



HÍBRIDO!

APLICAÇÕES DO PROCESSAMENTO DE LINGUAGEM NATURAL NO AMBIENTE EDUCACIONAL: UMA REVISÃO SISTEMÁTICA DA LITERATURA

Fernanda Pereira Santana¹; Leandro Corrêa Magalhães²

Resumo: A utilização de recursos tecnológicos vem acontecendo com uma frequência cada vez maior dentro do ambiente educacional e as técnicas de processamento de linguagem natural podem ser aplicadas neste contexto. Este artigo apresenta uma revisão sistemática da literatura com o objetivo de levantar e analisar as diferentes maneiras pelas quais o PLN está sendo utilizado no contexto educacional, bem como examinar as contribuições específicas que essas aplicações trazem para os processos de ensino e aprendizagem. A revisão revela que, embora o campo do PLN na educação seja ainda pouco explorado, houve um aumento notável no interesse nos últimos anos, impulsionado pelo avanço tecnológico e pelas novas exigências resultantes da pandemia de Covid-19.

Palavras-Chave: Ensino e Aprendizagem. Processamento de Linguagem Natural. Tecnologias da Informação e da Comunicação.

1. INTRODUÇÃO

No atual cenário educacional, a tecnologia se estabeleceu como um catalisador de transformação, redefinindo a maneira como educadores e alunos interagem com o conhecimento. Emergem, na contemporaneidade, diversas abordagens e possibilidades de se ampliar as ações pedagógicas em sala de aula, assim como seu potencial, dado o contexto social e tecnológico e os novos hábitos e práticas sociais. (FIGUEIREDO; PAZ; JUNQUEIRA, 2015, p.1156).

¹ Mestre em Sistemas de Informação e Gestão do Conhecimento. Professora EBTT no IFNMG – Campus Avançado Porteirinha. e-mail: fernanda.pereira@ifnmg.edu.br. <https://lattes.cnpq.br/8097705177197196>

² Especialista em Informática e Comunicação na Educação. Técnico de Tecnologia da Informação no IFNMG – Campus Avançado Porteirinha. e-mail: leandro.magalhaes@ifnmg.edu.br. <http://lattes.cnpq.br/7820233785316355>

Os recursos tecnológicos não apenas revolucionaram a dinâmica da sala de aula, mas também moldaram a abordagem dos professores nos processos de ensino e aprendizagem. Nesse contexto, é fundamental compreender que as tecnologias da informação e comunicação (TICs) não representam apenas ferramentas, mas sim impulsionadoras de transformação. Além de oferecer novas vias para o aprendizado e aquisição de conhecimento por parte dos alunos, podem também capacitar os educadores. Por meio das TICs, os professores têm acesso a uma vasta gama de recursos educacionais, descobrem novas práticas de ensino, e exploram uma infinidade de possibilidades pedagógicas.

A interseção entre tecnologia e educação não apenas redefine o processo de ensino-aprendizagem, mas também amplia os horizontes do conhecimento para educadores e alunos. Ao abraçar essas oportunidades proporcionadas pela tecnologia, educadores e discentes podem não apenas acompanhar a evolução do mundo digital, mas também se tornarem agentes ativos na construção do futuro educacional, explorando, inovando e enriquecendo suas experiências de aprendizado e ensino.

O Processamento de Linguagem Natural (PLN) emerge como uma força transformadora no campo da educação, oferecendo soluções inovadoras para desafios pedagógicos contemporâneos. Dessa forma, corrobora-se que o PLN lida com problemas relacionados à automação da interpretação e da geração da língua humana em aplicações como Tradução Automática, Sumarização Automática de Textos, Ferramentas de Auxílio à Escrita, Perguntas e Respostas, Categorização Textual, Recuperação e Extração de Informação, entre muitas outras.

Este domínio multidisciplinar, que combina linguística computacional, inteligência artificial e aprendizado de máquina, capacita computadores a compreender, interpretar e gerar linguagem humana de maneira natural. Em um mundo onde a tecnologia permeia todos os aspectos de nossa vida, o PLN tem se destacado como uma ferramenta necessária no cenário educacional, promovendo interações mais inteligentes e personalizadas entre educadores e alunos.

Esta revisão sistemática da literatura tem como objetivo levantar e analisar as diferentes maneiras pelas quais o PLN está sendo utilizado no contexto educacional, bem como examinar as contribuições específicas que essas aplicações trazem para os processos de ensino e aprendizagem.

Este exame crítico é vital não apenas para educadores e pesquisadores, mas também para todos aqueles interessados em explorar o vasto território onde a tecnologia e a educação se encontram.

2. METODOLOGIA

Para a condução deste estudo, optou-se por adotar a Revisão Sistemática de Literatura (RSL). Esta abordagem foi escolhida pela sua natureza imparcial, objetiva e sistemática, visando responder de maneira precisa às questões de pesquisa ao identificar todos os resultados pertinentes na busca pelos artigos. A RSL, como estratégia metodológica, proporciona uma análise criteriosa e abrangente, permitindo uma compreensão aprofundada do cenário existente na literatura.

Para Galvão e Ricarte (2019), uma revisão sistemática de literatura é

Uma modalidade de pesquisa, que segue protocolos específicos, e que busca entender e dar alguma logicidade a um grande corpus documental, especialmente, verificando o que funciona e o que não funciona num dado contexto. Está focada no seu caráter de reprodutibilidade por outros pesquisadores, apresentando de forma explícita as bases de dados bibliográficos que foram consultadas, as estratégias de busca empregadas em cada base, o processo de seleção dos artigos científicos, os critérios de inclusão e exclusão dos artigos e o processo de análise de cada artigo.

A utilização da RSL tem como objetivo apresentar uma avaliação justa sobre o tema usando uma metodologia confiável, rigorosa e auditável. (KITCHENHAM; CHARTERS, 2007).

Considerando que a qualidade de uma pesquisa está diretamente ligada à excelência do processo de revisão do estado da arte, como indicado por Dermeval et al. (2019), torna-se crucial estabelecer um método formal com critérios e etapas explícitas e bem definidas. Para garantir a qualidade deste trabalho, utilizou-se como referência o guia prático elaborado por Neiva e Silva (2016), baseado no protocolo proposto por Kitchenham e Charters (2007).

A seguir, serão detalhadas as etapas propostas por Neiva e Silva (2016) e como foram aplicadas nesta pesquisa.

1. Definir as questões de pesquisa principais da revisão: Nesta etapa, delineiam-se as principais questões que orientarão a pesquisa. No presente estudo, a

questão central é: "Como o Processamento de Linguagem Natural está sendo aplicado no contexto educacional?"

2. Definir as palavras-chave: Após a definição das questões, são identificadas as palavras-chave essenciais. A partir das questões de pesquisa, as palavras-chave identificadas foram: "Processamento de Linguagem Natural" e "educação".

3. Definir a *string* de busca: A construção da *string* de busca segue uma regra básica: as palavras-chave são separadas e, para cada uma delas, sinônimos relevantes são identificados. Foram definidas, neste estudo, as seguintes *strings*:

natural language processing education

nlp education

natural language processing teaching

nlp teaching

natural language processing learning

nlp learning

processamento de linguagem natural

pln

nlp

4. Definir as bases de busca: A escolha das bases de busca é uma decisão crucial, moldada pela área de foco da revisão sistemática. Para este estudo, foram selecionadas as seguintes bases de dados:

- DBLP (Digital Bibliography & Library Project) - É um serviço bibliográfico que lista artigos de conferências e revistas de ciência da computação, incluindo trabalhos relacionados ao Processamento de Linguagem Natural. Ele é amplamente utilizado por pesquisadores e oferece uma vasta gama de publicações nessa área.

- ACM (Association for Computing Machinery)- A ACM é uma organização líder em computação, e sua biblioteca digital contém uma rica variedade de literatura relacionada à ciência da computação, incluindo PLN.

A combinação dessas bases foi feita para garantir uma ampla cobertura de publicações acadêmicas, proporcionando uma visão abrangente e aprofundada sobre a interseção entre Processamento de Linguagem Natural e educação.

5- Refinamento da *string*: Nesta etapa, a estratégia é testar a *string* de busca previamente definida em uma das principais bases de dados. Após a execução da *string*, é crucial avaliar se os artigos recuperados são relevantes para a pesquisa. Caso os resultados não sejam satisfatórios em relação à relevância, é necessário revisar e calibrar a *string* de busca para melhor atender aos objetivos do estudo.

Durante as buscas utilizando as *strings* "natural language processing learning" e "nlp learning", observou-se a ausência de resultados relevantes para o trabalho. Os artigos recuperados estavam estritamente relacionados ao aprendizado de máquina, sem qualquer ligação com o contexto educacional. Diante dessa constatação, ambas as *strings* foram excluídas do processo de busca.

6- Execução da *string* de busca nas bases: Uma vez definida a *string* de busca, a mesma deverá ser adaptada para cada uma das bases, uma vez que a sintaxe utilizada por cada uma delas é diferente. Foram definidas as seguintes sintaxes de buscas:

Quadro 1: Sintaxe das *Strings*

String de Busca	Base	Critérios
natural language processing education	DBLP	--
nlp education	DBLP	--
natural language processing teaching	DBLP	--
nlp teaching	DBLP	--
"natural language processing" AND "education"	ACM	Palavras-chave
"natural language processing" AND "education"	ACM	Título
"nlp" AND "education"	ACM	Palavras-chave
"nlp" AND "education"	ACM	Título
"natural language processing" AND "teaching"	ACM	Palavras-chave
"natural language processing" AND "teaching"	ACM	Título
"nlp" AND "teaching"	ACM	Palavras-chave
"nlp" AND "teaching"	ACM	Título

Fonte: Elaborado pelos autores

7- Baixar e armazenar o resultado das buscas: Para preservar os dados obtidos durante as buscas, os resultados das principais bases foram exportados em formatos acessíveis. Utilizando os formatos BIB, os dados das buscas realizadas nas bases DBLP e ACM foram catalogados no Zotero e, posteriormente, armazenados em uma planilha.

8- Definir os critérios de inclusão e exclusão: Nesta fase, critérios rigorosos foram estabelecidos para determinar quais artigos avançariam para as etapas subsequentes da revisão. Artigos secundários, resumidos, duplicados e aqueles que não abordavam as aplicações do PLN no contexto educacional foram excluídos. Em contrapartida, foram incluídos apenas estudos primários que aplicassem técnicas específicas de PLN na educação.

9- Seleção de artigos (Primeira etapa: Análise por título e *abstract*): Iniciando com uma análise dos títulos e *abstracts*, a seleção inicial envolveu uma revisão criteriosa. Foram descartados os trabalhos que tratavam exclusivamente de um dos temas do estudo (PLN ou educação), bem como estudos que apenas faziam referências genéricas aos temas, sem aprofundamento em técnicas específicas.

10- Seleção de artigos (Segunda etapa: Análise por Introdução e Conclusão): Para refinar ainda mais a seleção, a análise se estendeu às introduções e conclusões dos artigos previamente selecionados. Utilizando os mesmos critérios da etapa anterior, essa fase visou garantir uma escolha robusta e embasada.

11- Seleção de artigos (Terceira etapa: Leitura completa): Na etapa final, critérios detalhados de qualidade foram definidos (checklist) para uma leitura aprofundada dos artigos. Estabelecendo um ponto de corte, artigos que não atendiam aos requisitos estabelecidos foram excluídos, assegurando a integridade e confiabilidade dos resultados.

12- Extração das respostas relacionadas às questões de pesquisa: Na última etapa, as questões de pesquisa foram respondidas por meio da análise minuciosa dos artigos selecionados anteriormente. Os dados extraídos foram organizados em uma planilha, incluindo informações como título, ano de publicação, país de origem, aplicação e técnica envolvida, permitindo uma análise detalhada e estruturada.

3. REFERENCIAL TEÓRICO

O emprego do PLN no âmbito educacional tem sido objeto de crescente interesse e pesquisa. Esta revisão sistemática da literatura visa compreender como o PLN tem sido aplicado, adaptado e refinado para melhorar os processos de ensino e aprendizagem.

Embora exista uma riqueza de literatura que explora diversas facetas do PLN, uma revisão sistemática dedicada a examinar suas aplicações específicas no contexto educacional, especialmente no que diz respeito à interação entre alunos e tecnologia, é notavelmente escassa. Este estudo visa preencher essa lacuna, proporcionando uma análise detalhada das tendências, desafios e inovações recentes relacionadas ao uso do PLN nas práticas pedagógicas.

De acordo com Costa, Ralha, Ralha (2005, p.19), PLN é um campo de estudos multidisciplinar, integrando Linguística, Matemática, Computação, dentre outras disciplinas afins. Sua aplicação em sistemas de cunho educacional diversifica ainda mais o conhecimento envolvido.

Segundo os mesmos autores, “atualmente, nenhum sistema é capaz de reconhecer com precisão a totalidade de uma língua natural, mas pesquisas recentes estão constantemente propondo novos formalismos que procuram suprir as necessidades identificadas nas abordagens anteriores.”. (COSTA; RALHA; RALHA, 2005, p.20).

Dermeval et al. (2019) abordam o mapeamento e a revisão sistemática da literatura da informática na educação, apontando que “revisão sistemática procura minimizar erros sistemáticos e aleatórios buscando definir claramente o procedimento a ser adotado na condução do levantamento do estado da arte de um tópico de pesquisa. Além disso, estes estudiosos ressaltam que qualquer revisão ou mapeamento sistemático na área de Informática na Educação deve, obrigatoriamente, incorporar pelo menos uma das seguintes bibliotecas digitais, tais como *ScienceDirect*, *ISI Web of Science*, *Scopus*, *SpringerLink*, *ACM Digital Library*, *IEEE Xplore*, *Compendex (Engineering Village)*, além de outras bases de outras áreas de conhecimento, dependendo da interdisciplinaridade do tópico da revisão sistemática.

No trabalho de Brasil (2018), é realizada uma revisão sistemática de literatura que tem como objetivo “identificar como a análise de aprendizagem tem sido aplicada a sistemas de apoio tecnológico para ambientes de ensino de instituições brasileiras”. É observado no trabalho que o protocolo para a execução da RSL, também é baseado na proposta de Kitchenham e Charters (2007), onde os autores afirmam que “a

principal diferença entre os métodos é que, ao invés de utilizar um processo de busca manual, foi utilizada uma busca automatizada para identificação dos trabalhos candidatos a pesquisa.”

Os autores Kuyven et al (2018) realizaram uma revisão sistemática de literatura relacionada ao uso de *chatbots* para fins educacionais e constataram que

apesar de a maior parcela de estudos serem voltados para o uso dos chatbots em áreas do ensino específicas, principalmente na área da Ciência da Computação, também existem pesquisas que buscam construir um sistema flexível, com uma lógica genérica, que pode servir de modelo ou se encaixar em diversas disciplinas, ou que tenha um propósito educacional.(...) De acordo com a análise dos artigos, entre os principais fatores que impedem a difusão do uso de chatbots na educação estão: a necessidade de uma base de conhecimento considerável para uma conversa satisfatória; a maior complexidade e imprevisibilidade do fluxo de diálogo entre o agente e o aluno em relação a uma conversação casual. (KUYVEN et al, 2018)

Foram encontrados artigos que realizam revisão sistemática de literatura sobre processamento de linguagem natural em outras áreas, como o trabalho dos autores De Almeida Bordigon et al (2018) que “apresenta técnicas de PLN aplicadas às fases do ciclo de vida *Business Process Management* da identificação de processos, descoberta de processos e análise de processos, bem como ferramentas para apoiar a descoberta de processos.”

Sobre conceitos relativos ao processamento de linguagem natural que foram utilizados no presente trabalho, é importante definir alguns como a *tokenização* que consiste em transformar o texto em apenas um conjunto de palavras (*tokens*) de forma limpa, eliminando caracteres inválidos e espaços eventualmente presentes no texto. (SILVA; MARTINS, 2017), a remoção de *stopwords* que elimina palavras irrelevantes em um processo de recuperação de informação, a análise sintática que consiste em analisar e estruturar frases, a análise semântica que realiza a verificação de aspectos relacionados ao significado das expressões e etiquetador que “recebe um texto e retorna o texto com suas palavras etiquetadas (classe gramatical, gênero, número etc).” (TAVARES et al, 2017).

Outras terminologias que precisam ser explicadas são as técnicas de aprendizado de máquina que podem ser classificadas em supervisionada, onde o aprendizado ocorre por meio da observação de exemplos rotulados, ou seja, a partir de dados de entrada e saída, não-supervisionada, na qual o algoritmo não recebe valores específicos de saída, apenas de entrada, com isso, a aprendizagem

deve ser capaz de identificar padrões e extrair conhecimento e aprendizagem por reforço, onde o algoritmo aprende a partir do reforço. (NAU et al, 2020).

4. ANÁLISE DOS DADOS

A execução das *strings* de busca ocorreu em 14 de abril de 2023, resultando em 137 trabalhos identificados. Após a eliminação de duplicatas, restaram 108 resultados, que foram submetidos à primeira etapa de seleção. Durante essa fase, os títulos e resumos dos trabalhos foram analisados, levando à exclusão de 75 artigos com base nos critérios de inclusão e exclusão. Dessa forma, 33 trabalhos prosseguiram para a segunda etapa de seleção.

Na fase subsequente, a leitura da introdução e do resumo permitiu excluir 9 artigos, culminando na análise completa de 24 trabalhos. Destes, 16 atenderam aos critérios de inclusão definidos. Estes trabalhos foram minuciosamente examinados, e as respostas às questões de pesquisa foram meticulosamente extraídas.

Uma análise inicial concentrou-se nas diversas aplicações do processamento de linguagem natural no contexto educacional, sendo que um único trabalho frequentemente abordava múltiplas aplicações. Destaca-se que 6 dos 16 artigos investigados se concentravam na correção de textos dissertativos, uma aplicação voltada para simplificar e acelerar as tarefas dos educadores. Além disso, a obtenção do perfil discente foi uma aplicação comum, revelando informações cruciais sobre as dificuldades dos alunos em áreas específicas e permitindo ajustes no planejamento educacional para melhorar o desempenho dos estudantes.

Outro tema recorrente nos trabalhos foi o uso de mapas conceituais, que apresentou duas abordagens distintas. Uma delas focou na análise e correção dos mapas, destinada aos professores, enquanto a outra envolveu a criação de mapas conceituais para apoiar os alunos em seu processo de aprendizado.

Além disso, outros contextos educacionais também foram explorados, incluindo a verificação de plágio (em 2 trabalhos), o ensino de português e Libras, a promoção de alternativas para o aprendizado e a identificação de ambiguidades em currículos, cada um abordado em um trabalho específico. Estas análises proporcionaram uma compreensão mais profunda das diversas maneiras pelas quais o processamento de linguagem natural está sendo integrado de forma significativa e inovadora no campo educacional.

Quadro 2: Aplicações do PLN no contexto educacional

Aplicação	Quantidade de artigos
Correção de textos dissertativos	6
Obtenção do perfil discente	3
Análise e/ou correção de mapas conceituais	2
Geração automática de questões e/ou mapas conceituais	2
Verificação de plágio	2
Classificação do material educacional pesquisado	1
Ensino de português e libras	1
Promoção de alternativas para o aprendiz	1
Verificação de ambiguidades em currículos	1

Fonte: Elaborado pelos autores

Verifica-se no Quadro 3 que a *tokenização* foi a técnica mais utilizada nos trabalhos analisados, sendo mencionada em 10 artigos. Uma justificativa para tal resultado é o fato da *tokenização* fazer parte da etapa de pré-processamento e assim precisar ser executada antes da aplicação de outras técnicas. A técnica de remoção de stopwords, apesar de englobar o pré-processamento, não foi citada com expressividade como a *tokenização*. Uma hipótese para esta análise é que os autores podem não ter deixado explícito a utilização da técnica que provavelmente foi aplicada em quase todos os artigos.

Ao analisar o Quadro 3, nota-se que técnicas como análises sintática e semântica podem ser observadas em 8 e 5 artigos, respectivamente e ocorrem após a aplicação de *tokenização*. Já as técnicas de aprendizagem supervisionada, similaridade (algoritmos baseados em distribuição de palavras) e utilização de dicionários léxicos são utilizadas em 5 dos 16 trabalhos.

Em relação às técnicas de aprendizagem semi-supervisionada, aprendizagem supervisionada – Regressão e similaridade - algoritmos baseados em tesouros, infere-se que são utilizadas com uma frequência mais baixa no ambiente educacional.

Quadro 3: Técnicas do PLN utilizadas nos artigos selecionados

Técnicas	Quantidade de artigos
<i>Tokenização</i>	10

Análise sintática	8
Análise semântica	5
Aprendizagem supervisionada - Classificação	5
Similaridade - Algoritmos baseados em distribuição de palavras	5
Utilização de dicionários léxicos	5
Aprendizagem não supervisionada - clusterização	4
Remoção de stopwords	4
Etiquetagem	3
Lematização	3
Ngramas/Modelo de Markov	2
Aprendizagem semi-supervisionada	1
Aprendizagem supervisionada - Regressão	1
Similaridade - Algoritmos baseados em tesauros	1

Fonte: Elaborado pelos autores

Quanto ao elo das técnicas e aplicações, foi elaborada a Quadro 4 com os resultados analisados, onde a relação dos números encontrados no quadro e as aplicações são mostradas abaixo:

- 1 = Correção de textos dissertativos
- 2 = Obtenção do perfil discente
- 3 = Análise e/ou correção de mapas conceituais
- 4 = Geração automática de questões e/ou mapas conceituais
- 5 = Verificação de plágio
- 6 = Classificação do material educacional pesquisado
- 7 = Ensino de português e libras
- 8 = Promoção de alternativas para o aprendizado
- 9 = Verificação de ambiguidades em currículos

Os trabalhos que têm como finalidade a correção de textos dissertativos (1) usaram 12 das 14 técnicas catalogadas, sendo a aplicação que mais utilizou os recursos elencados. As únicas técnicas que não foram empregadas na primeira aplicação foram a lematização e a similaridade - algoritmos baseados em tesauros. Dada a complexidade da tarefa de correção de textos dissertativos já era esperada a manipulação de tantas técnicas. Obtenção do perfil discente foi a segunda aplicação com uso de mais técnicas, não sendo observadas apenas as técnicas de

aprendizagem semi-supervisionada, lematização, ngramas/modelo de Markov, similaridade - algoritmos baseados em tesouros.

Nos artigos que são abordadas a classificação do material educacional pesquisado e promoção de alternativas para o aprendizado, foram empregadas somente uma técnica de forma explícita em cada artigo. Acredita-se que diversas técnicas tenham sido utilizadas nos referidos trabalhos, entretanto não foram mencionadas nos *papers*.

Identificou-se que as técnicas análise sintática e *tokenização* são as mais mencionadas nos artigos selecionados.

Quadro 4. Técnicas x Aplicações do PLN

Técnica	1	2	3	4	5	6	7	8	9
Análise semântica	x	x			x				x
Análise sintática	x	x	x	x			x	x	
Aprendizagem não supervisionada - clusterização	x	x			x				x
Aprendizagem semi-supervisionada	x								
Aprendizagem supervisionada - Classificação	x	x	x	x		x			
Aprendizagem supervisionada - Regressão	x	x							
Etiquetagem	x	x					x		
Lematização			x				x		
Ngramas/Modelo de Markov	x				x				
Remoção de s stopwords	x	x			x				
Similaridade - Algoritmos baseados em distribuição de palavras	x	x		x	x				
Similaridade - Algoritmos baseados em tesouros									x
<i>Tokenização</i>	x	x		x	x		x		x
Utilização de dicionários léxicos	x	x			x		x		x

Fonte: Elaborado pelos autores

A última análise incluiu uma comparação das publicações por país e ano. Dos 16 artigos selecionados, 5 foram publicados nos Estados Unidos, 3 na China, 2 no Japão, 2 na Índia, enquanto Brasil, Holanda e Rússia tiveram 1 publicação cada. A distribuição geográfica dos artigos destaca a presença significativa dos Estados Unidos e países asiáticos. As publicações abrangem um período de 2004 a 2023, sendo notável que a maioria delas (7) foi lançada nos últimos 4 anos, indicando uma tendência recente e crescente de estudos nesta área. Esse cenário reflete não apenas o avanço tecnológico, mas também a adaptação contínua dos sistemas educacionais às mudanças sociais e às necessidades dos alunos, ressaltando a importância cada

vez maior do PLN como uma ferramenta vital para inovação e melhoria na educação global.

5. CONSIDERAÇÕES FINAIS

Com base nos resultados obtidos, torna-se evidente que a maioria dos trabalhos analisados está centrada em pesquisas destinadas a auxiliar os professores. Apenas um trabalho contribui diretamente para a equipe pedagógica, enquanto cinco artigos oferecem aplicabilidades práticas para o cotidiano dos alunos.

Observa-se que o campo do processamento de linguagem natural na educação ainda é relativamente pouco explorado na literatura. No entanto, nos últimos quatro anos, tem-se notado um aumento considerável no interesse por esse tema. Este crescimento é, em parte, resultado do avanço tecnológico acelerado e das novas demandas impostas à educação devido à pandemia de Covid-19 que assolou o mundo entre 2020 e 2022. Diante desse cenário desafiador, é esperado que mais pesquisas sobre o processamento de linguagem natural na educação sejam publicadas em volumes significativamente maiores nos próximos anos.

É importante salientar que, devido às limitações de tempo para o desenvolvimento deste estudo, análises mais aprofundadas, como comparações detalhadas de aplicações, técnicas, países, anos e autores referenciados, não puderam ser realizadas. Questões importantes, como o nível de ensino para o qual os artigos são aplicáveis, as ferramentas, linguagens e corpora utilizados, não foram detalhadas na maioria dos artigos selecionados, o que impediu uma interpretação mais aprofundada desses aspectos.

Outra limitação encontrada neste trabalho é que a pesquisa se restringiu às ferramentas mencionadas nos artigos selecionados, excluindo, assim, outras técnicas, ferramentas e possíveis aplicações dentro do escopo da temática.

Para futuras pesquisas, pretende-se realizar buscas mais detalhadas em mais bases de dados, além de realizar seleções e interpretações mais aprofundadas, a fim de explorar com maior precisão as nuances do processamento de linguagem natural na educação.

6. REFERÊNCIAS

BARBOSA, M. L.; AMARAL, S. F. do. Aplicativos e gamificação na educação: possibilidades e considerações / Applications and gamification in education: possibilities and considerations. **Brazilian Journal of Development**, [S. l.], v. 7, n. 3, p. 23974–23987, 2021. DOI: 10.34117/bjdv7n3-210.

BRASIL, P. C.; MEDEIROS, Tainá Jesus; NUNES, Isabel Dillman. O uso de Learning Analytics no Brasil: uma revisão sistemática da literatura. In: **Revista Tecnologias na Educação – Ano 10 – Número/Vol.26**. Edição Temática VIII – III Congresso sobre Tecnologias na Educação, 2018.

COSTA, F. V; RALHA, José Carlos L.; RALHA, Célia G.. Aprendizagem de Língua Assistida por Computador: Uma Abordagem Baseada em HPSG. In: **XVI Simpósio Brasileiro de Informática na Educação**, 2005.

DE ALMEIDA BORDIGON, Ana Cláudia; THOM, Lucinéioa Heloisa ; SOARES SILVA, Thanner ; STEIN DANI, Vinicius ; FANTINATO, Marcelo ; BORGES PEREIRA, Renato Cesar. Processamento de Linguagem Natural na Identificação e Modelagem de Processos de Negócio: Uma Revisão Sistemática da Literatura . In: SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSI), 14. , 2018, Caxias do Sul. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2018 . p. 191-198.

DERMEVAL, Diego; COELHO, Jorge A. P. de M.; BITTENCOURT, Ig I. **Mapeamento Sistemático e Revisão Sistemática da Literatura em Informática na Educação**. In: JAQUES, Patrícia Augustin; SIQUEIRA; Sean; BITTENCOURT, Ig; PIMENTEL, Mariano. (Org.) Metodologia de Pesquisa Científica em Informática na Educação: Abordagem Quantitativa. Porto Alegre: SBC, 2020.

FIGUEIREDO, M.; PAZ, T.; JUNQUEIRA, E. Gamificação e educação: um estado da arte das pesquisas realizadas no Brasil. In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 4., 2015, Maceió. **Anais. Maceió: Anais**, 2015. p. 1155 – 1163

GALVÃO, M. C. B.; RICARTE, L. M. Revisão sistemática da literatura: conceituação, produção e publicação. In: **Logeion: Filosofia da Informação**, v. 6, n. 1, p. 57-73, 2019.

KITCHENHAM, B. and CHARTERS, S.. **Guidelines for performing systematic literature reviews in software engineering** (version 2.3), 2007.

KUYVEN, Neiva Larisane; ANTUNES, Carlos André; VANZIN, Vinicius João de Barros; SILVA, João Luis Tavares; KRASSMANN, Aliane Loureiro; TAROUÇO, Liane Margarida Rockenbach. Chatbots na educação: uma Revisão Sistemática da Literatura. **Revista Novas Tecnologias na Educação**, Porto Alegre, v. 16, n. 1, 2018. DOI: 10.22456/1679-1916.86019.

NAU, Jonathan; DAZZI, Rudimar Luís Scaranto; HAENDCHEN FILHO, Aluizio; FERNANDES, Anita. Processamento do Discurso em Textos Dissertativos-Argumentativos: Uma Abordagem Baseada em Mineração de Argumentos e Aprendizado Supervisionado de Máquina. *In: SEMINÁRIO INTEGRADO DE SOFTWARE E HARDWARE (SEMISH)*, 47.2020, Cuiabá. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 48-59. ISSN 2595-6205. DOI: <https://doi.org/10.5753/semish.2020.11316>.

NEIVA, Frâncila Weidt; SILVA, Rodrigo Luis de Souza. **Revisão Sistemática da Literatura em Ciência da Computação -- Um Guia Prático**, 2016.

SILVA, S. B.; MARTINS, C. A.. Pré-processamento textual para a extração de informação em bases de patentes. *In: Anais da Escola Regional de Informática da Sociedade Brasileira de Computação (SBC) – Regional de Mato Grosso*, v. 7, 2017.

TAVARES, Orivaldo; REINOSO, Luiz; ALMEIDA, Wanderson. CAP-APL: Plataforma para criação e uso de arquiteturas pedagógicas para aprendizagem de Português e Libras. *In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*