

INTELIGÊNCIA ARTIFICIAL, AS LIMITAÇÕES DA LINGUAGEM NATURAL E SEUS IMPACTOS: ESTADO DA ARTE

Datan Azevedo Aguiar^{1*}
Wesley Jorge Gomes de Souza Santana²

¹Bacharel em Sistemas da Informação, Ibiapina, Ceará, Brasil. *Autor para correspondência.
datan_x@yahoo.com.br

²Bacharel em Sistemas da Informação / Ibiapina, Ceará, Brasil, wesleyjdesouza@gmail.com

RESUMO: Atualmente, o mundo vive uma época de alto nível de comunicação, no qual todas as notícias circulam de forma rápida, gerando, cada vez mais, informação e surgindo, assim, a necessidade de respostas frágeis. Informações de todo o tipo estão disponíveis a um simples *click* em um *smartphone* ou qualquer dispositivo com um sistema responsivo. A *internet* está presente em quase todo aparelho fornecendo aos seus usuários um *feedback* de suas ações e otimizando suas vidas, podendo, assim, ser observada na sua vida particular e/ou no seu ambiente de trabalho. O presente artigo consiste em uma revisão de literatura do tipo sistemática, na qual tem como objetivo aprofundar o conhecimento acerca da linguagem natural e do Processamento de Linguagem Natural, com suas ferramentas e ontologias, sanando todas as dúvidas existentes acerca dos *softwares*, além de entender seu objetivo, suas estruturas e suas limitações. A estrutura desta pesquisa, deu-se, como foco principal de busca, no sítio do *Google Acadêmico*, nas áreas de ciências da informação e demais associações. Foram selecionados 55 documentos e classificados em artigos de periódicos, monografias, trabalhos de conclusão de curso, dissertações de mestrado, teses de doutorado, entre outras publicações, nacionais e internacionais. Sobre o estado da arte, foi colocado como análise principal o trabalho de Almeida e Dias (2019), que fez uma abordagem brilhante da análise de domínio, no campo da ciência de informação brasileira. Ele apresenta a questão acerca do que é domínio, que segundo suas pesquisas, são as construções de sistemas de organização da informação e que compartilham conhecimentos, aproveita-se, então, para enfatizar a necessidade de entender como um determinado domínio, chamado de PLN, auxiliou diversas áreas do conhecimento a alcançarem um resultado, sobre estas pesquisas, em especial, aquelas que utilizaram alguma ferramenta da Inteligência Artificial. Em suma, é possível concluir que o PLN é uma vertente da IA, no qual auxilia computadores a entender, interpretar e manipular a linguagem humana, resultando, assim, de diversas disciplinas, incluindo a ciência da computação e a linguística computacional, que buscam preencher a lacuna entre a comunicação humana e o entendimento dos computadores.

Palavras chaves: Ciência da computação, informação, linguística computacional, Processamento de Linguagem Natural, ontologias, *softwares*.

ARTIFICIAL INTELLIGENCE, THE LIMITATIONS OF NATURAL LANGUAGE AND ITS IMPACTS: STATE OF THE ART

ABSTRACT: *Currently, the world lives in a time of high level of communication, in which all the news circulates quickly, generating, more and more, information and, thus, the need for fragile answers. Information of all kinds is available at a simple click on a smartphone or any device with a responsive system. The internet is present in almost every device, providing its users with feedback on their actions and optimizing their lives, thus being able to be observed in their private lives and/or in their work environment. This article consists of a systematic literature review, which aims to deepen knowledge about natural language and Natural Language Processing, with its tools and ontologies, solving all existing doubts about the software, in addition to understanding its purpose, its structures and its limitations. The structure of this research took place, as the main focus of the search, on the Google Scholar website, in the areas of information sciences and other associations. 55 documents were selected and classified into journal articles, monographs, course conclusion works, master's dissertations, doctoral theses, among other national and international publications. Regarding the state of the art, the main analysis was the work of Almeida and Dias (2019), who made a brilliant approach to domain analysis in the field of Brazilian information science. He presents the question about what is a domain, which, according to his research, are the constructions of information organization systems that share knowledge, taking advantage, then, to emphasize the need to understand how a certain domain, called PLN, helped several areas of knowledge to achieve a result, on these researches, in particular, those that used some tool of Artificial Intelligence. In short, it is possible to conclude that NLP is a branch of AI, in which it helps computers to understand, interpret and manipulate human language, thus resulting from several disciplines, including computer science and computational linguistics, which seek to fill the gap between human communication and computer understanding.*

Keywords: *Computer science, information, computational linguistics, Natural Language Processing, ontologies, software.*

INTRODUÇÃO

Atualmente, o mundo vive uma época de alto nível de comunicação, no qual todas as notícias circulam de forma muito rápida, gerando, cada vez mais, informação e surgindo, assim, imediatamente, a necessidade de respostas frágeis. Informações de todo o tipo estão disponíveis a um simples *click* ou *touch* em um *smartphone* ou qualquer eletrônico com um sistema responsivo. A *Internet* das Coisas (IOT – *Internet Of Things*) está presente em quase todo aparelho, como geladeiras, micro-ondas, assistentes pessoais e até banheiros, fornecendo aos seus usuários um *feedback* de suas ações e otimizando suas vidas, podendo, assim, ser observada na sua vida particular e/ou no seu ambiente de trabalho.

O mundo das corporações está gerando cada vez mais dados e utilizando aqueles gerados pelas redes sociais como *facebook*, *instagram*, entre outros. Tecnologias como *Big Data* e *Data Mining* estão se tornando cada vez mais fontes de rendas necessárias, sendo um alvo a ser explorado pelas empresas. Contudo, as limitações do ser humano para analisar dados são facilmente verificáveis, porém, ao se utilizar uma tecnologia adequada, literalmente pode-se chegar a um resultado confiável, acima de tudo, de grande utilidade para estas corporações. As ferramentas adequadas para análise de grande quantidade de dados são *softwares* (SW) que utilizam Inteligência Artificial (IA – *Artificial Intelligence*), os quais provaram ser úteis e são continuamente estudados e desenvolvidos.

É exatamente nesse campo da IA, que este trabalho planeja fazer, em seu desenvolvimento presente e futuro, uma análise qualitativa, por meio da leitura de diversos artigos científicos, desenvolvendo um diálogo com seus autores para entender como funciona a linguagem natural, bem como suas ferramentas

de processamento e a forma em que são aplicadas nesses SW. Deste modo, se propõe a buscar seus significados e, posteriormente, os problemas e de que forma se estruturam, assim, a máquina e o ser humano irão estabelecer uma comunicação, surgindo assim, a grande questão, mas como entenderão?

As pessoas estão cada dia consumindo mais e mais tecnologias. Desta forma, pode-se observar que, dentro deste crescente universo em que o desenvolvimento de tecnologias é constante, podemos notar a presença da IA, que são SW que possuem a capacidade de responder e compreender a linguagem humana de forma autônoma, como pode-se observar em aplicativos de bancos, lojas e no *WhatsApp*. Estes são interativos e estão no alvo de diversos consumidores, pode-se então perceber um novo nicho de mercado, no qual tanto a dona de casa quanto as grandes empresas, buscam uma solução adequada a sua realidade.

As limitações linguísticas da IA podem afetar, não apenas, as pessoas comuns e suas necessidades cotidianas, mas também, as grandes corporações e sua renda. Os problemas podem afetar os seus funcionários, afetando diretamente suas interações, seja com os *frameworks* de gerenciamento, sejam com os projetos, ou pessoas.

A resposta se dá na compreensão, de forma adequada, ou seja, bem estruturada, nos quais se encontram os limites, que podem ser de uso, de produção ou manutenção de uma IA. Permitindo, assim, seu uso qualificado pelos profissionais envolvidos.

Agora será colocado em foco o seguinte exemplo, ao conversar com o assistente de voz como o do *Google*, por exemplo, as limitações do SW são claras, sendo elas: falta de compreensão do Processamento de Linguagem Natural (PLN) pelos programadores; formação de um dicionário lexical incompleto; estabelecimento das regras semânticas de forma falha; falta de uma construção adequada para captar ironia, sarcasmo ou outras expressões.

Dessa forma, é possível entender o objetivo do programa e tudo o que está sendo oferecido, como essa tecnologia pode compreender o que foi solicitado devido as limitações, e a grande questão, quais estruturas regem estes SW e seu banco de dados. Será que eles possuem uma margem na qual se permita dar uma resposta ou informação errada? Para se obter a resposta dessas perguntas, deve ser entendido a estrutura de linguagem natural e do PLN e suas ferramentas, bem como suas ontologias.

O presente artigo consiste em uma revisão de literatura do tipo sistemática, na qual tem como objetivo aprofundar o conhecimento acerca da linguagem natural e do PLN, com suas ferramentas e ontologias, sanando, assim, todas as dúvidas existentes acerca dos SW, além de entender seu objetivo, suas estruturas e suas limitações.

REVISÃO DE LITERATURA

Revisão Conceitual

O mundo tem acelerado a produção de diferentes tipos de dados, são uma infinidade de empresas ou corporações em diferentes locais do mundo captando, direta ou indiretamente, informações. Todo o tipo de dados de usuários, como gostos, desejos, curtidas, postagens, conversas, entre outras informações, são continuamente apresentadas a diferentes soluções digitais para filtrar, analisar e otimizar o uso dos mesmos.

Segundo Alvarenga (2019), o português possui uma reduzida quantidade de dados informatizados que podem ser anotados, ficando claro a qualidade inferior de ferramentas textuais de tal forma que os tradutores digitais, focados em mineração de opinião, em fóruns, por exemplo, se apresenta como um mercado inicial e com grande potencial de crescimento.

Uma área tão vasta e inexplorada logo atrai a comunidade científica que, de maneira dedicada, traz luz ao uso destas ferramentas de transformar dados em informação. Cada artigo produzido, nos últimos anos, sobre inteligência artificial, dialoga com problemas e soluções sobre os limites destes SW, desde sua construção sintática ao entendimento da linguagem natural. Pode ser relacionado a enorme quantidade de informações disponíveis na forma de textos diretamente a um potencial de conhecimento, aplicável por humanos no estudo da linguagem natural ou pelas máquinas com o processamento da mesma (KULTZAK, 2016).

O mundo da tecnologia, em sua constante e quase imparável evolução, apresenta inimagináveis formas de facilitar o dia a dia. Simplesmente, poder criar um novo hábito com consequências cada vez mais desconhecidas, e interações mais surpreendentes, dentre elas a IA, e com certeza uma das mais novas áreas da pesquisa intelectual, mas como diria Coppin (2015), suas bases surgem há milhares de anos.

Segue-se a afirmação feita por Coppin e alguns questionamentos foram respondidos sobre as diferentes áreas fundamentais. Desta forma, pode-se questionar ou afirmar com maior embasamento, partindo da construção do vocabulário, gramática e do uso das palavras dentro da epistemologia informacional,

buscando identificar as possibilidades de reconhecimento das informações (GRACIOSO; SALDANHA, 2016).

Ainda conforme os mesmos autores, os conhecimentos devem responder a primeira questão: o que é inteligência? Coppin (2013 p. 4) nos diz, inteligência é a capacidade de lidar com novas situações, de solucionar problemas. Ao responder a essas questões, é possível, também, ter a capacidade de criar e constituir um processo permanente de revisão e crítica do que foi visto na produção científica.

Inteligência Artificial

Segundo Coppin (2015), a IA, em seu estado mais básico, pode ser utilizada para resolver questionamentos, problemas simples, resolução de labirintos ou buscas. Em outras palavras, IA envolve métodos baseados no comportamento inteligente de humanos e de outros animais para solucionar problemas complexos.

Deste ponto de entendimento, podemos começar a analisar sistemas que podem perceber declarações humanas, ou segundo Coppin (2015), aqueles que são capazes de extrair algum significado de declarações humanas e agir baseados nestas. Sempre partindo da noção de que a linguagem é, em todas as suas concepções e, primordialmente, o sistema de informação mais primário que possuímos (GRACIOSO; SALDANHA, 2016).

Sabemos que a linguagem é o primeiro passo para a compreensão da realidade e que também é, em suas diferentes concepções e variedades de formas, nossa principal maneira de declarar o mundo. Dentro do universo da ciência da informação, pode-se definir, de forma mais clara, o uso de SW de IA, cita-se assim, Gracioso e Saldanha (2016), sobre a questão nas narrativas do pragmatismo informacional. Necessariamente, há o desligamento de uma noção definitiva de informação e a busca desta, a partir dos contextos dos usos dos saberes.

Dessa forma, é possível perceber o crescimento contínuo do entendimento de IA e que tais sistemas podem ser projetados, como diria Coppin (2015, p. 4), não para se comportar de modo inteligente, mas, simplesmente, para dar conta de alguma função útil ou, como supracitado, de forma pragmática.

Além destas questões envolvendo linguagem e resolução de problemas, a maioria dos pesquisadores podem dividir as técnicas de IA em fortes e fracas, no qual pode claramente utilizar a declaração de que os seguidores da IA forte acreditam que, dispondo de um computador com suficiente capacidade de processamento e, fornecendo a ele suficiente inteligência, pode-se criar um computador que possa literalmente pensar e ser consciente do mesmo modo que um humano é consciente. Por outro lado, a IA fraca é, simplesmente, a visão de que comportamento inteligente pode ser modelado e utilizado por computadores para solucionar problemas complexos (COPPIN, 2015).

Devemos lembrar que IA forte ou fraca é diferente de métodos fortes e fracos em uso para o funcionamento das IA. Os métodos fracos podem ser entendidos, como regras gerais, lógica em que se segue as questões *true or false* e raciocínio automatizado que podem ser usados para resolver um grande conjunto de problemas. Já os métodos fortes, demandam conhecimentos sofisticados sobre a realidade e seus problemas, necessitando, assim, da metodologia criada pelos fracos (COPPIN, 2015).

Dessa forma, estrutura-se de forma básica o que se pode entender com IA levando-se em consideração a fala de Russell e Norvig (2013) sobre a diferença entre humano e máquina.

Diante disso, é possível observar os tópicos da abordagem que o teste de *Turing Total* exige, sendo eles: PLN, representação de conhecimento, raciocínio automatizado, aprendizado de máquina, visão computacional e robótica. O mesmo tem por objetivo descobrir se uma máquina pode ter comportamento similar ao humano. Para ilustrar, de forma simples, como este ocorre, será utilizado o exemplo original. Durante um jogo, um ser humano será o *player*, no qual entrará em uma conversa com outro humano e uma máquina, todos utilizando linguagem natural. Os três participantes estão isolados, conversando por texto, se o player não for capaz de distinguir quem é o humano ou não, então a máquina passou no teste.

Linguagem Natural

Comunicar e se fazer entender é um princípio básico do entendimento entre as pessoas, mas mesmo entre elas, surgem enormes dificuldades de compreensão que são amenizadas por regras e estruturas formais. Logo, tem-se o problema dentro da IA que é a compreensão dessa estrutura usada cotidianamente pelos seres humanos, mas que será esclarecido com as referências dos autores a seguir.

Conforme a Stefanini Group (2019), a comunicação em suas diferentes aplicações é um meio de aprender e transmitir conhecimentos. Graças a ela, é possível que informações, independentes de suas

formas, sejam desejos, notícias ou qualquer outra coisa que se busque, sejam passados para outros ou recebidos, bem como o contrário se apresenta como verdadeiro.

De acordo com o pensamento supramencionado sobre o PLN, podem ser configuradas como técnicas para analisar e representar naturalmente textos, isto realizado com o propósito de alcançar um processamento de linguagem semelhante ao humano. Dentro do universo da IA, as decisões são tomadas levando em consideração os significados na linguagem humana e o PLN. É uma área que permite a IA ter capacidade de ler, entender e derivar linguagens humanas.

Seguindo esse pensamento, ao se analisar o mundo entre humanos e máquinas, pode-se afirmar que as linguagens naturais, como o português, mandarim ou inglês, evoluem, como por exemplo as gírias, jargões ou outras variantes da linguagem informal. Adiciona-se ainda os sotaques entre outras transformações constantes.

Como foi abordado por Simões e Farinha (2010), a língua portuguesa tem evoluído, desta forma, está sujeita a várias reformas, as quais foram feitas no recente acordo ortográfico atualizando as normas que a regem.

Deste modo, pode-se observar que a linguagem natural é bem diferente das linguagens artificiais como a lógica matemática e a programação que são estabelecidas por estruturas bem rígidas. A maioria dos estudos em linguagem natural é baseada em construções matemáticas, para modelar os relacionamentos existentes nos textos estudados. As técnicas consistem, em sua ampla maioria, na análise do número de incidências e da proximidade das palavras em uma mensagem (PEREIRA; RIGO).

Quanto a essas mesmas palavras ou mensagens, segundo Gracioso e Saldanha (2016), os limites do universo de uma escola científica, filosófica ou de programação são os limites de sua linguagem. Com esta compreensão sobre limite, é possível entender a relação entre o homem e seu conhecimento ou o homem e sua conversa. Como por exemplo: um indivíduo fala com sua assistente pessoal, a Siri ou Cortana, nesse diálogo pode-se minimizar alguns prováveis problemas de significado, utilizando um conjunto de regras. Separando e definindo as sentenças qualificadoras (PEREIRA; RIGO, 2013), aproxima-se, desta forma, as duas linguagens.

Para Ludwig Wittgenstein (2013), os limites do meu mundo são os limites da minha linguagem. A busca pela compreensão e significação adequadas aumentaram os resultados positivos de compreensão da linguagem natural pelos SW. Mas, se for levado em consideração apenas a análise de relevância, logo, ela será limitada pela presença de palavras chaves, abandonando a estrutura frasal do texto. Portanto, para as ferramentas de IA uma frase com um conteúdo conceitualmente incorreto, pode apresentar um alto fator de relevância o que poderia levar a resultados incorretos (PEREIRA; RIGO, 2013).

Chega-se, assim, a possível conclusão que para a linguagem natural ser captada de forma adequada, pelos SW de IA, inúmeras variáveis devem ser levadas em consideração. Segundo Coppin (2015), linguagens humanas são extremamente complexas e cheias de armadilhas que quase parecem feitas para pegar qualquer um, seja humano ou computador, inexperiente no uso da linguagem.

Processamento de Linguagem Natural

Em posse de todas essas afirmações e problemas, ficou claro que a linguagem natural, como subárea da IA, é um campo necessário, visto que a máquina necessita interpretar a fala. Apenas o gramatical é insuficiente, ela deve levar em consideração o tom de voz, o duplo sentido, entre outros sistemas de significação, pertencentes a linguagem humana.

Pode-se afirmar que, alguns níveis de processamentos são necessários para o PLN, como: fonológico, interpretação dos sons; morfológico, composição das palavras; lexical, interpreta os significados das palavras; sintático, composição da frase; semântico, significado completo da frase; discurso, observa o texto como um todo, e, por último, pragmático, que busca o que existe nas entrelinhas (STEFANINI, 2019).

Segundo Bulegon e Moro (2010), o processamento de linguagem natural trata-se de um mecanismo criado, não só para extrair as informações de textos, mas que, também, pode facilitar a entrada de dados nos sistemas e a estruturação dos mesmos, permitindo, assim, o uso de textos e escritos, tanto antigos quanto recentes, as abordagens podem variar, sendo elas: a simbólica, que utiliza regras linguísticas estruturadas sem ambiguidades; a estatística, que utiliza modelos matemáticos; a conexionista, mistura o estatístico com representações do conhecimento e a híbrida que combina as demais. Proporcionando, assim, uma forma ilimitada, sendo seu uso amplamente recomendado (STEFANINI, 2019).

Para Russell e Norvig (2013), um texto escrito é composto, quase que em sua totalidade, por caracteres. Os quais podem ser identificados como letras, dígitos, pontuação, caracteres especiais e espaços. Uma sequência de símbolos de comprimento n , chamado de algoritmo de n -grama, é uma coleção de n letras. Para uma IA, como o google tradutor, identificar um determinado idioma em língua natural, o SW utiliza

o n-grama para identificar alguns caracteres, que juntos pertencem, exclusivamente, a um idioma, alguns exemplos da língua inglesa são: *and, the* e *of*.

O que pode ser feito com este modelo de caracteres de n-gramas? O sistema calcula um vetor para cada idioma, no qual armazena informações sobre quantas vezes cada n-grama ocorre naquele idioma. Para se obter a resposta, os softwares de reconhecimentos de linguagens terão uma alta taxa de reconhecimento de qual linguagem natural está escrita nele (COPPIN, 2015).

O PLN também pode ser utilizado para encontrar *spams* ou qualquer outra estrutura textual, utilizando o mesmo cálculo e levando em consideração estruturas chaves, por meio de um modelo de aprendizagem supervisionada.

Para a boa compreensão deste aprendizado, é importante prever o comportamento da máquina, no qual ocorre por meio de variáveis. Algumas destas são independentes, como exemplo o histórico escolar, e as dependentes, a nota do Enem, por exemplo. A principal característica dá-se quando, os dados utilizados para treiná-las, contém a variável dependente, resultante das variáveis independentes observadas. Nesse caso, é dito que os dados são anotados com as respostas ou classes a serem previstas. Algumas destas técnicas são: regressão linear, logística, árvores de decisão, redes neurais, entre outras.

Seguindo essa abordagem, para uma máquina representar uma mensagem, ela necessita de um conjunto, composto de pares formados por característica/valor. Após esta identificação, deve-se aplicar um algoritmo, classificado como *h*, o qual irá buscar as características de um vetor de caracteres *X*. A partir deste ponto, a unificação entre as abordagens de modelagem da linguagem e as técnicas de aprendizagem de máquina, são compatíveis. (RUSSELL; NORVIG, 2013).

Seguindo o raciocínio do PLN segundo Russell e Norvig (2013), o uso de algoritmo de compressão, sem perdas de caracteres, considera uma sequência de símbolos e detecta padrões repetidos. Como resultado, o algoritmo escreve uma sequência mais compacta de um texto, o autor deu como exemplo, o seguinte texto "0,142857142857142857", podendo ser comprimido para "0, [142857] *3".

Ao se apropriar das técnicas, regras e definições acima, pode ser abordado os diferentes SW de IA, e afirmar que, necessariamente, devem utilizar das ferramentas linguísticas descritas, alcançando como resultado o máximo pra ciência e o mínimo percentual de erros em suas análises, ao somar, a esses resultados, outras áreas da IA. Algumas destas são: a representação do conhecimento e a ontologia, ambas abordadas neste estudo.

PROCEDIMENTOS METODOLÓGICOS

Grande parte do problema de um estudo, principalmente sobre o estado de arte, se encontra na necessidade de abordar uma grande quantidade de artigos, teses, entre outros

Os SW são cada vez mais complexos e as engrenagens invisíveis que os tornam tão interessantes são com certeza absolutos objetivos a serem buscados por todo acadêmico de tecnologia da informação. Buscar a compreensão de funcionamento do uso de IA para análise da linguagem, definiu a necessidade desta pesquisa exploratória do estado da arte.

A estrutura desta pesquisa, desde sua concepção para avaliar o estado atual da arte da produção de conteúdo voltados para PLN em aplicações da IA, deu-se, como foco principal de busca, no sítio do *Google Acadêmico*, nas áreas de ciências da informação e demais associações. Foram colocados, sobre a luz deste estudo, 55 documentos, resultantes de pesquisas, e classificados em artigos de periódicos, monografias, trabalhos de conclusão de curso, dissertações de mestrado, teses de doutorado, entre outras publicações, nacionais e internacionais.

Como abordado anteriormente na metodologia, para a compreensão adequada do que está sendo pesquisado, no decorrer do momento em que este trabalho foi escrito, foi levado em consideração as novas tecnologias abordadas. A efetivação das buscas realizadas, originou-se os melhores resultados, atualização constante, e utilizou as informações necessárias, com foco na clareza dos documentos abordados.

Os descritores utilizados foram: "Estudo da Arte" e "Mapeamento Sistemático", foram acompanhados do uso das palavras-chave, "Inteligência Artificial", "Processamento de Linguagem Natural" e "Ontologia", e para o uso correto dos buscadores, foi utilizada as seguintes expressões booleanas na *String* de busca, *AND* e/ou *OR*, alternando entre os termos chave já definidos, nas buscas automáticas.

A necessidade de uma boa fundamentação teórica, e uma necessidade em trabalhos de conclusão de curso, especificamente, em pesquisa envolvendo o estado da arte, se torna obrigatório. Abordando questões similares de diferentes artigos, e concluindo em um mapeamento sistemático.

Além disso, utilizou-se a base de dados de documentos em bibliotecas digitais brasileiras e estrangeiras, como: *Science Electronic Library* (LILACS) e *Google Scholar* (*Google Acadêmico*), voltado para pesquisa em periódicos e revistas científicas.

RESULTADOS E DISCUSSÃO

Sobre o estado da arte, foi colocado como análise principal o trabalho de Almeida e Dias (2019), que fez uma abordagem brilhante da análise de domínio, no campo da ciência de informação brasileira. Ele apresenta a questão acerca do que é domínio, que segundo suas pesquisas, são as construções de sistemas de organização da informação e que compartilham conhecimentos, aproveita-se então para enfatizar a necessidade de entender como um determinado domínio, chamado de PLN, auxiliou diversas áreas do conhecimento a alcançarem um resultado, sobre estas pesquisas, em especial, aquelas que utilizaram alguma ferramenta da IA.

Ainda citando o mesmo autor, o estado da arte deve apontar inovações e soluções, para uma determinada área do conhecimento, bem como identificar as restrições e limitações daquelas produções teóricas e/ou práticas.

Desta forma, uma compreensão do que foi escrito sobre determinado assunto ganha uma importância única ao registrar os trabalhos anteriores como premissa necessária para realização de novas pesquisas. Com o estado da arte, será vital evidenciar o diálogo do pensamento do pesquisador com os autores citados para uma nova concepção científica (ALMEIDA; DIAS, 2019).

Após as pesquisas efetuadas, o grande volume de documentos encontrados obrigou a criação de critérios para excluir alguns trabalhos, embora se apresentem como resultados das pesquisas feitas, não agregavam valor a esta pesquisa. O critério de exclusão I, foi utilizado quando o trabalho não estava corretamente alocado, nos sítios de busca.

Seguindo a criação do critério de exclusão II, no qual aborda trabalhos que tenham surgido nas pesquisas, passando pela leitura do resumo e, após a leitura do mesmo na íntegra, se encontrava fora do espectro dos critérios desenvolvido para inclusão. O último critério de exclusão, o III, resulta após a leitura da pesquisa, embora o tema esteja dentro dos critérios de inclusão, os experimentos realizados em sua metodologia, foram mal elaborados ou inconclusivos. O resultado de todas essas construções se encontra no quadro 1.

Quadro 1. Mapeamento Sistemático de Literatura

Critérios	Documentos Analisados
I	38
II	1
III	14
IV	23
V	20
Excluídos	4
Total	55

Fonte: Autoria Própria (2023).

Analisando o quadro 1, é possível compreender uma totalidade de 55 documentos abordados, nas diferentes bases de dados, dos quais 38 foram escritos entre os anos de 2018 e 2021, proporcionando, assim, uma atualidade do tema abordado. Segundo Almeida e Dias (2019), a análise contínua deste panorama, fortalece a necessidade de verificar continuamente as produções científicas na área de Ciência da Informação.

Dentre os artigos produzidos, apenas um deles teve, como principal foco de estudo, a questão da abordagem da ontologia dentro da área estudada, no qual será abordado nas seções à frente, permitindo a compreensão do conhecimento dentro da tecnologia da informação.

Alguns dos documentos, analisados no quadro 1, podem assumir como verdadeiros mais de um critério de aceitação. Um exemplo que se pode citar, foi o caso das PLN, no qual 14 artigos, teses e outras produções científicas, tem o mesmo como parte de sua estrutura principal ou como ponto chave para trabalhar o real objetivo da pesquisa.

Os trabalhos abordados podem estar presentes nos critérios IV e V e como fortalecimento dessa afirmação, tem-se a fala de Lopes e Pereira (2020): “Sendo a inclusão de novos componentes e quantidade

de possibilidades de combinações de características, pode chegar facilmente à casa de milhões de resultados, o que não gera uma exclusão, mas sim fortalecimento do tema”, como será visto mais à frente.

No que se refere ao mapeamento sistemático, o estabelecimento de critérios apropriando-se, também, de outras formas de fundamentações metodológicas, tornaram-se mais presentes, visto que as análises dos artigos passaram por uma filtragem qualitativa, permitindo, assim, uma conclusão sobre o foco da pesquisa e posteriores resultados.

Em relação ao uso da IA na Educação, identificou-se que, entre os anos de 2019 e 2021, o mundo se viu sobre uma nova ameaça e, devido a esta situação, o ensino presencial teve que ser suspenso. O surgimento de novas formas de aprender tornaram-se necessárias, como pesquisas e métodos que permitissem que a educação ocorresse em casa.

Diversas técnicas foram implementadas, muitas vezes sem sucesso. Tentativas anteriores de abordar o tema ensino a distância esteve sobre a comunidade acadêmica, que propôs abrir novos caminhos utilizando tecnologia na educação.

Pode-se, assim, citar Amaral, Meira e Tavares (2020), quando ele afirma que soluções deste tipo começaram por volta dos anos 1980, inicialmente no ensino da aritmética, matemática e afins. Levando em consideração as pesquisas em IA visando desenvolver o ensino/aprendizagem nas últimas 3 décadas, os resultados não têm sido conclusivos.

Tem-se observado, recentemente, maiores avanços nas limitações encontradas pelas máquinas, ao lidar com a grande variedade de formas de aprendizado, trazendo, assim, a necessidade do critério I (quadro 1).

Numa tentativa de definir estas questões, em sua produção Gonzalez e Lima (2003), afirma que as estruturas sintáticas de uma sentença, podem ser obtidas através do processamento morfossintático, sendo as representações desta estrutura regidas por leis gramaticais.

Em se tratando do processo de aprendizado de máquina, um ponto em destaque que cruzou diversos textos encontrados nos artigos pesquisados, refere-se a questão do processo de aprendizagem de máquina.

Desta forma, o exemplo citado por Canto (2020), dá a seguinte relação, o desempenho em “T” é medido por “P” e melhora com a experiência “E”. Neste exemplo, seria desejado um programa de computador que aprendesse a classificar palavras. Para encontrar os vetores de *string* positivo e/ou negativo, como exemplo pode-se fazer a seguinte associação: “E” é igual a experiência de ver o usuário classificar palavras, como positivo e/ou negativo.

Transformando em algoritmo o momento analisado pela máquina, “T” é igual a tarefa de classificar as palavras. Segundo ponto, “P” é igual ao número ou fração de palavras, corretamente classificadas, como positivo/negativo, concluindo com “E” que incrementa e reinicia o loop.

Segundo esse exemplo, Amaral, Meira e Tavares (2020), fortalece a ideia de que um dos usos mais comuns para as técnicas de aprendizado de máquina é utilizar os algoritmos, fundamentados em *Machine Learning*. Garantindo, assim, sua aplicação na atuação de situações já conhecidas, para prever e classificar novas situações dentro de um mesmo contexto.

Esta estrutura supracitada, permite uma visualização clara do processo de aprendizagem supervisionado, já para a segunda etapa, alguns pontos devem ser garantidos para que possa correr tudo o que estava previsto e, para tal, as PLN são fundamentais. Segundo Evangelista, Sassi, Silva (2020), o PLN, também conhecido como linguística computacional, é um subcampo da IA, no qual visa aprender, compreender, reconhecer e produzir conteúdo em linguagem humana.

A aprendizagem de máquina não supervisionada, pelo que se pode observar, até o atual estado de arte, necessita de um léxico em constante atualização, precisa de um PLN bem estruturado, e talvez seja a parte mais importante.

Sobre Dados e *Games*, como consequência dos diálogos anteriores que, os dados gerados na web, nos dias de hoje, são uteis com valor de mercado e podem gerar informações extremamente relevantes. As informações supramencionadas, revelam um ponto importante, no qual não importando o processo abordado na web, dados são gerados. Os questionamentos que podem-se obter são: Como tratar esses dados? Como capítulos?

Quando dados são abordados, surgem as questões permanentemente citadas, ao abordar tópicos como banco de dados e as estruturas que modelam esses bancos. Para fortalecer a compreensão deste assunto, pode-se citar a afirmação de Moura *et al.* (2019), uma das principais estruturas componentes de um banco de dados, e a modelagem entidade-relacionamento.

Após a escolha do modelo relacional a ser usado no banco de dados, o próximo passo, de acordo com Moura *et al.* (2019), para a construção do banco de dados, é os usuários definirem consultas utilizando o SQL. Para a construção do SW, o uso da sintaxe correta é obrigatório. O algoritmo pode possuir comandos, para inserir, alterar, ou recuperar os dados, isso torna-se um grande desafio para os usuários dos bancos de

dados que não dominam a linguagem de programação adequada ou compreendem o mínimo, como por exemplo, a estrutura de um banco de dados.

Alguns problemas podem ser abordados no decorrer dos estudos, envolvendo dados, segundo Evsuko e Gomes (2019), a necessidade de definir informações não estruturadas, é urgente. Se as informações se apresentam como resultado, mas não atendem a um modelo relacional, tendo como origem dados específicos, já pré-determinados, a conclusão é de que essas informações são inválidas.

Durante a pesquisa, obteve-se um foco especial sobre documentos produzidos na web, com estruturação de dados variados, resultando em dados compostos, em sua maioria, de documentos em construção no formato de texto. Essas informações resultantes, possivelmente ainda não estruturadas, tratam-se de: relatórios técnicos, artigos científicos, blogs de operação, análises laboratoriais, pareceres técnicos, periódicos, entre outros. Dessa forma, essa vasta disponibilidade de dados não resulta em informações utilizáveis sem a análise adequada.

Os pesquisadores demonstraram conhecimentos bem abrangentes, abordaram os fundamentos da ciência da informação, fazendo uso de SW já disponíveis e, a partir deles, conseguiram teorizar sobre seus questionamentos.

Ao se utilizar destas estruturas, criaram projetos práticos que construíram resultados altamente confiáveis, dentre eles, a pesquisa de Coneglian (2020), o qual abordou todos os temas, detalhadamente, e com aplicação prática.

Os trabalhos excluídos, vistos anteriormente, embora seus focos sejam de excelência acadêmica, foram necessários exatamente para fortalecer proposições apresentadas e objetivos estabelecidos. A limitação dos questionamentos feitos e os contrapontos apresentados, resultaram em um trabalho mais abrangente e claro.

Embora em algumas pesquisas possam existir variáveis não verificadas, não houve uso de SW equivalentes. Para se obter resultados similares, a determinação do tipo de máquina utilizada e/ou a definição do ambiente de controle, podem ter sido minimizados pelo autor.

Ademais, pode ser abordado, a utilização de redes neurais para alcançar resultados semelhantes aos dos artigos mencionados durante a MS. A atualização de como se encontra o estado da arte, pode se dar a cada momento. Pode-se observar que já se produziram jogos, inclusive, um dos artigos tem em seu foco principal esta questão. Podem, também, ser abordados pela ótica da IA, filmes, simuladores, *chatbox*, robôs e outras infinitudes de SW, para a comodidade e lucro.

Justifica-se, também, um novo estudo como produção sobre os processos abordados, sendo possível recorrer a um estudo de caso focado. As áreas que podem ser abordadas são: gestão de projetos frágeis, utilizando de IA e *frameworks* atuais; e, análise de repercussões e impactos, com as diferentes equipes em evidência, resultando na afirmação ou negação das questões abertas anteriormente.

CONSIDERAÇÕES FINAIS

Em suma, é possível concluir que o PLN é uma vertente da IA, no qual auxilia computadores a entender, interpretar e manipular a linguagem humana, resultando, assim, de diversas disciplinas, incluindo a ciência da computação e a linguística computacional, que buscam preencher a lacuna entre a comunicação humana e o entendimento dos computadores.

Dessa forma, é importante se ter uma fundamentação de boas regras lexicais, dentro do PLN aplicado a IA, com a programação adequada da área para o bom uso da mesma. Utilizando-se técnicas adequadas dentro de uma boa estrutura composta de aprendizagem supervisionada e somando-se a um banco de dados adequado, todo e qualquer dado pode resultar em uma informação com alto valor.

REFERÊNCIAS

- AFONSO, A. R.; DUQUE, C. G. (2019). Análise de sentimentos em comentários de vídeos do youtube utilizando aprendizagem de máquinas supervisionadas. *Ciência da Informação*, v. 48, n. 3.
- ALMEIDA, J. F. V. R.; DIAS, G. A. (2019). Estado da arte sobre análise de domínio no campo da ciência da informação brasileira. *Brazilian Journal of Information Science*, Universidade Estadual Paulista – UNESP, v. 13, n. 3, p. 26-45.
- ALVARENGA, J. P. R. (2019). *Avaliação de métodos de transferência de aprendizado aplicados a problemas de processamento de linguagem natural em textos da língua portuguesa*.
- ARAUJO, G. R. D. O.; VITTORAZZI, W. D. O. (2018). *A aplicação de redes neurais artificiais recorrentes no processamento de linguagem natural*.

BARBOSA, A. F.; CAMPELO, C. (2020). Processamento de linguagem natural em artefatos textuais educacionais: Um mapeamento sistemático no contexto brasileiro. In: *SBC. Anais do XXXI Simpósio Brasileiro de Informática na Educação*. [S.l.], p. 1433-1442.

BULEGON, H.; MORO, C. M. C. (2010). Mineração de texto e o processamento de linguagem natural em sumários de alta hospitalar. *Journal of Health Informatics*, v. 2, n. 2.

CANTO, L. G. (2020). Análise de notícias do mercado financeiro utilizando processamento de linguagem natural e aprendizado de máquina para decisões de swing trade. *Universidade Federal do Rio de Janeiro – UFRJ*. (Doctoral dissertation).

CONEGLIAN, C. S. (2020). Recuperação da informação com abordagem semântica utilizando linguagem natural: a inteligência artificial na ciência da informação. *Universidade Estadual Paulista – UNESP*.

COPPIN, B. (2015). Inteligência artificial. [S.l.]: Grupo Gen-LTC.

EVANGELISTA, J. R. G.; SILVA, E. M. L. da; SASSI, R. J. (2020). Enriquecimento de base de dorks com processamento de linguagem natural. *Brazilian Journal of Development*, v. 6, n. 3, p. 10763-10780.

GOMES, D. d. S. M.; EVSUKOFF, A. G. (2019). *Processamento de linguagem natural em português e aprendizagem profunda para o domínio de óleo e gás*.

GONZALEZ, M.; LIMA, V. L. S. (2003). Recuperação de informação e processamento da linguagem natural. In: *XXIII Congresso da Sociedade Brasileira de Computação*. [S.l.: s.n.], v. 3, p. 347-395.

GRACIOSO, L. de S.; SALDANHA, G. S. (2016). *Ciência da informação e filosofia da linguagem: da pragmática informacional a web pragmática*. [S.l.]: Junqueira&Marin Editores.

KULTZAK, A. F. (2016). Categorização de textos utilizando algoritmos de aprendizagem de máquina com WEKA. *Universidade Tecnológica Federal do Paraná – UTFPR*. (Masters dissertation).

KUYVEN, N. L.; VANZIN, V. J. de B.; ANTUNES, C. A.; CEMIN, A.; SILVA, J. L. T.; TAROUÇO, L. M. R. (2020). Assistente conversacional para resolução de problemas trigonométricos em linguagem natural. *Revista Brasileira de Informática na Educação*, v. 28, p. 208-228.

LOPES, A. S. B.; PEREIRA, M. M. (2020). Exploração de espaço de projeto para multicóres heterogêneos com o uso aprendizado de máquina: o estado da arte. *Brazilian Journal of Development*, v. 6, n. 5, p. 26730-26749.

MINUSSI, S.; MOURA, A. A.; JARDIM, M. L. G.; RAVASIO, M. H. (2018). Considerações sobre estado da arte, levantamento bibliográfico e pesquisa bibliográfica: relações e limites. *Revista Gestão Universitária*, v. 9.

MORAIS, E. A. M.; AMBROSIO, A. P. L. (2007). *Ontologias: conceitos, usos, tipos, metodologias, ferramentas e linguagens*. (Technical report).

MOURA, E. A. *et al.* (2019). *Consultas sql utilizando linguagem natural para um contexto acadêmico*.

NETO, J. P. de B. Gestão de pessoas 4.0: liderança e aprendizado diante da inteligência artificial e da quarta revolução industrial. *Gestão de Pessoas*, p. 13.

OLIVEIRA, D. C. D. *et al.* (2019). *Aplicação das técnicas de processamento de linguagem natural cosine similarity e word mover's distance na automatização da correção de questões discursivas no sistema tutor inteligente mazk*. Araranguá, SC.

_____ (2020). Aplicação das técnicas de processamento de linguagem natural cosine similarity e word movers distance para auxiliar na correção de questões discursivas em um tutor inteligente. In: *SBC. Anais do XXXI Simpósio Brasileiro de Informática na Educação*. [S.l.], p. 1243-1252.

PEREIRA, F. R.; RIGO, S. J. (2013). Utilização de processamento de linguagem natural e ontologias na análise qualitativa de frases curtas. *RENOTE*, v. 11, n. 3.

RUSSELL, S.; NORVIG, P. (2013). *Artificial intelligence: Pearson new international edition: A modern approach*. [S.l.]: Pearson Higher Ed.

SCHRAMM, F. R. (2020). A sociedade globalizada contemporânea de risco e o aprimoramento humano no transhumanismo e posthumanismo. *Revista Iberoamericana de Bioética*, n. 13, p. 01-16.

SIMÕES, A.; FARINHA, R. (2010). Dicionário aberto: um recurso para processamento de linguagem natural. Vice-versa. *Revista galega de traducion*, p. 159-172.

STARKS, M. R. (2020). *Bem-vindos ao inferno na terra-inteligência artificial, bebes, bitcoin, carteis, china, democracia, diversidade, disgenia, igualdade, hackers, direitos humanos, islamismo, liberalismo, prosperidade*.

TAVARES, L. A.; MEIRA, M. C.; AMARAL, S. F. (2020). Inteligência artificial na educação: Survey. *Brazilian Journal of Development*, v. 6, n. 7, p. 48699-48714.

VIEIRA, R.; LOPES, L. (2010). *Processamento de linguagem natural e o tratamento computacional de linguagens científicas*. EM CORPORA, p. 183.

WALTRICK, S. A. *et al.* (2012). *Cr terios para a sele o de fontes de informa o cient fica multim dia em acesso livre na internet: cria o de acervo digital para cursos de gradua o a dist ncia.*

WITTGENSTEIN, L. (2013). *Tractatus logico-philosophicus.* s o Paulo: Companhia editora nacional.

XAVIER, J. (2019). *O uso da an lise de sentimentos no twitter para avaliar a opini o do p blico consumidor a respeito do sistema operacional.*

Received on 02, 2023.

Accepted on 06, 2023.