Utilização de Voyant Tools para Compreensão de Documentos Não Estruturados

GABRIELA ALVES DA ROCHA
ELOÍZA MARTINS PRIMO CAPELOCI

RESUMO

A análise e a mineração de dados textuais têm desempenhado um papel fundamental na interpretação de informações geradas em um cenário cada vez mais informatizado, no qual dados são produzidos constantemente. Estima-se que, diariamente, aproximadamente 2,0 quintilhões de *bytes* de dados assumam a forma de texto. A mineração de dados não estruturados consiste no processo de extração de informações úteis e conhecimento a partir de uma grande quantidade de textos em linguagem natural, como documentos em formato PDF, *e-mails* e postagens em redes sociais. Desde o surgimento da computação, tanto linguistas quanto especialistas em recuperação da informação têm desenvolvido *softwares* com o objetivo de identificar padrões linguísticos que não podem ser detectados em uma leitura convencional. Este estudo aborda a análise de conteúdo textual por meio da ferramenta web *Voyant Tools*, que explora técnicas de análise textual e mineração de dados textuais. O objetivo central é investigar como essa ferramenta pode contribuir para a compreensão de documentos não estruturados, que não seguem um padrão definido devido à sua variedade linguística, ao contexto e à intenção comunicativa, ao estilo de escrita e à inclusão de elementos não textuais, como gráficos, tabelas, imagens e vídeos. Para a análise, também foram incorporadas algumas técnicas de Processamento de Linguagem Natural (PLN), essenciais para uma interpretação mais precisa e significativa dos textos. Dentre essas técnicas, destacam-se a remoção de *stopwords*, a contagem de palavras e a análise da frequência de cada termo dentro de um *corpus* textual.

Palavras-chaves: Processamento de Linguagem Natural; Análise Textual; Mineração de dados; Dados não Estruturados.

Abstract

Textual data analysis and mining have played a fundamental role in the interpretation of information generated in an increasingly computerized environment, in which data are constantly being produced. It is estimated that approximately 2.0 quintillion bytes of data are generated in the form of text every day. Unstructured data mining consists of the process of extracting useful information and knowledge from a large amount of natural language texts, such as PDF documents, emails, and social media posts. Since the emergence of computing, both linguists and information retrieval specialists have developed software with the aim of identifying linguistic patterns that cannot be detected in conventional reading. This study addresses the analysis of textual content through the Voyant Tools web tool, which explores textual analysis and textual data mining techniques. The main objective is to investigate how this tool can contribute to the understanding of unstructured documents, which do not follow a defined pattern due to their linguistic variety, context and communicative intention, writing style and the inclusion of non-textual elements, such as graphs, tables, images and videos. For the analysis, some Natural Language Processing (NLP) techniques were also incorporated, which are essential for a more precise and meaningful interpretation of the texts. Among these techniques, the following stand out: the removal of stopwords, word counting and the analysis of the frequency of each term within a textual corpus.

Keywords: Natural Language Processing; Textual Analysis; Data Mining; Unstructured Data.

1. INTRODUÇÃO

A análise textual e a mineração de textos desempenham um papel fundamental na compreensão de dados textuais em um mundo cada vez mais voltado à informação. De acordo com um estudo da Gartner (2022), estima-se que, em 2023, 80% dos dados gerados no mundo sejam estruturados, incluindo dados textuais. Isso significa que, diariamente, aproximadamente 2,0 quintilhões de *bytes* de informações são produzidos sob a forma de texto.

Atualmente, nota-se, cada vez mais, a importância dos dados para as empresas e para a

sociedade em geral. Esses dados são gerados em volumes crescentes e, segundo Grego (2014), se todo o conteúdo digital produzido pela humanidade, músicas, livros, filmes, documentos e dados fosse armazenado em *iPads* e empilhado, a altura da pilha corresponderia a cerca de dois terços da distância entre a Terra e a Lua. O autor complementa que a tendência é que essa pilha imaginária alcance 6,6 vezes essa distância nos próximos anos.

Diversos conceitos foram desenvolvidos para lidar com o crescimento exponencial dos dados. O termo Big Data, por exemplo, refere-se à grande quantidade de dados gerados diariamente, dos quais 80% estão na forma de texto. A mineração de textos possibilita a extração de informações por meio da análise quantitativa e qualitativa de grandes coleções documentais, como artigos científicos, notícias jornalísticas, páginas da *web* e redes sociais. Esse método permite a identificação de padrões e o estabelecimento de relações entre eles, considerando a frequência e a temática dos termos (Pezzini, 2016).

2. REVISÃO DA LITERATURA

2.1 Ferramenta Voyant Tools

O *Voyant Tools* é uma plataforma *on-line* e gratuita de analise textual que oferece uma variedade de ferramentas para explorar e visualizar dados textuais de maneira eficiente. Desenvolvido para atender às necessidades de pesquisadores, estudantes e profissionais que lidam com análise textual, o *Voyant Tools* procura simplificar a interpretação de grandes conjuntos de dados textuais.

Uma das características notáveis do *Voyant Tools* é a sua interface intuitiva e amigável, tornando a análise textual acessível mesmo para usuários sem experiência técnica avançada. A plataforma permite a análise de textos brutos, oferecendo informações sobre, por exemplo, padrões, frequência de palavras e estruturas linguísticas. Entre as ferramentas disponíveis, destacam-se a análise de frequência de palavras, que revela quais termos são mais relevantes em um texto, e a nuvem de palavras, que apresenta visualmente as palavras mais utilizadas.

Além disso, o *Voyant Tools* permite a criação de gráficos interativos, facilitando a compreensão de tendências e padrões ao longo do texto. A capacidade de realizar análises mais avançadas, como a exploração de relações entre termos, faz do *Voyant Tools* uma ferramenta versátil para estudos qualitativos e quantitativos. A possibilidade de integrar diferentes tipos de documentos, como artigos, relatórios e, até mesmo, redes sociais, amplia ainda mais o escopo de aplicação dessa ferramenta. O *Voyant Tools* se apresenta como uma contribuição para a pesquisa em humanidades digitais, oferecendo uma abordagem pertinente e útil para a análise de grandes volumes de texto. Sua flexibilidade, combinada com uma variedade de recursos, o posiciona como uma ferramenta apreciável para aqueles que buscam explorar e compreender o significado subjacente a dados textuais de maneira abrangente e interativa.

2.2 Mineração de textos

I. Aprimoramento e Mineração de Texto

A evolução das técnicas de mineração de texto investigadas em melhorias significativas. Esses progressos possibilitaram a automatização de rotinas e processos que, anteriormente, exigiam esforço humano. Isso, por sua vez, foi elaborado em maior eficiência, contribuindo para a redução do tempo necessário para a execução dessas tarefas.

De maneira geral, a mineração de textos pode ser entendida como uma subárea da

Recuperação da Informação (RI) (Salton; Mcgill, 1983). Através de um conjunto de rotinas de processamento e análise de padrões, a informação é recuperada a partir de dados textuais, gerando, consequentemente, conhecimento. Assim, destaca-se que a fundamentação dessa área está ligada às definições de dados, informação e conhecimento.

II. Definições de Dados, Informação e Conhecimento

Buscando um melhor alinhamento sobre os conceitos relacionados à mineração de textos, é interessante pontuar as suas definições. De acordo com Silva, Peres e Boscarioli (2016), dado pode ser descrito como algo bruto, sem contexto, ou seja, um símbolo ou um conjunto de símbolos quantificados ou quantificáveis. Por outro lado, a informação pode ser descrita como dados tratados, os quais possuem significados. Deve-se observar que nem toda informação gerada é necessariamente útil e utilizada, e que nem tudo dado processado é garantia de informação. Por fim, o conhecimento pode ser definido como uma informação explorada com algum propósito específico, ou seja, utilizado para, por exemplo, tomada de decisão, construção de cenários, entre outros.

Nesse sentido, pode-se entender dados como matéria-prima indispensável para a análise. Ainda de acordo com Silva, Peres e Boscarioli (2016), os dados podem ser classificados de duas formas: estruturados e não estruturados. A identificação do tipo de dado é essencial para que o processo de mineração possa ser aplicado, uma vez que os examinados de cada tipo de dado exigem rotinas específicas para seu processamento. De modo geral, os dados estruturados são aqueles que se referem ao resultado de transações, ou ainda, de medição ou observação, podendo ser armazenados em uma tabela ou em um formato que siga um padrão pré-definido, facilmente compreensível pela máquina. Enquanto os dados não estruturados referem-se àqueles que não apresentam padrões pré-definidos, sendo necessária a aplicação de rotinas para o seu tratamento e processamento.

III. Análise de dados não estruturados e técnicas de mineração

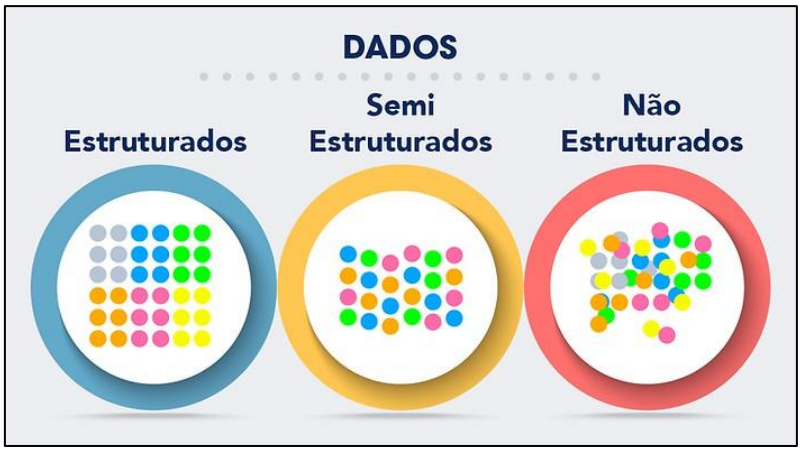
Uma análise de dados armazenados em formato não estruturado é considerada uma atividade mais complexa, se comparada à análise de dados estruturados. Isso se dá pelo fato de os dados possuírem características de não estruturação. Logo, são permitidas técnicas específicas para tratamento deste tipo de dados. Este conjunto de técnicas e ferramentas faz parte da área de Recuperação de Informações, mais conhecido como Descoberta de Conhecimento em Textos (*Knowledge Discovery from Text – KDT*). De acordo com Beppler *et al.* (2005), KDT engloba técnicas e ferramentas inteligentes e automáticas que auxiliam na análise de grandes volumes de dados com o intuito de "garimpar" conhecimento útil, beneficiando não somente usuários de documentos eletrônicos da Internet, mas qualquer domínio que utilize textos não estruturados.

Existem muitas técnicas para mineração de dados, sejam eles dados estruturados, semiestruturados ou não estruturados, que é o caso dos arquivos que foram usados para o desenvolvimento deste trabalho. Um PDF (*Portable Document Format*) é considerado um dado não estruturado porque, ao contrário de dados estruturados, como banco de dados relacionais ou planilhas, as informações contidas em um arquivo PDF não são organizadas em tabelas, linhas e colunas. A imagem abaixo representa como os dados são organizados.

Para compreender as diversas formas pelas quais a informação pode ser organizada e processada, é fundamental distinguir entre os diferentes tipos de dados. A maneira como eles são estruturados impacta diretamente seu armazenamento, análise e utilização. Existem fundamentalmente três categorias principais de dados: dados estruturados, semiestruturados e não estruturados. Cada uma dessas categorias possui características específicas que determinam

sua aplicabilidade e os métodos mais adequados para o seu gerenciamento. A Imagem 1 ilustra a diferença entre essas classificações, sendo essencial para a compreensão dos conceitos que serão incluídos ao longo deste trabalho.

Imagem 1 – Tipos de Dados



Fonte: DATASIDE (2021)

Na Imagem 1, os dados estruturados são caracterizados por um alto grau de organização, geralmente dispostos em tabelas com linhas e colunas bem definidas, como ocorre em bancos de dados relacionais, por exemplo. Já os dados semiestruturados possuem alguma estrutura, mas não seguem um modelo rígido. Exemplos incluem arquivos XML ou JSON, onde a organização é hierárquica e *tags* ou marcadores ajudam a definir os elementos de dados. Por fim, os dados não estruturados representam a maior parte dos dados que são gerados atualmente. Eles não possuem um formato predefinido, sendo armazenados em seu formato nativo, texto. Textos livres, áudios, vídeos e imagem são alguns exemplos.

2.3 Processamento de linguagem natural

Processamento de língua natural é uma subárea da Ciência da Computação, Inteligência Artificial e da Linguística que estuda os problemas da geração e compreensão automática de línguas humanas naturais.

O Processamento de Linguagem Natural (PLN) é complementado pela análise de textos, que conta, agrupa e categoriza palavras para extrair estruturas e significados de grandes volumes de conteúdo. As tarefas fundamentais do PLN incluem a tokenização e análise sintática, lematização ou sistematização, marcação dos componentes do discurso, detecção de idioma e identificação de relações semânticas. As tarefas de PLN desmembram a linguagem em partes menores e essenciais, procurando compreender as relações entre elas e explorar como esses elementos colaboram para criar significado (SAS, 2023).

Essas tarefas fundamentais são frequentemente utilizadas em níveis mais avançados de

PLN, tais como a categorização de conteúdo para resumos baseados em Linguística, a descoberta e modelagem de tópicos para identificar significados temas em coleções de texto, e a extração contextual para obter informações estruturadas de fontes textuais. Em todos esses casos, o objetivo é utilizar técnicas linguísticas e algoritmos para aprimorar ou enriquecer o texto, resultando em uma interpretação mais precisa e valiosa das entradas brutas (SAS, 2023). Por exemplo, ele permite que os computadores leiam textos, escutem e interpretem discursos, identifiquem emoções e determinem quais partes são relevantes.

Alguns dos objetivos comuns no Processamento de Linguagem Natural incluem a recuperação de informações a partir de textos, a tradução automática, a interpretação de textos e a capacidade de realizar inferências a partir de textos (Liddy, 2003).

O Processamento de Linguagem Natural incorpora técnicas diversas para interpretar a linguagem humana, desde métodos estatísticos e de *machine learning* a abordagens algorítmicas e baseadas em regras e *Deep Learning*. O PLN é importante porque ajuda a resolver a ambiguidade na linguagem e adiciona uma estrutura numérica útil aos dados para muitas aplicações *downstream*, como reconhecimento de fala ou análise de texto (SAS, 2023).

O objetivo mais desafiador do Processamento de Linguagem Natural é, atualmente, a criação da Web Semântica, que busca conectar o vasto volume de dados disponíveis na internet às necessidades de seus milhões de usuários. Dada a imensa quantidade de dados não estruturados gerados diariamente, que vão desde registros médicos até conteúdos em redes sociais, a automação torna-se essencial para uma análise completa e eficiente de texto e fala (SAS, 2023).

Em termos de Processamento de Linguagem Natural (PLN), o *Voyant Tools* aplica técnicas automatizadas e superficiais para analisar e entender a linguagem humana. Isso inclui a tokenização (divisão de textos em palavras ou frases), a análise de frequência de termos, remoção de *stopwords*, palavras que geralmente não contribuem para o significado do texto e podem ser removidas para focar nas palavras-chave mais relevantes. A técnica de análise textual tem uma grande relevância, pois envolve contar a frequência de cada ocorrência de palavras no texto e palavras mais frequentes podem indicar tópicos na interpretação dos dados textuais.

3. MATERIAL E MÉTODOS

Para a realização deste projeto, foram efetuadas buscas sistematizadas em bases de dados eletrónicas. Como SCIELO E Google Acadêmico, além de consultar sites, livros e jornais especializados no tema. As pesquisas não tiveram restrição de idioma ou data de publicação, abrangendo dados e estudos de julho a setembro de 2023.

Foram selecionados 20 trabalhos focados em inteligência artificial, algoritmos de aprendizado de máquina, riscos e impactos sociais da IA, big data e tecnologia da informação. As palavras-chave utilizadas para a busca bibliográfica foram: inteligência artificial; aprendizado de máquinas; algoritmos; sociedade; Big Data; Tecnologia da Informação.

Os PDFs e artigos pesquisados abordam a importância da inteligência artificial, o uso de algoritmos de aprendizado de máquina e os impactos dessa tecnologia na sociedade.

3.1 Analise dos dados

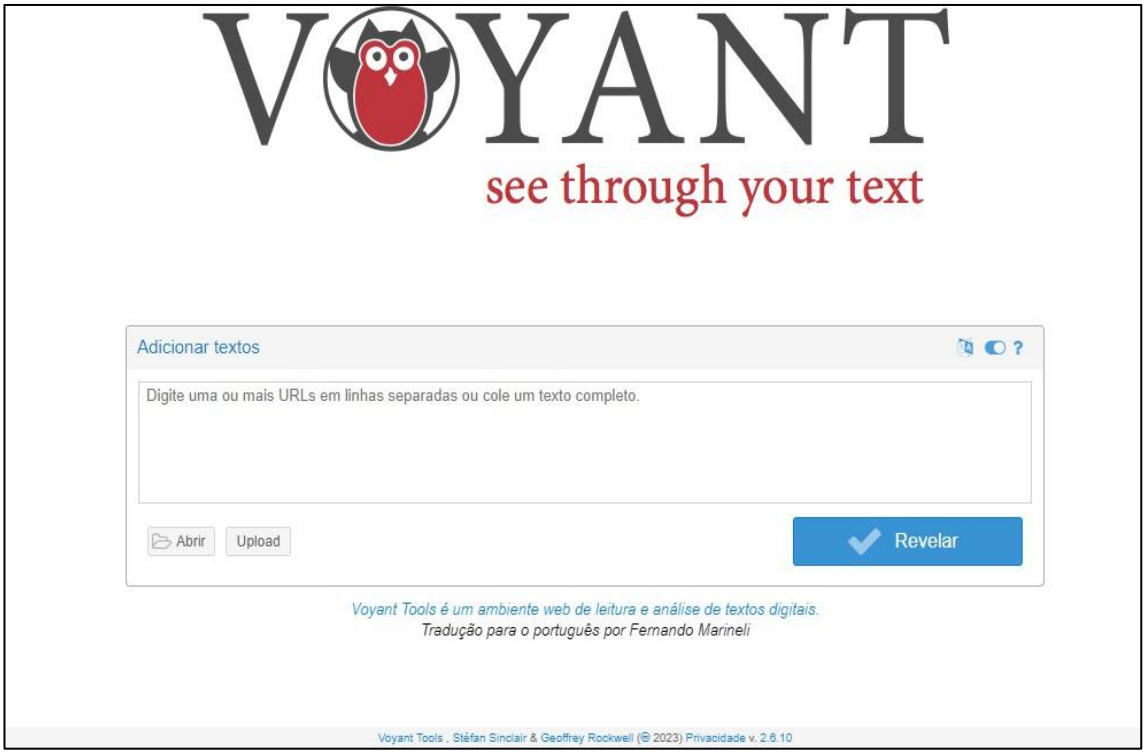
Os arquivos foram carregados na ferramenta *Voyant Tools*. Após o *upload*, fórmulas, tabelas imagens e *stopwords* foram removidas dos textos. Esse procedimento foi crucial para identificar as palavras-chave e os termos mais relevantes dentro do tema escolhido. A ferramenta oferece diversas opções e possui uma interface intuitiva, facilitando a análise.

4. RESULTADOS E DISCUSSÃO

Os resultados das análises feitas foram representados nas imagens abaixo para mostrar a relação estre os termos, a correlação das palavras no corpus textual e a frequência de cada termo gerado pelas análises.

A imagem abaixo mostra a página inicial da ferramenta, onde são carregados os arquivos e os textos para análise, a interface gráfica é bastante intuitiva e de fácil utilização.

Imagem 2 – Página Inicial da ferramenta *Voyant Tools*



Fonte: elaborada pelas autoras através da plataforma *Voyant Tools* (2023)

Como ilustrado na Imagem 2, o *Voyant Tools* oferece um campo central onde o usuário pode digitar ou colar textos diretamente, inserir *URLs* (uma por linha) ou carregar arquivos de texto do próprio computador. A simplicidade na entrada de dados é um ponto forte da ferramenta, pois facilita o início da análise para diferentes tipos de *corpus*. Uma vez inseridos os dados, o botão "Revelar" inicia o processamento, levando o usuário a uma página de resultados com diversas visualizações interativas e painéis de análise.

Após uma análise detalhada, foi gerada uma nuvem de palavras contendo os 245 termos mais recorrentes. A disposição das palavras segue um critério visual em que os termos mais frequentes são posicionados centralmente e exibidos em tamanhos maiores, facilitando a identificação de padrões de destaque. À medida que o algoritmo percorre a lista, ele distribui os termos de maneira a preencher os espaços disponíveis, garantindo que palavras menores sejam acomodadas entre as palavras de maior relevância.

É fundamental compreender que a cor das palavras e sua posição na visualização não possuem significado específico. De fato, ao redimensionar a janela ou recarregar a página, as palavras podem ser redistribuídas em diferentes locais. No entanto, a interação com a nuvem de palavras é intuitiva: ao posicionar o mouse sobre um termo, uma caixa informativa é exibida, revelando sua frequência dentro do conjunto de dados analisado. Esse recurso permite uma exploração mais detalhada dos termos presentes na análise.

Imagem 3 – Nuvem de palavras

Fonte: elaborada pelas autoras através da plataforma *Voyant Tools* (2023).

O gráfico de bolhas exibe a frequência e a distribuição de termos em um corpus. Cada documento no corpus é representado como uma linha horizontal e dividido em segmentos de igual comprimento (50 segmentos por padrão). Cada palavra selecionada é representada como uma bolha sendo que o tamanho da bolha indica a frequência da palavra no segmento de texto correspondente. Quanto maior a bolha mais frequentemente a palavra ocorre. Colocar o cursor sobre um local na linha do documento fará com que uma bolha apareça com frequências de termo para esse segmento.

No final da linha do documento, há um rótulo que indica a contagem de todos os termos selecionados para esse documento. Deixar o cursor sobre esse rótulo mostra as frequências distribuídas entre termos mais frequentes desse documento.

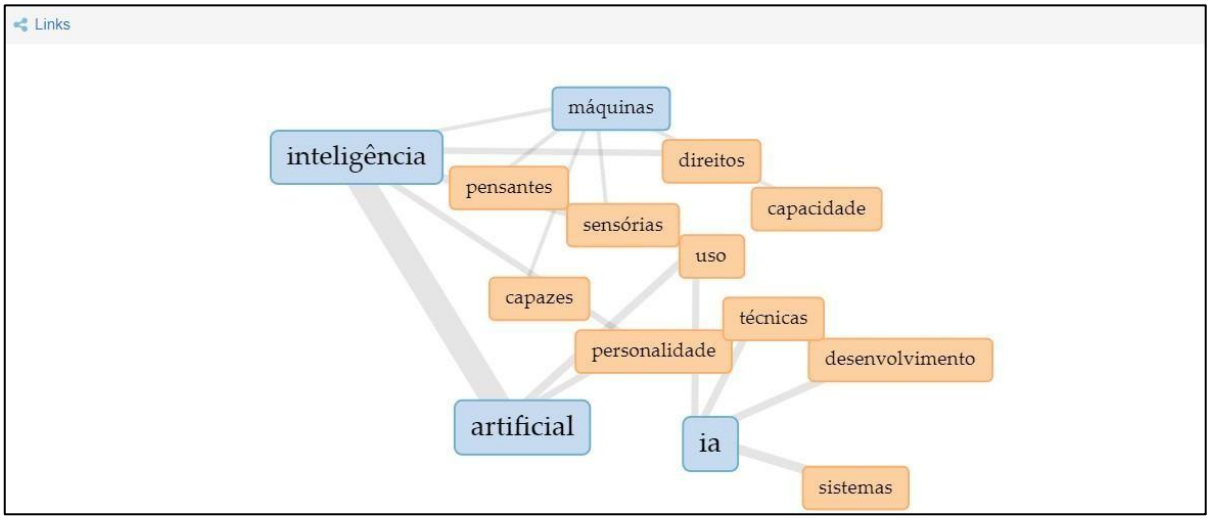
Imagem 4 – Gráfico de bolhas

Fonte: elaborada pelas autoras através da plataforma *Voyant Tools* (2023).

A Imagem 5 apresenta um mapa conceitual que constrói relações semânticas a partir do termo central **"inteligência"**, conectando-o a diversas palavras e expressões que compõem um corpus textual. Os termos estão organizados em caixas coloridas, destacando diferentes dimensões da inteligência e sua aplicação, com destaque para "inteligência artificial".
A disposição visual do mapa evidencia conexões fundamentais entre conceitos como máquinas pensantes, capacidade sensorial, sistemas inteligentes, desenvolvimento de técnicas e até discussões sobre personalidade e direitos no contexto da inteligência artificial. As linhas que ligam esses termos indicam relações semânticas, ajudando a compreender como eles se encaixam em um universo maior de significados.

Essa abordagem permite uma interpretação estruturada do tema, facilitando a identificação de padrões e interligações dentro do corpus textual. Além disso, a organização gráfica contribui para uma análise intuitiva dos elementos mais relevantes, tornando o conteúdo visualmente acessível e didático.

Imagem 5 – Links

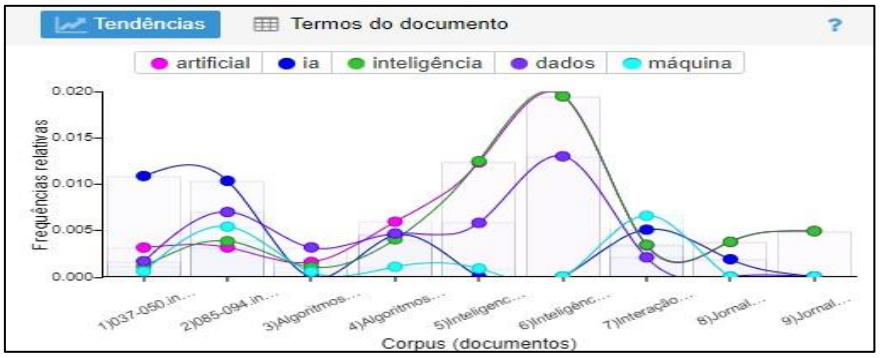


Fonte: elaborada pelas autoras através da plataforma *Voyant Tools* (2023).

Este gráfico de distribuição representa visualmente a frequência dos termos mais utilizados em todo o corpus textual analisado. A ferramenta mapeia essas palavras, permitindo que se observe quais são os conceitos mais recorrentes e como eles se distribuem ao longo dos documentos.

Quando o corpus completo é visualizado, é possível identificar padrões gerais de uso da linguagem, destacando termos-chave que aparecem com maior frequência. Isso oferece *insights* sobre a estrutura do texto, facilitando a compreensão dos temas predominantes na análise. Caso um único documento seja selecionado, a ferramenta refina essa abordagem, apresentando a distribuição específica dos termos dentro daquele contexto, permitindo comparações mais detalhadas entre diferentes fontes.

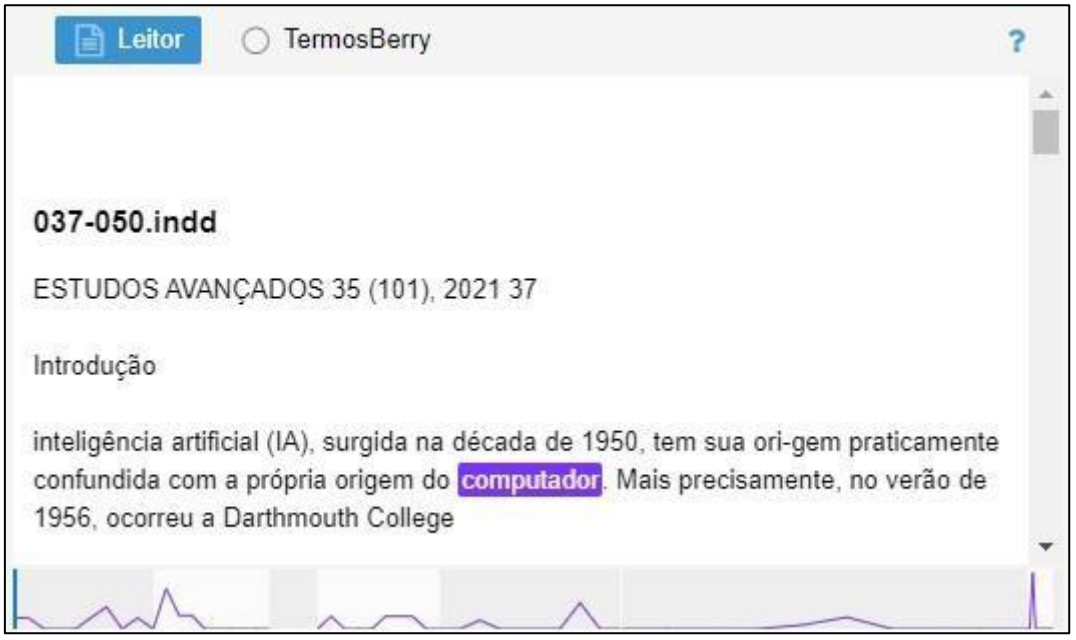
Imagem 6 – Gráfico de tendências



Fonte: elaborada pelas autoras através da plataforma *Voyant Tools* (2023).

A aba ‘Leitor” é reservada para a revisão dos textos completos e logo abaixo pode ser observado um pequeno gráfico de barras que indica a quantidade de texto que cada documento possui.

Imagem 7 – Conteúdo da aba ‘Leitor’



Fonte: elaborada pelas autoras através da plataforma *Voyant Tools* (2023).

Esses indicadores ajudam a compreender a estrutura do texto, identificando padrões de linguagem, frequência de termos e complexidade textual. A comparação entre documentos mais longos e mais curtos, bem como entre textos mais fáceis ou difíceis de ler.

A ferramenta “Sumário” fornece uma visão geral textual simples do corpus sob análise, incluindo número de palavras, número de palavras exclusivas, documentos mais longos e mais curtos, maior e menor densidade de vocabulário, número médio de palavras por frase, palavras mais frequentes, picos notáveis em frequência, e palavras distintas.

A seguir, a imagem apresenta um resumo das principais características e especificações textuais do corpus textual apresentado ao longo desse trabalho, incluindo a extensão dos documentos, a densidade vocabular, a média de palavras por frase e o índice de leiturabilidade (Índice de legibilidade), destacando os documentos com os valores mais altos e mais baixos para cada métrica.

Imagem 8 - Sumário (parte 1)

Fonte: elaborada pelas autoras através da plataforma *Voyant Tools* (2023).

A segunda parte do Sumário, fornece uma outra visão geral da segunda parte das características estatísticas no corpus textual.

Imagem 9 – Sumário (parte 2)

Palavras mais frequentes no *corpus*

Palavra	**Frequência**
Artificial	208
IA	206
Inteligência	192
Dados	142
Máquina	140

Criada pela autora através da plataforma Voyant Tools (2023)

Palavras distintivas (comparadas com o restante do *corpus*):

Categoria/Contexto	**Palavras Distintas**
1. 037-050.indd	Ia (59), agentes (12), agente (12), sichman (10), 101 (13)
2. 085-094.indd	Ia (46), sucesso (12), fev (8), am (8), 2021 (16)
3. Algoritmos e IA	frazão (9), ana (9), jota (7), constituição (7), colunas (7)
4. Algoritmos para construção	2014 (44), especialista (19), paciente (29), prudente (14), vol (13)
5. Inteligência Artificial	2018 (50), 133 (15), univ (14), privacidade (28), jan (14)
6. Inteligência Artificial	concretas (2), vislumbrem (1), vislumbra (1), sofisticação (1), resguardando (1)
7. Interação homem máquina	2022 (40), teccogsn (35), patamar (35), interação (42), jul (35)
8. Jornal Do Brasil	pessoa (12), testamento (4), morte (4), edyanne (4), direitos (8)
9. Jornal O Dia	colégio (8), escolar (4), medidas (5), santo (3), nota (3)

Fonte: elaborada pelas autoras através da plataforma *Voyant Tools* (2023).

A ferramenta *Voyant Tools* desempenhou ao longo desta pesquisa um papel importante na análise textual e mineração de textos, ofereceu uma plataforma eficaz e intuitiva no quesito exploração e compreensão de dados textuais. Sua utilidade ofereceu as visualizações interativas, ferramentas de análise estatística e recursos básicos, mas, não menos importantes de linguagem natural.

Ao permitir a visualização dinâmica de padrões, termos-chave e relações dentro do

corpus textual analisado, o *Voyant Tools* facilitou a identificação de *insights*, padrões e tendências, tornando-se uma ferramenta importante para um grande conjunto de dados textuais.

Embora a ferramenta seja útil para análise de dados textuais e mineração de texto, é importante reconhecer os seus pontos negativos, como o fato de operar principalmente como uma ferramenta on-line, o que pode representar um desafio se trabalha com documentos sensíveis ou em ambientes onde não se tem uma boa conectividade.

Outro ponto negativo se dá pela limitação a grandes conjuntos de dados, neste trabalho foram analisados 20 artigos com uma média de 20 páginas cada um, que se mostrou satisfatório na análise, porém, para conjunto de dados muito extensos, o desempenho da ferramenta pode ser afetado, resultando em um tempo maior de processamento. Para algumas tarefas avançadas de mineração de texto, como análise profunda de tópicos ou reconhecimento de entidades, a ferramenta pode não oferecer todas as funcionalidades necessárias, exigindo o uso de ferramentas mais especializadas.

5. CONCLUSÃO

A ferramenta utilizada neste trabalho demonstrou-se eficaz na análise e mineração de dados textuais, evidenciando sua relevância na interpretação e exploração de grandes volumes de informação. Além de comprovar a funcionalidade do *Voyant Tools*, este estudo ressaltou a importância das abordagens metodológicas adotadas para transformar dados brutos em conhecimento significativo, especialmente em um cenário onde a velocidade e a quantidade de informações representam desafios constantes.

A capacidade da ferramenta em fornecer *insights* profundos e visualizações intuitivas reforça sua utilidade na compreensão da crescente complexidade dos dados textuais. Em um contexto onde a confiabilidade da informação e a rapidez de acesso são essenciais, o *Voyant Tools* se mostrou uma solução eficaz para facilitar a análise, promovendo estratégias mais estruturadas na extração de significados relevantes.

Ao aprofundar a exploração de suas funcionalidades, este trabalho destacou não apenas a eficiência da ferramenta, mas também a necessidade de aprimoramento contínuo das abordagens para lidar com o volume expressivo de informações presentes no ambiente digital atual.

Com base nos resultados obtidos, este estudo abre caminhos para futuras pesquisas e aplicações em diversas áreas. A análise semântica pode ser aprofundada com técnicas avançadas de processamento de linguagem natural, permitindo uma compreensão mais refinada das relações entre os termos. Além disso, a comparação com outras ferramentas de análise textual pode ajudar a identificar abordagens mais eficazes. Outro possível avanço seria a aplicação dos métodos utilizados em diferentes domínios, como educação, saúde e inteligência artificial, tema tão relevante nos dias de hoje, explorando a interpretação de grandes volumes de texto em contextos específicos. Por fim, estudos futuros podem investigar como ferramentas como o *Voyant Tools* influenciam processos de decisão e automação na extração de conhecimento.

REFERÊNCIAS

ADMINISTRADORES. **Leitura, análise, interpretação e síntese textual: leitura.** Disponível em: https://administradores.com.br/artigos/leitura-analise-interpretacao-e-sintese-textual-leitura. Acesso em: 19 set. 2023.

BEPPLER, M; FERNANDES, A. Aplicação de text mining para a extração de conhecimento jurisprudencial. In: **PRIMEIRO CONGRESSO SUL CATARINENSE DE EDUCAÇÃO**, 2005.

DATASIDE. **Tipos de dados: estruturados, semiestruturados e não estruturados**. Disponível em: https://www.dataside.com.br/dataside-community/big-data/tipos-de-dados-estruturados-semi-estruturados-e-nao-estruturados. Acesso em: 4 nov. 2023.

ENCONTROGRAFIA. Ebook: Análise Textual Discursiva - Mosaico de Pesquisas Autorais. Disponível em: https://encontrografia.com/wp-content/uploads/2023/03/Ebook_Analise-Textual-Discu rsiva-mosaico-de-pesquisas-autorais.pdf. Acesso em: 19 set. 2023.

GARTNER, Inc. **The Future of Data and Analytics: Trends and Predictions 2022-2023**. Stamford, CT: Gartner. 2022.

GREGO, Maurício. Conteúdo digital dobra a cada dois anos no mundo. Exame, 2014. Disponível em: https://exame.com/tecnologia/conteudo-digital-dobra-a-cada-dois-anosno-mundo/. Acesso em: 7 nov. 2023.

IDC (International Data Corporation). **Análise textual e sua relevância para o mundo atual.** Disponível em: https://www.idc.com/. Acesso em: 21 set. 2023.

INF.UFPR. **A importância da mineração de dados na prevenção de acidentes de trânsito.** Disponível em: https://www.inf.ufpr.br/sbbd-sbsc2014/sbbd/proceedings/artigos/pdfs/127.pdf. Acesso em: 21 set. 2023.

LARHUD. **Voyant Tools.** Disponível em: http://www.larhud.ibict.br/index.php?title=Voyant_Tools. Acesso em: 21 set. 2023.

LIDDY, E. D. **Natural Language Processing**. In: Encyclopedia of Library and Information Science, 2nd ed. New York: Marcel Decker, Inc., 2003.

MORAIS, R. V.; AMBRÓSIO, L. A. **Mineração de textos: uma abordagem baseada em conceitos de inteligência artificial**. São Paulo: Editora UNESP, 2007.

PEZZINI, Marco. **Mineração de textos: conceitos e aplicações**. São Paulo: Saraiva, 1. ed., 2016. 182 p.

SALTON, G.; MCGILL, M. J. **Introduction to Modern Information Retrieval**. New York: John Wiley & Sons, 1983.

SAS. **Processamento de Linguagem Natural:** o que é e qual sua importância? Disponível em: https://www.sas.com/pt_br/insights/analytics/processamento-de-linguagem-natural.html. Acesso em: 19 set. 2023.

SCIELO. **A análise textual discursiva: mosaico de pesquisas autorais**. Disponível em: https://www.scielo.br/j/ciedu/a/wvLhSxkz3JRgv3mcXHBWSXB/?format=pdf&lang=pt. Acesso em: 19 set. 2023.

SILVA, L. A.; PERES, S. M.; BOSCARIOLI, C. **Introdução à Mineração de Dados**: com

aplicações em R. Rio de Janeiro: Elsevier, 2016.

TIBCO. **What is Real-Time Data?** Disponível em: https://www.tibco.com/pt-br/reference-center/what-is-real-time-data. Acesso em: 19 set. 2023.

TIBCO. **What is Text Analytics?** Disponível em: https://www.tibco.com/pt-br/reference-center/whatistextanalytics#:~:text=A%20an%C3%A1lise%20de%20texto%20combina,para%20derivar%20percep%C3%A7%C3%B5es%20e%20padr%C3%B5es. Acesso em: 19 set. 2023.

UFRN. Aplicação de Mineração de Texto na Análise de Sentimentos em Comentários de Política. Disponível em: https://repositorio.ufrn.br/bitstream/123456789/31417/1/Aplicacaomineracaotexto_Benicio_2020.pdf. Acesso em: 21 set. 2023.

APENDICE
Referências dos textos utilizados para análise com a ferramenta Voyant Tools

ALAVES, J. S., SILVA, M. M., SANTOS, J. L. (2023). Aplicação de redes neurais artificiais para classificação de imagens médicas. **Revista Redes**, n. 129, p. 87. Acesso em: 5 nov. 2023.

AMARAL, G. R.; XAVIER, F. A inteligência artificial e o novo patamar da interação humano-máquina. **TECCOGS: Revista Digital de Tecnologias Cognitivas**, n. 26, p. 06-43, 2022. Acesso em: 5 nov. 2023.

BRASIL, J. D. **Avanço da inteligência artificial gera busca por proteção de direitos**. Disponível em: https://www.jb.com.br/ciencia-e-tecnologia/2023/09/1045775-avanco-da-inteligencia-artificial-gera-busca-por-protecao-de-direitos.html. Acesso em: 4 nov. 2023.

FONSECA, R. P. (2019). **Fundos de Investimentos Baseados em Machine Learning**. Disponível em: https://repositorio.utfpr.edu.br/jspui/bitstream/1/24679/1/fundosinvestimentosmachinelearning.pdf. Acesso em: 5 nov. 2023.

FRAZÃO, A. **Algoritmos e inteligência artificial Repercussões da sua utilização sobre a responsabilidade civil e punitiva das empresas**. [s.l: s.n.]. Disponível em: https://www.professoraanafrazao.com.br/files/publicacoes/2018-05-16-Algoritmos_e_inteligencia_artificial.pdf. Acesso em: 5 nov. 2023.

FROZZA, R., Oliveira, M. V., & GATTI, M. S. V. (ano). **Fundos de Investimentos: uma abordagem utilizando Machine Learning**. Disponível em: https://repositorio.utfpr.edu.br/jspui/bitstream/1/24679/1/fundosinvestimentosmachine learning.pdf. Acesso em: 5 nov. 2023.

GABRIEL, A. et al. **Modelo de machine learning para detecção e contagem de pés de café (Coffea sp.) por análise de vídeo.** [s.l: s.n.].
Disponível em: https://ric.cps.sp.gov.br/handle/123456789/6472 . Acesso em: 5 nov. 2023.

HENRIQUES, I. Inteligência artificial e publicidade dirigida a crianças e adolescentes. [s.l: s.n.]. Disponível em: https://revista.internetlab.org.br/wp-content/uploads/2022/03/Inteligencia-artificial-e- publicidade-dirigida-a-criancas-e-adolescentes.pdf. Acesso em: 5 nov. 2023.

HOFFMANN-RIEM, W. BIG DATA E INTELIGÊNCIA ARTIFICIAL: desafios para o Direito. **REI - REVISTA ESTUDOS INSTITUCIONAIS**, v. 6, n. 2, p. 431–506, 23 set. 2020. Acesso em: 5 de nov. de 2023.

JORNAL O DIA. **Alunos de colégio particular usam inteligência artificial para criar imagens de colegas nuas**. Disponível em: https://odia.ig.com.br/rio-de-janeiro/2023/11/6734390-alunos-de-colegio-particular-us am-inteligencia-artificial-para-criar-imagens-de-colegas-nuas.html. Acesso em: 4 nov. 2023.

LUDERMIR, T. B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Estudos Avançados**, v. 35, n. 101, p. 85–94, abr. 2021. Acesso em: 5 nov. 2023.

PINOTTI, B.; OLIVEIRA, G. **Inteligência artificial e proteção de dados:** sobre a audodeterminação informativa e a manipulação informacional por machine learning. v. 26, p. 1809–1628, [s.d.]. Acesso em: 5 nov. de 2023.

RSD - Journal of Research in Dental Sciences. Evaluation of the effectiveness of two different anesthesia techniques in maxillary lateral incisors: a randomized clinical trial. RSD **Journal of Research in Dental Sciences**, [S.l.], v. 10, n. 4, p. 1737-1743, jul. 2021. Disponível em: https://rsdjournal.org/index.php/rsd/article/view/15296. Acesso em: 5 nov. 2023.

RIC - Rede de Computadores e Sistemas, São Paulo. **Big Data e o Modelo de Machine Learning Andrew**: aplicação no contexto educacional. 2021. Disponível em: https://ric.cps.sp.gov.br/bitstream/123456789/6472/1/bigdata_2021_2_andrew_model odemachinelearning.pdf. Acesso em: 5 nov. 2023.

SICHMAN, J. S. Inteligência Artificial e sociedade: avanços e riscos. **Estudos Avançados**, v. 35, n. 101, p. 37–50, abr. 2021. Acesso em: 5 de nov. 2023.

TOMASEVICIUS FILHO, E. Inteligência artificial e direitos da personalidade. **Revista da Faculdade de Direito**, Universidade de São Paulo, v. 113, p. 133–149, 5 ago. 2018. Acesso em: 5 nov. 2023.

UNOESTE. **Estudo da inteligência artificial aplicada na área da saúde**. Disponível em: https://unoeste.br/site/enepe/2014/suplementos/area/Exactarum/Computa%C3%A7%C3%A3o/ESTUDO%20DA%20INTELIG%C3%8ANCIA%20ARTIFICIAL%20APLIACADA%20NA%20%C3%81REA%20DA%20SA%C3%9ADE.pdf. Acesso em: 5 nov. 2023.

VALLE, B. DE M. **Tecnologia da informação no contexto organizacional**. Ciência da Informação, v. 25, n. 1, 1996. Acesso em: 5 nov. 2023.