos especializados pelos usuários. Além disso, o estudo destaca a importância de criar conjuntos de dados específicos e a relevância de abordagens centradas em definições para a construção de ontologias em contextos como o ISO 26262.</p>	<p>ontologias específicas para classificação de dados podem facilitar a indexação e busca de documentos em repositórios, tornando o acesso a informações mais eficiente e preciso. Além disso, a geração de grafos de conhecimento a partir dos textos pode auxiliar na visualização e compreensão das relações entre os dados armazenados nos repositórios institucionais, proporcionando uma experiência mais enriquecedora para os usuários que buscam informações técnicas nesses ambientes.</p>
2023	<p>Problematic Unordered Queries in Temporal Moment Measurement by Using Natural Language</p>	<p>O problema de pesquisa abordado no artigo é a dificuldade de medir momentos temporais em vídeos não editados usando linguagem natural. Os pesquisadores exploram como as consultas em linguagem natural desordenadas podem afetar a capacidade de um sistema de identificar com precisão o momento desejado no vídeo. Eles propõem métodos para lidar com consultas desordenadas e destacam a importância de compreender a relação entre modalidades visual e textual para melhorar a recuperação de informações em vídeos.</p>	<p>O objetivo do estudo é desenvolver uma abordagem eficaz para quantificar momentos temporais em vídeos não editados usando linguagem natural. Os pesquisadores buscam criar um sistema capaz de identificar com precisão momentos importantes, reconhecer indivíduos e objetos referenciados e encontrar o tempo significativo que corresponde à consulta de entrada. Eles também procuram comparar sua abordagem com benchmarks existentes para avaliar seu desempenho.</p>	<p>Os resultados do estudo incluem a proposta de uma solução para medir momentos temporais em vídeos não editados usando consultas em linguagem natural. Os pesquisadores desenvolveram um sistema que pode identificar o momento relevante dentro de um vídeo longo e não editado que corresponde à consulta. Eles também abordaram a questão das consultas desordenadas, destacando como a compreensão inadequada de consultas sem uma estrutura gramatical correta pode levar a uma degradação no desempenho do sistema. Além disso, o estudo apresenta uma análise detalhada dos</p>	<p>O estudo sobre a medição de momentos temporais em vídeos usando linguagem natural pode ser aplicado em repositórios institucionais para melhorar a recuperação de informações em vídeos não editados. Ao implementar a abordagem proposta pelos pesquisadores, os repositórios institucionais podem permitir que os usuários realizem consultas em linguagem natural para encontrar momentos específicos em vídeos longos. Isso facilitaria a busca e a navegação em conteúdos audiovisuais, tornando a experiência do usuário mais eficiente e intuitiva. Além disso, a capacidade de lidar com consultas</p>

		<p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>		<p>experimentos realizados, incluindo a avaliação de métricas e a comparação com benchmarks estabelecidos, como o conjunto de dados Charades-STA.</p>	<p>desordenadas pode melhorar a precisão e relevância dos resultados obtidos, aumentando a utilidade e eficácia dos repositórios institucionais para a pesquisa e o acesso a conteúdo multimídia.</p>
2023	<p>Research on the Construction and Application of Knowledge Graph in the Ceramic Field Based on Natural Language Processing</p>	<p>O problema de pesquisa abordado neste estudo é a necessidade de coletar, organizar e gerenciar o conhecimento disperso no campo da cerâmica, bem como apoiar a pesquisa científica nesse domínio. Os autores destacam a complexidade e dispersão dos dados na área da cerâmica, o que dificulta a aplicação direta de técnicas de aprendizado de máquina. Além disso, a escassez de amostras de dados e a falta de pesquisas sobre extração de conhecimento específico do campo cerâmico representam desafios adicionais para a aplicação de tecnologias avançadas, como os grafos de conhecimento, nesse contexto.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo deste estudo é aplicar a tecnologia de grafos de conhecimento no campo da cerâmica, utilizando técnicas de processamento de linguagem natural para construir grafos de conhecimento específicos desse domínio. Os autores buscam explorar as vantagens dos grafos de conhecimento na aprendizagem, organização e inferência de conhecimento no campo da cerâmica, visando superar os desafios enfrentados pelo aprendizado de máquina nesse domínio. Além disso, o estudo visa fornecer serviços de recuperação de dados mais precisos para profissionais da área de cerâmica, contribuindo assim para a pesquisa e prática nesse campo.</p>	<p>Os resultados da pesquisa incluem a construção de um grafo de conhecimento no campo da cerâmica, a realização de experimentos de reconhecimento de relacionamentos e atributos, e o desenvolvimento de um serviço de perguntas e respostas inteligente para conhecimento cerâmico. Os autores realizaram a aquisição de dados relevantes, a construção de conjuntos de dados, o reconhecimento de relacionamentos e atributos, e a construção do grafo de conhecimento do campo cerâmico. Além disso, desenvolveram um serviço de perguntas e respostas inteligente para fornecer recuperação de dados mais precisa no campo da cerâmica. Os resultados indicam uma abordagem inovadora na construção e aplicação de grafos de conhecimento no campo cerâmico, destacando a aquisição em larga escala de dados não estruturados, a construção semiautomática do</p>	<p>O estudo sobre a construção e aplicação de grafos de conhecimento no campo da cerâmica, baseado em processamento de linguagem natural, pode ser aplicado em repositórios institucionais para melhorar a organização e recuperação de informações específicas desse domínio. Ao utilizar tecnologias avançadas de grafos de conhecimento, é possível estruturar e relacionar o conhecimento cerâmico de forma mais eficiente, facilitando a busca e a análise de dados relevantes. Isso pode beneficiar repositórios institucionais ao oferecer uma maneira mais precisa e inteligente de acessar e explorar informações sobre cerâmica, contribuindo para a pesquisa e o desenvolvimento nesse campo específico.</p>

				grafo de conhecimento e a melhoria na precisão da recuperação de dados cerâmicos.	
2023	Retrieval Contrastive Learning for Aspect-Level Sentiment Classification	<p>O problema de pesquisa abordado no documento é a melhoria da classificação de sentimentos em nível de aspecto em tarefas de Processamento de Linguagem Natural (PLN) por meio do uso de Aprendizado Contrastivo de Recuperação (RCL). O objetivo é aprimorar a capacidade dos modelos de PLN em atribuir sentimentos específicos a sentenças em relação a diferentes aspectos, superando desafios como a compreensão de aspectos não treinados, filtragem de informações irrelevantes e detecção de opiniões implícitas nos textos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo principal do estudo é propor e investigar um novo framework chamado Retrieval Contrastive Learning (RCL) para aprimorar a representação de instâncias em tarefas de classificação de sentimentos em nível de aspecto. O RCL visa melhorar a capacidade dos modelos existentes de PLN em prever sentimentos de forma mais precisa, compreendendo aspectos não treinados, filtrando informações irrelevantes e identificando opiniões implícitas nos textos. Além disso, o estudo busca demonstrar a eficácia do RCL experimentalmente e analisar seu desempenho em comparação com abordagens tradicionais de classificação de sentimentos.</p>	<p>Os resultados da pesquisa demonstraram que a abordagem proposta, o framework Retrieval Contrastive Learning (RCL), foi eficaz em melhorar a representação de instâncias para a classificação de sentimentos em nível de aspecto. Após treinar os modelos existentes de classificação de sentimentos com o RCL, observou-se um aumento significativo na precisão das previsões de sentimentos, especialmente em cenários onde os modelos tradicionais enfrentavam desafios, como compreensão de aspectos não treinados, filtragem de informações irrelevantes e detecção de opiniões implícitas nos textos. Além disso, os experimentos realizados com o RCL demonstraram que os modelos aprimorados foram capazes de entender aspectos não vistos anteriormente, filtrar ruídos nos textos e capturar opiniões implícitas, resultando em representações semânticas mais ricas para a previsão de sentimentos. Os resultados obtidos com o RCL superaram as abordagens tradicionais e</p>	<p>O estudo sobre Retrieval Contrastive Learning (RCL) para classificação de sentimentos em nível de aspecto pode ser aplicado em repositórios institucionais para aprimorar a análise de sentimentos em textos acadêmicos, relatórios de pesquisa, feedback de alunos, entre outros conteúdos. Ao utilizar o RCL, os repositórios podem melhorar a compreensão de opiniões específicas em relação a diferentes aspectos, filtrar informações irrelevantes e identificar opiniões implícitas nos textos. Isso pode resultar em uma análise mais precisa e detalhada dos sentimentos expressos nos documentos armazenados nos repositórios institucionais, auxiliando na tomada de decisões e na compreensão do feedback e das opiniões dos usuários.</p>

				estabeleceram novos padrões de desempenho em tarefas de classificação de sentimentos em nível de aspecto.	
2023	SAM: cross-modal semantic alignments module for image-text retrieval	<p>O problema de pesquisa abordado neste estudo é o desafio de realizar a recuperação de informações entre imagens e textos em diferentes modalidades. Especificamente, o foco está na necessidade de desenvolver um método eficaz para alinhar semanticamente as representações visuais e textuais, a fim de melhorar o desempenho da recuperação de informações em tarefas de busca de imagem-texto.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo deste estudo é propor o módulo SAM (Cross-Modal Semantic Alignments Module) para a recuperação de imagem-texto, que utiliza uma estrutura de grafo para aprimorar o alinhamento das relações intermodais. O SAM visa explicitamente alinhar regiões de uma imagem com palavras em um texto, utilizando a estrutura de grafo bipartido para realizar atenção entre os nós conectados de diferentes modalidades. O objetivo é melhorar a precisão e eficácia da recuperação de informações entre imagens e textos.</p>	<p>Os resultados da pesquisa demonstraram que o módulo SAM proposto superou os modelos de recuperação existentes em termos de desempenho. Ao utilizar o SAM para a recuperação de imagem-texto, o estudo mostrou melhorias significativas na capacidade de capturar representações conjuntas entre as duas modalidades. Além disso, os experimentos realizados nos conjuntos de dados MS-COCO e Flickr30K mostraram que o SAM foi capaz de ser integrado com sucesso às redes de recuperação existentes, resultando em uma melhoria geral no desempenho da recuperação de informações entre imagens e textos.</p>	<p>O estudo sobre o módulo SAM para recuperação de imagem-texto pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de informações multimodais. Ao integrar o SAM em sistemas de repositórios, é possível melhorar a precisão e relevância dos resultados de pesquisa ao alinhar semanticamente as representações visuais e textuais. Isso pode beneficiar usuários que buscam informações específicas por meio de consultas que envolvem tanto imagens quanto texto, proporcionando uma experiência de busca mais eficaz e abrangente.</p>
2023	SEARCHFORMER: Semantic patent embeddings by siamese transformers for prior art search	<p>O problema de pesquisa abordado no artigo é a melhoria da eficiência e precisão na busca de arte anterior relevante em aplicações de patentes. O estudo propõe o uso de um modelo específico de BERT otimizado para gerar embeddings semânticos que facilitem a busca de informações relevantes em documentos de patentes, visando identificar</p>	<p>O objetivo do estudo é desenvolver e avaliar o SEARCHFORMER, um modelo baseado em transformers siameses, para gerar embeddings semânticos de patentes que facilitem a busca de arte anterior relevante. O modelo é projetado para melhorar a eficiência e precisão na identificação de documentos de patentes que possam negar a</p>	<p>Os resultados da pesquisa indicam que o SEARCHFORMER, um modelo baseado em transformers siameses otimizado para busca de arte anterior em patentes, apresentou melhorias significativas na classificação de documentos relevantes em comparação com baselines tradicionais. Através de experimentos de classificação</p>	<p>O estudo sobre o SEARCHFORMER, um modelo baseado em transformers siameses para busca de arte anterior em patentes, pode ser aplicado em repositórios institucionais para aprimorar a eficiência na busca e recuperação de informações relevantes. Ao utilizar embeddings semânticos gerados pelo</p>

		<p>textos que possam negar a novidade ou o passo inventivo de uma reivindicação de patente.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>novidade ou o passo inventivo de uma reivindicação de patente, contribuindo para o trabalho de examinadores de patentes.</p>	<p>retrospectiva, o SEARCHFORMER demonstrou uma melhoria estatisticamente significativa na capacidade de posicionar citações do tipo 'X' (que negam a novidade ou o passo inventivo) em posições de destaque. Além disso, o modelo foi capaz de gerar embeddings semânticos eficazes para a busca de informações relevantes em documentos de patentes, contribuindo para a eficiência e precisão do processo de busca de arte anterior.</p>	<p>SEARCHFORMER, os repositórios podem melhorar a precisão na identificação de documentos relacionados, facilitando a pesquisa e análise de conteúdo. Isso pode beneficiar instituições acadêmicas, empresas e organizações que lidam com grandes volumes de documentos patentários, agilizando o processo de busca de informações cruciais para tomada de decisões e pesquisa.</p>
2023	Semantic Tagging for the Urdu Language: Annotated Corpus and Multi-Target Classification Methods	<p>O problema de pesquisa abordado no documento é a necessidade de desenvolver ferramentas de marcação semântica eficazes para a língua urdu. Isso inclui a construção de um corpus anotado semanticamente e a aplicação de métodos de classificação de vários alvos para melhorar a compreensão e extração de informações significativas do idioma urdu. A pesquisa visa superar os desafios específicos associados à marcação semântica em uma língua morfológicamente rica e com recursos limitados, como o urdu.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>O objetivo do estudo é abordar a lacuna na pesquisa existente sobre marcação semântica para a língua urdu, desenvolvendo um corpus anotado semanticamente e aplicando métodos de classificação de vários alvos para lidar com a tarefa de desambiguação semântica. O estudo visa fornecer uma base sólida para a atribuição de tags semânticas em urdu, utilizando uma taxonomia semântica ampla e abordando a questão da desambiguação semântica de maneira mais robusta e abrangente. Além disso, o objetivo é disponibilizar o corpus anotado publicamente para facilitar o desenvolvimento e avaliação de métodos de marcação semântica em urdu.</p>	<p>O estudo resultou no desenvolvimento do primeiro corpus grande semanticamente anotado para o urdu, disponibilizado gratuitamente para a comunidade de pesquisa. Além disso, a aplicação bem-sucedida de vários classificadores de aprendizado de máquina de vários alvos para a tarefa de marcação semântica foi realizada pela primeira vez em qualquer idioma. Recursos de suporte, como a ferramenta de anotação e o léxico semântico de palavras únicas, foram criados. Os resultados demonstraram uma alta precisão de 94% no corpus proposto, destacando a eficácia dos métodos propostos para a marcação semântica em urdu. Esses avanços contribuem</p>	<p>O estudo sobre marcação semântica para a língua urdu pode ser aplicado em repositórios institucionais para enriquecer a capacidade de processamento de linguagem natural em idiomas de recursos limitados. Ao disponibilizar o corpus anotado semanticamente publicamente, instituições podem utilizar esses dados para desenvolver e avaliar ferramentas de marcação semântica em urdu, melhorar a extração de informações significativas e facilitar a pesquisa em áreas como extração de entidades nomeadas, extração de informações e classificação de documentos. Isso pode contribuir para o avanço da pesquisa em linguagem natural</p>

				significativamente para a pesquisa em processamento de linguagem natural e abrem caminho para futuras aplicações e estudos no campo da marcação semântica em urdu.	e beneficiar aplicações como tradução automática, detecção de plágio e recuperação de informações em idiomas de baixa disponibilidade de recursos.
2023	Slovak Dataset for Multilingual Question Answering	<p>O problema de pesquisa abordado no documento é a falta de conjuntos de dados de perguntas e respostas em eslovaco para sistemas de questionamento e resposta multilíngues. A necessidade de um conjunto de dados anotado manualmente em eslovaco é crucial para o desenvolvimento e aprimoramento de sistemas de questionamento e resposta nesse idioma, bem como para a pesquisa em processamento de linguagem natural multilíngue.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	O objetivo principal do estudo é criar o SK-QuAD, o primeiro conjunto de dados de perguntas e respostas para o idioma eslovaco. Esse conjunto de dados visa fornecer uma base sólida para o desenvolvimento de sistemas de geração de respostas a perguntas em linguagem natural, permitindo tanto o treinamento quanto a verificação desses sistemas. Além disso, o SK-QuAD possibilita a comparação de modelos de linguagem neurais atuais e futuros em eslovaco e em outros idiomas, contribuindo para avanços na área de processamento de linguagem natural.	O resultado da pesquisa foi a criação do SK-QuAD, o primeiro conjunto de dados de perguntas e respostas manualmente anotado em eslovaco. Este conjunto de dados contém mais de 91 mil perguntas e respostas factuais de diversas áreas, com cada pergunta tendo uma resposta marcada no parágrafo correspondente. Além disso, o SK-QuAD inclui exemplos negativos na forma de "perguntas não respondidas" e "respostas plausíveis". O conjunto de dados foi disponibilizado gratuitamente para uso científico e tem o potencial de melhorar o desempenho de sistemas de questionamento e resposta em eslovaco e em sistemas multilíngues.	O estudo sobre o SK-QuAD, o conjunto de dados de perguntas e respostas em eslovaco, pode ser aplicado em repositórios institucionais como uma importante contribuição para a pesquisa em processamento de linguagem natural e sistemas de questionamento e resposta multilíngues. Ao disponibilizar o SK-QuAD gratuitamente para uso científico, os repositórios institucionais podem incorporar esse recurso valioso em seus acervos, permitindo que pesquisadores e desenvolvedores acessem e utilizem o conjunto de dados para avançar em suas próprias investigações e projetos relacionados à linguagem natural e inteligência artificial.
2023	SoundexGR: An algorithm for phonetic matching for the Greek language	O problema de pesquisa abordado no artigo é a necessidade de desenvolver um algoritmo de correspondência fonética específico para a língua grega. Os autores identificaram a falta de um algoritmo eficaz de Soundex para o idioma grego, o que dificulta tarefas de	O objetivo do estudo é introduzir um algoritmo de correspondência fonética projetado especificamente para a língua grega. Os autores buscam adaptar e estender o conceito original do algoritmo Soundex para acomodar as regras fonéticas do idioma	O resultado da pesquisa sobre o algoritmo SoundexGR para correspondência fonética na língua grega demonstrou que o algoritmo é eficaz e supera outras abordagens, como stemming, transcrição fonêmica completa e distância de edição, em termos de F-Score. Os	O estudo sobre o algoritmo SoundexGR para correspondência fonética na língua grega pode ser aplicado em repositórios institucionais para melhorar a precisão da recuperação de informações e do processamento de texto em documentos em grego. Ao

		<p>Processamento de Linguagem Natural e Recuperação de Informação que dependem da correspondência fonética de palavras. Assim, o objetivo principal do estudo é propor e avaliar o algoritmo SoundexGR como uma solução para melhorar a correspondência fonética no idioma grego e, conseqüentemente, a precisão de tarefas de processamento de texto.</p> <p><b>Categoria:</b> Recurso de Informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>grego, resultando em uma família de algoritmos denominada SoundexGR. Eles pretendem demonstrar a eficácia do algoritmo em diferentes cenários, fornecendo resultados experimentais comparativos para diversos parâmetros do algoritmo, a fim de revelar o equilíbrio entre precisão e recall em conjuntos de dados com diferentes tipos de erros. Além disso, os autores comparam o desempenho do SoundexGR com outras abordagens, como stemming, transcrição fonêmica completa e distância de edição, para destacar a superioridade do SoundexGR em termos de F-Score, especialmente em coleções de palavras com sons semelhantes.</p>	<p>experimentos realizados mostraram que o SoundexGR alcançou uma pontuação média de F-Score de 95% em coleções de palavras com sons semelhantes. Isso indica que o algoritmo é altamente preciso na identificação e correspondência de palavras com base em sua pronúncia fonética no idioma grego. Além disso, os resultados experimentais revelaram o equilíbrio entre precisão e recall em diferentes cenários e tipos de erros, destacando a eficácia e adaptabilidade do SoundexGR para uma variedade de tarefas de processamento de texto na língua grega.</p>	<p>implementar o SoundexGR em repositórios, é possível aprimorar a correspondência de palavras com base em sua pronúncia fonética, facilitando a busca por informações relevantes e corrigindo erros de digitação. Isso pode resultar em uma recuperação mais eficiente e precisa de documentos, beneficiando pesquisadores, estudantes e demais usuários que utilizam os repositórios institucionais para acessar informações em grego.</p>
2023	SRU-based Multi-angle Enhanced Network for Semantic Text Similarity Calculation of Big Data Language Model	<p>O problema de pesquisa abordado neste artigo é a dificuldade em calcular a similaridade semântica de textos em chinês devido à complexidade e ambigüidade da semântica chinesa. A pesquisa busca desenvolver um modelo computacional, denominado SRU-based Multi-angle Enhanced Network (SMAEN), que possa capturar efetivamente as características de interação semântica do texto chinês a partir de múltiplos ângulos, visando melhorar a precisão e eficácia dos cálculos de</p>	<p>O objetivo deste estudo é propor e desenvolver um modelo computacional avançado, o SRU-based Multi-angle Enhanced Network (SMAEN), para calcular a similaridade semântica de textos em chinês. O modelo SMAEN visa abordar as deficiências existentes na calculação de similaridade de texto em chinês, capturando informações de interação e semântica de sentenças em profundidade a partir de três ângulos diferentes: semântica, tempo e espaço. Além disso, o estudo busca melhorar a análise</p>	<p>O resultado da pesquisa apresentada no artigo é a proposta e validação do modelo SRU-based Multi-angle Enhanced Network (SMAEN) para calcular a similaridade semântica de textos em chinês. O SMAEN demonstrou um desempenho excelente em termos de precisão ao ser avaliado em conjuntos de dados públicos chineses, como o LCQMC e o PAWS-X, em comparação com o método de linha de base. Houve um aumento na precisão de 1,56% e 3,11% nos conjuntos de dados</p>	<p>O estudo sobre o SRU-based Multi-angle Enhanced Network (SMAEN) para calcular a similaridade semântica de textos em chinês pode ser aplicado em repositórios institucionais de diversas formas. Por exemplo, ao utilizar o modelo SMAEN, os repositórios institucionais podem aprimorar a análise de similaridade entre textos em chinês, facilitando a recuperação de informações relevantes e a recomendação de conteúdo personalizado. Além disso, a capacidade do SMAEN de capturar informações de</p>

		<p>similaridade de texto em grandes modelos de dados linguísticos.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)</p> <p><b>Subcategoria:</b> Interação e IA</p>	<p>global da similaridade de sentenças, combinando Bi-SRU e CNN para calcular a similaridade de texto a partir de múltiplos ângulos, utilizando uma abordagem de atenção suave e agregação de correspondência em um framework interativo.</p>	<p>LCQMC e PAWS-X, respectivamente. O modelo SMAEN combina embeddings de caracteres e palavras, utiliza Bi-SRU para extrair informações sintáticas e semânticas das sentenças, e modela a similaridade global das interações de sentenças a partir de três perspectivas: semântica, tempo e espaço. Esses resultados indicam que o SMAEN é capaz de capturar informações de similaridade de texto de forma mais precisa e eficaz, fornecendo uma base sólida para pesquisas futuras em aplicações de big data.</p>	<p>interação e semântica de sentenças em múltiplos ângulos pode melhorar a precisão na identificação de padrões e relações em grandes conjuntos de dados linguísticos, auxiliando na organização e categorização eficiente de documentos nos repositórios institucionais.</p>
2023	Tashaphyne0.4: a new arabic light stemmer based on rhizome modeling approach	<p>O problema de pesquisa abordado no artigo é a melhoria dos processos de recuperação de informações em processamento de linguagem natural, especificamente no contexto da extração de raízes e hastes precisas de texto em árabe. O objetivo é desenvolver um novo algoritmo, o Tashaphyne0.4, que supere as limitações de outros stemmers existentes e forneça resultados mais precisos e eficazes na extração de raízes e hastes em textos árabes.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade</p>	<p>O objetivo do estudo é propor e desenvolver o algoritmo Tashaphyne0.4, um stemmer leve para a língua árabe baseado no modelo Rhizome. Este algoritmo visa extrair raízes e hastes de forma eficiente e precisa a partir de palavras árabes, contribuindo para aprimorar os sistemas de recuperação de informações, indexação de texto e classificadores em mineração de dados. O Tashaphyne0.4 busca superar as limitações de outros stemmers existentes e demonstrar sua robustez e eficácia por meio de experimentos e comparações com outros stemmers bem estabelecidos.</p>	<p>O resultado da pesquisa demonstrou que o algoritmo Tashaphyne0.4 superou outros seis stemmers árabes bem conhecidos em termos de extração de raízes e hastes precisas de palavras árabes. O Tashaphyne0.4 obteve resultados significativamente melhores do que os stemmers Khoja, ISRI, Motaz/Light10, Tashaphyne0.3, FARASA e Assem, conforme avaliado em quatro conjuntos de dados de benchmark árabes abrangentes. A comparação foi realizada com base em métricas de precisão, recall e F1, demonstrando a eficácia e superioridade do Tashaphyne0.4 na extração de raízes e hastes em textos árabes</p>	<p>O estudo sobre o algoritmo Tashaphyne0.4, um stemmer leve para a língua árabe, pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações em textos em árabe. Ao utilizar o Tashaphyne0.4, os repositórios podem melhorar a indexação de documentos, facilitar a busca por informações precisas e relevantes em árabe e aprimorar a organização e categorização de conteúdo. Isso contribui para uma melhor acessibilidade e eficiência na busca e recuperação de documentos em língua árabe em repositórios institucionais.</p>

		percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA		em comparação com os stemmers existentes.	
2023	Task-like training paradigm in CLIP for zero-shot sketch-based image retrieval	<p>O problema de pesquisa abordado neste estudo é a melhoria do desempenho da recuperação de imagens baseada em esboços (SBIR) em um cenário de zero-shot, onde o modelo deve ser capaz de recuperar imagens com base em esboços sem a necessidade de exemplos de treinamento específicos para cada classe. O foco está em superar as lacunas semânticas e modais entre esboços e imagens, bem como entre classes vistas e não vistas, a fim de melhorar a precisão e a generalização do modelo de recuperação de imagens.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é propor e desenvolver um paradigma de treinamento denominado Task-Like Training (TLT) para aprimorar a recuperação de imagens baseada em esboços em um cenário de zero-shot. O TLT visa abordar os desafios de percepção cruzada entre modalidades e da lacuna semântica entre classes vistas e não vistas. Para alcançar esse objetivo, o estudo utiliza o codificador de texto do CLIP e mecanismos de aprendizado inovadores, como aprendizado de identificação baseado em texto, tutoria de prompt de texto e aprendizado de consistência cruzada modal. O objetivo final é melhorar significativamente o desempenho do modelo em comparação com os métodos existentes em várias bases de dados de referência.</p>	<p>Os resultados da pesquisa demonstram que a abordagem proposta, o paradigma de treinamento Task-Like Training (TLT), supera significativamente os métodos do estado-da-arte em recuperação de imagens baseada em esboços em um cenário de zero-shot. Através da utilização do TLT, o modelo apresenta melhorias na percepção cruzada entre modalidades, na redução da lacuna semântica entre classes vistas e não vistas, e na consistência entre modalidades. Os experimentos abrangentes realizados nas bases de dados Sketchy, Sketchy-No, Tuberlin e QuickDraw demonstram que a abordagem proposta supera de forma significativa os métodos existentes, proporcionando uma recuperação mais precisa e eficaz de imagens com base em esboços.</p>	<p>O estudo sobre o paradigma de treinamento Task-Like Training (TLT) para recuperação de imagens baseada em esboços em um cenário de zero-shot pode ser aplicado em repositórios institucionais para aprimorar a busca e recuperação de imagens com base em esboços. Ao implementar as técnicas inovadoras propostas no estudo, como aprendizado de identificação baseado em texto, tutoria de prompt de texto e aprendizado de consistência cruzada modal, os repositórios institucionais podem melhorar a precisão e eficácia da recuperação de imagens com base em esboços, facilitando a busca e organização de conteúdo visual de forma mais eficiente e precisa.</p>
2023	The power and potentials of Flexible Query Answering Systems: A critical and comprehensive analysis	<p>O problema de pesquisa abordado neste documento é a análise crítica e abrangente dos sistemas de resposta a consultas flexíveis (FQAS), com foco nas abordagens conceituais e de implementação, bem como</p>	<p>O objetivo deste estudo é fornecer uma análise crítica e abrangente dos sistemas de resposta a consultas flexíveis (FQAS), destacando as tendências, desenvolvimentos recentes e desafios enfrentados</p>	<p>Os resultados da pesquisa incluem insights sobre o estado-da-arte, tendências, novas propostas e direções futuras de pesquisa no campo dos sistemas de resposta a consultas flexíveis. Foram</p>	<p>O estudo sobre sistemas de resposta a consultas flexíveis (FQAS) pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações relevantes e a interação com os usuários. A</p>

		<p>nos desafios futuros enfrentados por esses sistemas. A pesquisa visa identificar os principais métodos, descobrir e rastrear abordagens convencionais e emergentes, e direcionar futuras pesquisas na área, especialmente diante do rápido desenvolvimento das ciências de dados e inteligência artificial.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>por esses sistemas. O estudo visa identificar novas direções de pesquisa promissoras e tendências emergentes na área de FQAS, a fim de contribuir para moldar o futuro desses sistemas e tópicos relacionados.</p>	<p>identificadas novas tendências em sistemas de resposta a consultas e questões, como métodos de indexação para big data, abordagens baseadas em dados para classificação de documentos relevantes e recomendação de documentos, e a criação de corpora de benchmark especializados para treinar e testar modelos baseados em dados. Além disso, foram destacadas necessidades e desafios emergentes, como a identificação de viés e aspectos desequilibrados em coleções de benchmark, a conformidade com regulamentações de explicabilidade de modelos e privacidade do usuário, e a mudança de perspectiva em relação aos FQAS.</p>	<p>análise crítica e abrangente dos FQAS pode ajudar a desenvolver sistemas mais inteligentes e eficazes, capazes de interpretar consultas, selecionar fontes confiáveis, analisar e apresentar informações de forma apropriada. Isso pode resultar em uma melhor experiência do usuário, maior eficiência na busca por informações e na gestão de dados, e na garantia da qualidade e veracidade dos resultados obtidos. Além disso, a consideração de tendências emergentes, como a modelagem de dados e a conformidade com regulamentações de privacidade, pode contribuir para a evolução e aprimoramento contínuo dos repositórios institucionais.</p>
2023	The Power of Selecting Key Blocks with Local Pre-ranking for Long Document Information Retrieval	<p>O problema de pesquisa abordado neste documento é a eficácia da seleção de blocos-chave em documentos longos para a recuperação de informações. A pesquisa investiga como a pré-classificação local de blocos pode ajudar na identificação e seleção dos blocos mais relevantes em documentos extensos, visando melhorar a precisão e eficiência dos modelos de processamento de linguagem natural, como o</p>	<p>O objetivo deste estudo é demonstrar a eficácia da seleção de blocos-chave com pré-classificação local em documentos longos para a recuperação de informações. O estudo visa melhorar a capacidade dos modelos de processamento de linguagem natural, como o BERT, em lidar com documentos extensos, selecionando os blocos mais relevantes para formar uma representação mais concisa e informativa do documento. A pesquisa busca mostrar como</p>	<p>Os resultados da pesquisa demonstraram a eficácia da abordagem proposta de selecionar blocos-chave com pré-classificação local em documentos longos para a recuperação de informações. Os experimentos realizados mostraram que o método desenvolvido para aprender a selecionar blocos superou todos os outros métodos avaliados, incluindo os apresentados em trabalhos anteriores. Além disso, a seleção de passagens-chave com pré-classificação local</p>	<p>O estudo sobre a seleção de blocos-chave com pré-classificação local em documentos longos para a recuperação de informações pode ser aplicado em repositórios institucionais para aprimorar a eficiência e precisão da busca e recuperação de informações em documentos extensos. Ao implementar a abordagem de selecionar os blocos mais relevantes em documentos longos, os repositórios institucionais podem melhorar a experiência do</p>

		<p>BERT, em tarefas de recuperação de informações.</p> <p><b>Categoria:</b> Recursos de informação <b>Subcategoria:</b> Documento</p>	<p>essa abordagem pode aprimorar a precisão e eficiência dos sistemas de recuperação de informações ao lidar com documentos extensos.</p>	<p>tornou o modelo PARADE mais eficiente e preciso, com resultados gerais melhores ou pelo menos semelhantes. A pesquisa também comparou a abordagem proposta com modelos de transformadores de atenção esparsa e demonstrou que a nova abordagem alcançou resultados superiores sem a necessidade de kernels CUDA personalizados.</p>	<p>usuário ao apresentar resultados mais precisos e relevantes em suas pesquisas. Isso pode resultar em uma recuperação de informações mais eficaz, economizando tempo e esforço dos usuários ao localizar conteúdo relevante nos repositórios institucionais.</p>
2023	Transformer-Based Discriminative and Strong Representation Deep Hashing for Cross-Modal Retrieval	<p>O problema de pesquisa abordado neste trabalho é a melhoria do desempenho da recuperação cruzada de informações em diferentes modalidades (como texto e imagem) por meio do desenvolvimento de um framework de hashing profundo baseado em transformadores. O objetivo é explorar a correlação entre dados de diferentes modalidades e gerar representações distintas e discriminativas para melhorar a eficiência da recuperação cruzada de informações.</p> <p><b>Categoria:</b> Recurso de Informação <b>Subcategoria:</b> Informação</p>	<p>O objetivo deste trabalho é propor um framework de hashing profundo denominado Transformer-Based Discriminative and Strong Representation Deep Hashing (TDSRDH) para a recuperação cruzada de informações em diferentes modalidades, como texto e imagem. O TDSRDH visa obter representações fortes e discriminativas das características dos dados, preservando a similaridade entre as modalidades durante o processo de aprendizado. Além disso, o estudo busca demonstrar a eficácia do framework por meio de experimentos em conjuntos de dados reais e ablações para validar as contribuições propostas.</p>	<p>Os resultados da pesquisa demonstraram que o framework proposto, Transformer-Based Discriminative and Strong Representation Deep Hashing (TDSRDH), superou outros métodos de referência em termos de desempenho de recuperação cruzada de informações. Por meio de experimentos realizados em três conjuntos de dados amplamente utilizados (MIRFLICKR-25K, IAPR TC-12 e NUS-WIDE), o TDSRDH obteve melhorias de 1% a 3% em relação aos baselines. Além disso, os experimentos de ablação realizados confirmaram a eficácia das contribuições propostas, demonstrando melhorias de 2% e 1% em diferentes aspectos avaliados. Portanto, os resultados indicam que o TDSRDH é uma abordagem promissora para a recuperação cruzada de</p>	<p>O estudo sobre Transformer-Based Discriminative and Strong Representation Deep Hashing (TDSRDH) pode ser aplicado em repositórios institucionais para aprimorar a recuperação de informações em diferentes modalidades, como texto e imagem. Ao utilizar o framework TDSRDH, os repositórios podem melhorar a eficiência da busca e recuperação de conteúdo heterogêneo, permitindo uma correlação mais precisa entre dados de diferentes tipos. Isso pode resultar em uma experiência de busca mais eficaz para os usuários, facilitando a localização de informações relevantes em repositórios institucionais.</p>

				informações em diferentes modalidades.	
2023	Trilingual conversational intent decoding for response retrieval	<p>O problema de pesquisa abordado neste trabalho é a decodificação de intenções conversacionais trilingues para a recuperação de respostas. A pesquisa visa desenvolver um sistema de Perguntas e Respostas Conversacionais Trilingues (TIACQA) capaz de responder a perguntas em inglês, hindi e iorubá. O foco está na complexidade reduzida e na baixa sobrecarga de recursos para tarefas de modelagem trilingue, com desempenho promissor em conjuntos de dados de referência grandes e pequenos. O objetivo é permitir que o sistema compreenda a linguagem utilizada, identifique a intenção por trás de uma pergunta e forneça respostas relevantes com base nessa intenção prevista.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query)  <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste trabalho é desenvolver um sistema de Perguntas e Respostas Conversacionais Trilingues (TIACQA) capaz de responder a perguntas em inglês, hindi e iorubá. O sistema proposto utiliza uma abordagem híbrida de aprendizado extremo que combina técnicas de recuperação de informações e aprendizado supervisionado. Além disso, o objetivo é avaliar o desempenho do sistema em relação a conjuntos de dados de referência e ao conjunto de dados trilingue desenvolvido, demonstrando uma melhoria significativa em comparação com outros modelos existentes. O trabalho visa contribuir para o avanço de sistemas conversacionais capazes de lidar com múltiplos idiomas, especialmente aqueles com recursos limitados, com potenciais aplicações em diversas áreas, como educação, saúde e atendimento ao cliente.</p>	<p>Os resultados da pesquisa indicam que o modelo proposto de Perguntas e Respostas Conversacionais Trilingues (TIACQA) obteve um desempenho superior em termos de métricas de sensatez e especificidade, com uma pontuação de 92% e 90%, respectivamente. Além disso, o modelo demonstrou ser capaz de identificar a linguagem utilizada, determinar a intenção por trás de uma pergunta e fornecer respostas relevantes na mesma língua. A avaliação do modelo em três conjuntos de dados de referência para detecção de intenções e em um conjunto de dados trilingue para recuperação de respostas mostrou que o modelo superou abordagens existentes em termos de métricas de desempenho e tempo de treinamento mais rápido. Esses resultados destacam a eficácia do modelo proposto em lidar com conversas multi-dialetos de forma eficiente e eficaz, mesmo em cenários com dados limitados, com uma arquitetura leve e algoritmos eficientes contribuindo para seu alto desempenho e versatilidade.</p>	<p>O estudo sobre decodificação de intenções conversacionais trilingues para recuperação de respostas pode ser aplicado em repositórios institucionais para melhorar a interação com usuários que acessam informações em diferentes idiomas. Ao implementar um sistema de Perguntas e Respostas Conversacionais Trilingues (TIACQA), os repositórios podem oferecer respostas precisas e relevantes em inglês, hindi e iorubá, facilitando a comunicação e a busca por informações em múltiplos idiomas. Isso pode aprimorar a experiência do usuário, aumentar a eficiência na recuperação de informações e promover a acessibilidade em ambientes institucionais multilingues.</p>

2023	Universal Multimodal Representation for Language Understanding	<p>O problema de pesquisa abordado neste trabalho é a necessidade de desenvolver métodos eficazes para incorporar informações visuais em tarefas de compreensão de linguagem natural. A maioria dos estudos anteriores se concentra em modelos conjuntos de imagens e textos, como pré-treinamento de visão-linguagem e tarefas de aplicação multimodal. No entanto, esses estudos geralmente dependem de grandes conjuntos de dados anotados texto-imagem para tarefas específicas, como legendagem de imagens e resposta a perguntas visuais. O desafio essencial é a falta de um corpus alinhado texto-imagem de alta qualidade para aplicativos de NLP baseados apenas em texto. Portanto, o problema central é investigar um método geral que aproveite informações visuais em uma ampla gama de tarefas monomodais, como processamento de linguagem baseado apenas em texto, e compreender o papel das imagens na representação da linguagem, bem como a aplicação da multimodalidade em cenários padrão de NLP.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de</p>	<p>O objetivo deste trabalho é apresentar um método de representação visual universal que se baseia apenas em um conjunto inicial de anotações independentes da tarefa, em vez de depender de grandes conjuntos de dados anotados texto-imagem específicos da tarefa. Isso visa superar a limitação do uso de informações visuais em tarefas padrão de NLP baseadas apenas em texto. O método proposto busca conectar caminhos monomodais de incorporação de texto e imagem, recuperando imagens diversas de um lookup table de tópicos-imagem leve ou de um espaço de incorporação texto-visual pré-treinado em pares texto-imagem em grande escala. A fusão das representações de texto e imagem é realizada por meio de uma camada de atenção simples e eficaz. O objetivo é aplicar facilmente essa abordagem a tarefas baseadas apenas em texto sem a necessidade de corpora paralelos multimodais anotados manualmente.</p>	<p>Os resultados da pesquisa demonstram a eficácia do método proposto de representação visual universal em uma ampla gama de tarefas de processamento de linguagem natural. Foram realizados experimentos em 14 conjuntos de dados de referência envolvendo inferência de linguagem natural, similaridade semântica, classificação de texto e tradução automática. A análise e os experimentos confirmam a eficácia do método, mostrando melhorias significativas em diversas tarefas e idiomas. Além disso, a pesquisa revela que a incorporação de informações visuais enriquece as representações textuais de palavras de conteúdo, fornece informações detalhadas sobre a relação entre conceitos e eventos e pode contribuir para a desambiguação. A abordagem proposta supera a necessidade de grandes conjuntos de dados anotados texto-imagem específicos da tarefa, tornando-a universal em termos de requisitos de tarefa. A pesquisa também destaca a flexibilidade e controle do processo de recuperação de imagens, bem como a capacidade de aplicar o método a tarefas baseadas apenas em texto em diferentes idiomas.</p>	<p>O estudo sobre a representação visual universal para compreensão de linguagem pode ser aplicado em repositórios institucionais para melhorar a indexação e recuperação de informações. Ao incorporar informações visuais em textos, os repositórios podem oferecer uma representação mais rica e contextualizada do conteúdo, facilitando a busca e a compreensão dos usuários. Além disso, a abordagem proposta pode enriquecer as representações de palavras-chave e metadados, tornando a organização e a recuperação de documentos mais eficazes.</p>
------	--	---	---	--	--

		informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA			
2023	Unveiling the inventive process from patents by extracting problems, solutions and advantages with natural language processing	<p>O problema de pesquisa abordado neste estudo é a extração de problemas, soluções e vantagens de patentes por meio de processamento de linguagem natural. Os autores buscam desenvolver um sistema baseado em modelos de linguagem transformadora para identificar esses elementos-chave do processo inventivo a partir de um grande volume de sentenças de patentes. Eles destacam que a análise manual de patentes é uma tarefa tediosa e propensa a viés, e que a análise automática de patentes é crucial para auxiliar empresas, pesquisadores, advogados e formuladores de políticas.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>	<p>O objetivo deste estudo é aprimorar um modelo de linguagem baseado em transformadores para extrair de forma eficaz e precisa os elementos fundamentais do processo inventivo - problemas técnicos, soluções para esses problemas e efeitos vantajosos da invenção - a partir do texto de patentes. A meta é alcançar um alto desempenho na identificação desses conceitos em diversas áreas tecnológicas, contribuindo para a análise automatizada de patentes e fornecendo insights valiosos para mapeamento de arte prévia, geração de novas ideias e identificação de evolução tecnológica.</p>	<p>O modelo desenvolvido neste estudo alcançou um F1 score de 90% na extração de problemas, soluções e vantagens de patentes em diversas áreas tecnológicas. Esse resultado é considerado notável em comparação com outros sistemas de ponta capazes de analisar documentos técnicos complexos, como é o caso de patentes. A precisão e eficácia do modelo demonstram sua capacidade de identificar com sucesso os elementos-chave do processo inventivo a partir do texto das patentes, contribuindo significativamente para a análise automatizada e a extração de informações valiosas contidas nesses documentos.</p>	<p>O estudo propõe um modelo baseado em processamento de linguagem natural para extrair problemas, soluções e vantagens de patentes, o que pode ser aplicado em repositórios institucionais para automatizar a análise de documentos técnicos, como patentes. Essa aplicação pode facilitar a identificação de informações relevantes, apoiar a geração de novas ideias, mapear a arte prévia e acompanhar a evolução tecnológica, beneficiando pesquisadores, inventores, advogados de patentes e formuladores de políticas.</p>
2023	Urdu Named Entity Recognition System	<p>O problema de pesquisa abordado neste artigo é a</p>	<p>O objetivo deste estudo é compreender o estado da arte</p>	<p>O estudo apresentou resultados promissores, demonstrando um</p>	<p>O estudo sobre o Reconhecimento de Entidades</p>

	Using Deep Learning Approaches	<p>necessidade de aprimorar os sistemas de Reconhecimento de Entidades Nomeadas (NER) em Urdu, uma língua morfológica rica e escrita da direita para a esquerda. Os autores destacam a dependência massiva de engenharia manual de recursos e pré-processamento de dados nos sistemas de UNER existentes, e propõem uma abordagem baseada em deep learning para superar esses desafios e melhorar a eficiência e precisão do reconhecimento de entidades nomeadas em Urdu.</p> <p><b>Categoria:</b> Representação do problema do usuário (necessidade real de informação, necessidade percebida de informação, solicitação e query) <b>Subcategoria:</b> Interação e IA</p>	<p>dos sistemas de Reconhecimento de Entidades Nomeadas (NER) em Urdu e propor um sistema baseado em deep learning que possa automaticamente aprender recursos a partir dos dados, eliminando a necessidade de engenharia manual de recursos. Os pesquisadores buscam superar as limitações dos modelos baseados em recursos existentes e demonstrar a eficácia das abordagens de deep learning na tarefa de NER em Urdu. Além disso, o estudo visa apresentar um conjunto de dados de tweets em Urdu anotados manualmente para cinco classes de entidades nomeadas e comparar o desempenho do sistema proposto com os sistemas de NER existentes, mostrando um progresso notável em relação aos métodos atuais.</p>	<p>progresso significativo em relação aos métodos atuais de Reconhecimento de Entidades Nomeadas (NER) em Urdu. O sistema proposto alcançou uma melhoria de 6,26% no escore F1 em comparação com os métodos de estado da arte. Os pesquisadores desenvolveram várias arquiteturas baseadas em deep learning, incluindo a combinação de redes neurais convolucionais (CNN) e redes neurais recorrentes (RNNs) para representações em nível de caractere e palavra, além de utilizar informações de Part-of-Speech (POS) e diferentes arquiteturas como Long Short-Term Memory (LSTM) e Gated Recurrent Unit (GRU). Os resultados foram avaliados em quatro conjuntos de dados de referência e demonstraram a eficiência do modelo proposto sem a necessidade de engenharia manual de recursos.</p>	<p>Nomeadas (NER) em Urdu usando abordagens de deep learning pode ser aplicado em repositórios institucionais para melhorar a extração de informações de documentos em Urdu. Ao implementar o sistema proposto, os repositórios podem automatizar a identificação e classificação de entidades nomeadas em textos em Urdu, facilitando a recuperação de informações relevantes e a organização eficiente dos documentos. Isso pode aprimorar a acessibilidade e a usabilidade dos repositórios institucionais para usuários que buscam informações específicas em Urdu.</p>
2023	Video Moment Localization Network Based on Text Multi-semantic Clues Guidance	<p>O problema de pesquisa abordado neste estudo é a localização de momentos em vídeos com base em pistas textuais multi-semânticas. O objetivo é desenvolver uma rede de localização de momentos de vídeo que utilize informações textuais para guiar a extração e codificação de características de vídeo, a fim de aprimorar a</p>	<p>O objetivo deste estudo é propor uma nova rede de localização de momentos de vídeo baseada em pistas textuais multi-semânticas, que integre fatores-chave para a tarefa desafiadora de localização de momentos em vídeos. A rede visa melhorar a representação das características semânticas do texto, guiar a extração de</p>	<p>Os resultados da pesquisa incluem a proposta de uma nova rede de localização de momentos de vídeo baseada em pistas textuais multi-semânticas, que demonstrou melhorias significativas em relação aos métodos state-of-the-art. A abordagem desenvolvida neste estudo proporcionou avanços na precisão e eficiência da</p>	<p>O estudo sobre a rede de localização de momentos de vídeo baseada em pistas textuais multi-semânticas pode ser aplicado em repositórios institucionais para aprimorar a indexação e recuperação de vídeos com base em consultas textuais. Ao integrar pistas textuais para guiar a extração de características de vídeo, essa</p>

		<p>correspondência entre o texto da consulta e os momentos relevantes no vídeo.</p> <p><b>Categoria:</b> Recursos de informação</p> <p><b>Subcategoria:</b> Documento</p>	<p>características de vídeo relevantes para o texto e promover interações eficazes entre modalidades de texto e vídeo para aprimorar a precisão e eficiência da localização de momentos em vídeos.</p>	<p>localização de momentos em vídeos, destacando a importância da integração de pistas textuais multi-semânticas para guiar a extração de características de vídeo e melhorar a correspondência entre texto de consulta e momentos relevantes no vídeo. Os experimentos realizados em dois conjuntos de dados, Charades-STA e ActivityNet Captions, validaram a eficácia da abordagem proposta em comparação com métodos existentes.</p>	<p>abordagem pode melhorar a precisão na correspondência entre consultas de usuários e momentos relevantes nos vídeos armazenados nos repositórios. Isso pode facilitar a busca e recuperação de informações específicas em vídeos, tornando os repositórios institucionais mais eficientes e acessíveis para os usuários.</p>
--	--	---	--	--	--