

# Utilização de Voyant Tools para Compreensão de Documentos Não Estruturados

GABRIELA ALVES DA ROCHA<sup>1</sup>; ELOÍZA MARTINS PRIMO CAPELOCI<sup>2</sup>

<sup>1</sup> Discente em Big Data No Agronegócio na FATEC Pompeia “Shunji Nishimura”, Pompeia-SP. [gabriela.rocha3@fatec.sp.gov.br](mailto:gabriela.rocha3@fatec.sp.gov.br)

<sup>2</sup> Docente do curso de Big Data no Agronegócio, FATEC Pompeia “Shunji Nishimura”, Pompeia-SP. [eloiza.capeloci@fatec.sp.gov.br](mailto:eloiza.capeloci@fatec.sp.gov.br)

## RESUMO

A análise e mineração de dados textuais tem sido importante na interpretação de dados textuais gerados em um cenário cada vez mais informatizado, onde dados são gerados a todo tempo. Estima-se que aproximadamente 2,0 quintilhões de *bytes* de dados assumem a forma de texto diariamente. A mineração de dados não estruturados representa o processo de extrair informações úteis e conhecimento a partir de uma certa quantidade de dados textuais, constituído por texto em linguagem natural, como por exemplo documentos em formato PDF, e-mails e postagens em redes sociais. Desde a origem da computação, tanto linguistas quanto especialistas em recuperação da informação têm criado *softwares* com o intuito de identificar padrões de linguagem que não são identificados em uma leitura convencional. Este estudo aborda a análise de conteúdo textual por meio da utilização da ferramenta *web* Voyant Tools que explora análise textual e mineração de dados textuais. O objetivo central é explorar como esta ferramenta pode contribuir para a compreensão de documentos não estruturados, que não seguem um padrão definido por conta da sua variedade linguística, contexto e intenção, estilo de escrita e a inclusão de elementos não textuais, como gráficos e tabelas, imagens e vídeos. Para a análise também foram incorporadas algumas técnicas de Processamento de Linguagem Natural (PLN) relevantes para uma análise textual mais precisa e significativa. Dentre as técnicas, destacam-se a remoção das *stopwords*, contagem de palavras e análise de frequência de cada termo dentro de um corpus textual.

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

## INTRODUÇÃO

A análise textual e mineração de textos desempenha um papel fundamental na compreensão de dados textuais, em um mundo cada vez mais voltado a informação. De acordo com um estudo da Gartner (2022), em 2023, 80% dos dados gerados no mundo serão dados estruturados, incluindo dados de textos. Isso significa que cerca de 2,0 quintilhões de *bytes* de dados por dia são gerados sob forma textual. Atualmente, nota-se cada vez mais a importância dos dados para as empresas e para

toda sociedade em geral. Estes dados são gerados em volumes cada vez maiores e de acordo com Grego (2014), se todo conteúdo digital, tais como músicas, livros, filmes, documentos e dados, produzido pela humanidade fosse colocado em *iPads* e empilhados, a altura da pilha seria de cerca de dois terços da distância entre a Terra e a Lua. E complementa que a tendência é que essa pilha imaginária possa alcançar 6,6 vezes essa distância nos anos que virão.

Existem vários conceitos para lidar com o crescimento exponencial dos dados, o termo *Big Data* por exemplo trata desse conceito de grande quantidade de dados, e como já citado 80% dos dados gerados no mundo diariamente estão sobre a forma de texto.

A mineração de textos possibilita a extração de informações, através da análise quantitativa e qualitativa, de grandes coleções de documentos como, por exemplo, artigos científicos, notícias jornalísticas, páginas *web*, redes sociais, entre outros. O método permite a identificação de padrões e o estabelecimento de relações entre eles com base na frequência e temática de termos (Pezzini, 2016). Na prática, “a mineração de textos define um processo que auxilia na descoberta de conhecimento inovador a partir de documentos textuais, que pode ser utilizado em diversas áreas do conhecimento” (MORAIS; AMBRÓSIO, 2007).

A mineração de dados – expressos sob a forma de textos, imagens, vídeos, tabelas, grafos, sinais, outros – é um método que auxilia no processo analítico de como a pesquisa é conduzida, produzida, compartilhada, consumida e apreendida pelas subjetividades de quem a acessa ou que por ela é (re) classificada. Nesse sentido, é necessária sua indagação à luz de um regime sociotécnico de informação marcado por uma visibilidade informacional atualmente (PIMENTA, 2017).

Neste estudo, utilizamos a ferramenta *web* Voyant Tools para analisar um corpus textual específico. Para Trask (2004), corpus é “um conjunto de textos escritos ou falados numa língua, disponível”, para análise tendo como objetivo principal explorar e extrair *insights* a partir do corpus selecionado, extraindo informações que normalmente não são visíveis em uma análise manual.

Esta ferramenta oferece uma série de recursos para analisar textos de maneira eficiente, por meio da utilização superficial de algoritmos de linguagem natural, sendo possível explorar a frequência de cada palavra no texto, analisar os tópicos mais frequentes em cada um deles. Isso ajuda a descobrir informações ocultas, padrão de

linguagem e tendências que podem informar a compreensão do corpus e do contexto ao qual ele pertence.

A análise textual e mineração de dados textuais no Voyant Tools oferece uma abordagem baseada em dados para explorar e interpretar textos, utilizando algoritmos de linguagem natural, permitindo interpretar o significado de dados textuais, usada para analisar documentos, a fim de extrair o máximo de conhecimento e informações que na maioria das vezes estão ocultas ou passam despercebidas em uma primeira análise ou leitura.

Este trabalho tem como objetivo demonstrar como o Voyant Tools pode ser usado para realizar análise de conteúdo textual eficaz e fornecer informações significativas a partir de documentos não estruturados.

## **MINERAÇÃO DE TEXTOS**

A evolução das técnicas de mineração de texto resultou em melhorias significativas. Esses progressos possibilitaram a automatização de rotinas e processos que, anteriormente, demandava esforço humano e isso resultou em uma maior eficiência, contribuindo para a redução de 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; MCGIL, 1983), na qual através de um conjunto de rotinas de processamento e análise de padrões, a informação é recuperada a partir de dados textuais, gerando, conseqüentemente conhecimento. Assim destaca-se que a fundamentação dessa área está ligada às 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 Boscaroli (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 todo dado processado é garantia de informação. Por fim, conhecimento pode ser definido como a informação explorada com algum propósito

específico, ou seja, utilizada para, por exemplo, tomada de decisão, construção de cenários, entre outros.

Nesse sentido, pode-se entender dados como matéria-prima indispensável para análise. Ainda de acordo com Silva, Peres e Boscaroli (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 as peculiaridades de cada tipo de dado demandam rotinas específicas para seu processamento. De forma 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 por 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.

A análise de dados armazenados em formato não estruturado é considerada uma atividade mais complexa, se comparada à análise de dados estruturados, isto se dá pelo fato de os dados possuírem características da não estruturação. Logo, são necessárias técnicas específicas para tratamento deste tipo de dados. Este conjunto de técnicas e ferramentas fazem parte da área de recuperação de Informações, mais conhecida como Descoberta de Conhecimento em Textos (*Knowledge Discovery from Text – KDT*). De acordo com Bepler et al, 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 utiliza 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.

### **Imagem 1 – Tipos de Dados**



Fonte: DATASIDE (2021)

A partir da identificação de padrões, palavras-chave e relações semânticas nos textos, a mineração de texto permite categorizar e classificar os dados de forma coerente e acessível. Isso facilita a busca, análise e interpretação das informações, proporcionando *insights* valiosos para a tomada de decisões em uma variedade de contextos, como análise de sentimentos, processamento de documentos legais, extração de informações em pesquisas acadêmicas, entre outros.

## 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 e 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.

## **FERRAMENTA VOYANT TOOLS**

A ferramenta Voyant Tools que é uma ferramenta on-line e *open source* de análise de dados textuais, desenvolvido pelos professores canadenses Stéfán Sinclair e Geoffrey Rockwell em 2003. O *software* Voyant Tools foi desenvolvido principalmente usando tecnologias da web, como HTML, *JavaScript* e CSS. Portanto, não é criado em uma única linguagem de programação, mas é uma aplicação que funciona em navegador da *web* (SINCLAIR; ROCKWELL, 2003).

A ferramenta foi desenvolvida para facilitar a leitura e análise de textos digitais, mas rapidamente se tornou popular entre estudantes e público em geral, dada a facilidade de sua utilização. A intenção da ferramenta é fornecer uma maneira fácil e acessível para visualizar e explorar grandes volumes de dados textuais.

O Voyant Tools oferece muitos recursos para realizar análise textual, incluindo: geração de nuvens de palavras; visualização de frequência; identificação de palavras relacionadas e análise de tendências e padrões. Nos últimos anos a ferramenta tem sido usada para uma variedade de propósitos, incluindo pesquisa acadêmica, ensino de aprendizagem e análise de dados sociais.

Pesquisadores têm usado o Voyant Tools para analisar textos em amplos contextos, entre eles: literatura, ensino de idiomas, saúde e arquitetura de sistemas.

## **MATERIAL E MÉTODOS**

Para realizar este projeto foram feitas buscas sistematizadas em bases de dados eletrônicas (SCIELO e Google Acadêmico), além de sites e livros e jornais especializados no tema. As buscas não foram limitadas por língua ou data de publicação e foram encontrados dados e pesquisas de julho a setembro de 2023.

Foram pesquisados 20 trabalhos relacionados a inteligência artificial, algoritmos de aprendizado de máquinas, os riscos e impactos da inteligência artificial na sociedade, *big data* e tecnologia da informação para a busca bibliográfica foram utilizadas as palavras-chave: inteligência artificial; aprendizado de máquinas; algoritmos; sociedade; *Big Data*; Tecnologia da Informação.

Os PDFs e artigos jornalísticos pesquisados têm como tema a importância da inteligência artificial, uso de algoritmos de aprendizagem de máquinas e os impactos dessa tecnologia na sociedade.

Os *uploads* desses artigos foram realizados na Ferramenta Voyant Tools e após o *upload* destes arquivos foram removidas, fórmulas, tabelas, imagens e as *stopwords*

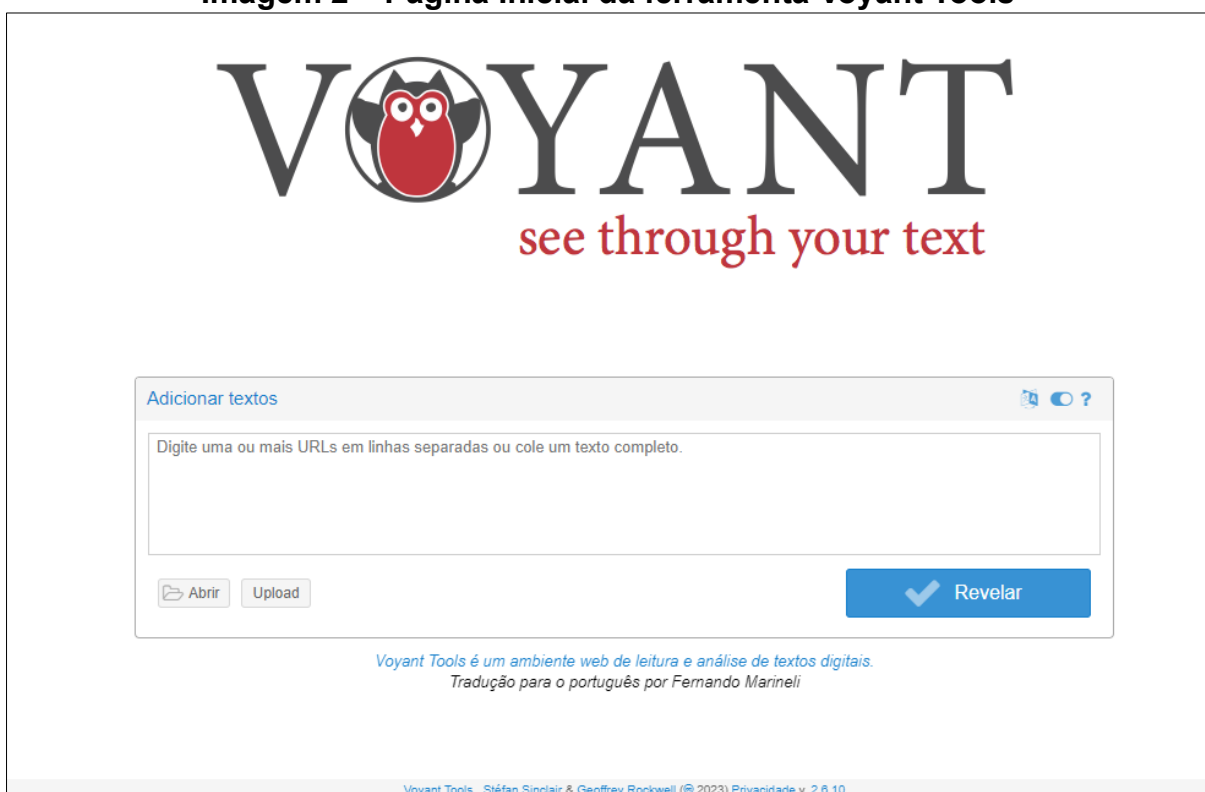
dos textos para identificar dentre o tema escolhido quais as palavras chaves e quais as palavras mais relevantes nos textos. A ferramenta oferece diversas opções e possui uma interface intuitiva para facilitar a utilização.

## RESULTADOS E DISCUSSÃO

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

A Imagem 2 mostra a página inicial da ferramenta, onde são carregados os arquivos e os textos para análise, a interface gráfica é bastante intuitiva.

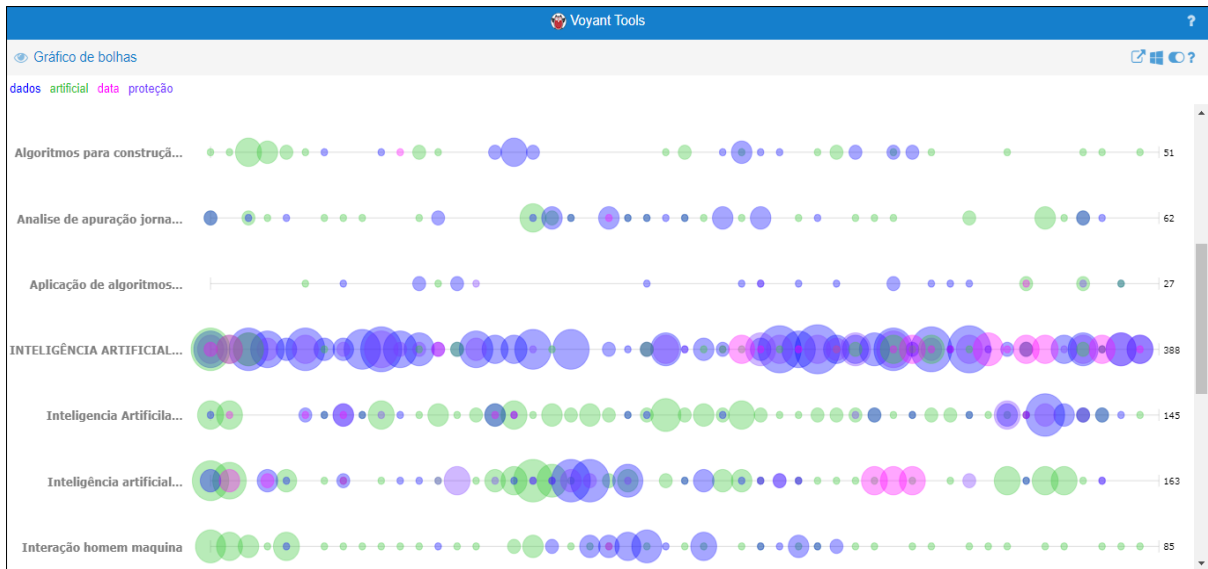
**Imagem 2 – Página Inicial da ferramenta Voyant Tools**



Fonte: Criada pela autora através da plataforma Voyant Tools (2023)

Após análise foi gerada uma nuvem de palavras e destas foram analisados 245 termos mais comuns. A nuvem de palavras posiciona as palavras de tal forma que os termos mais frequentes são posicionados centralmente e com tamanhos maiores. Como o algoritmo atravessa a lista e continua desenhando mais palavras o mais

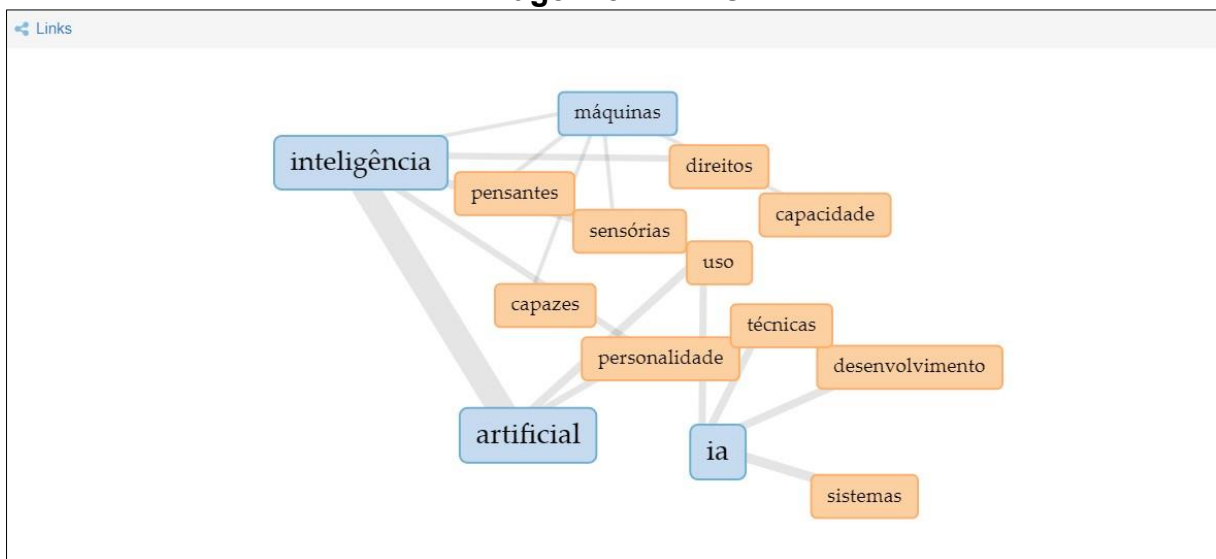




Fonte: Criada pela autora através da plataforma Voyant Tools (2023)

A Imagem 5 estabeleceu através da palavra ‘inteligência’ uma relação semântica entre as demais palavras e frases que estão associadas a um corpus textual.

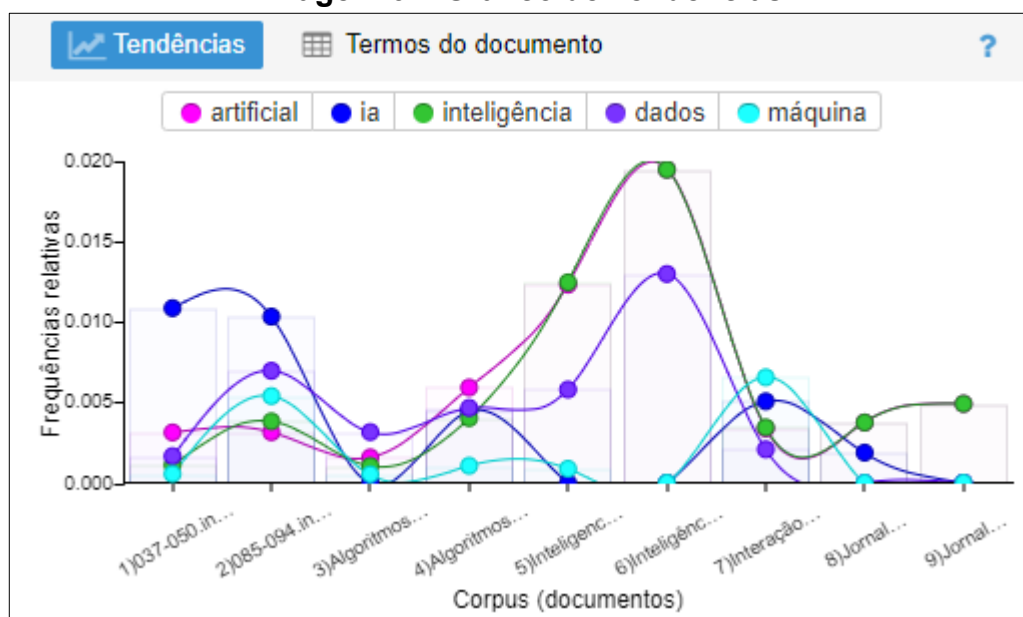
### Imagem 5 – Links



Fonte: Criada pela autora através da plataforma Voyant Tools (2023)

Neste gráfico de distribuição gerado pela ferramenta é possível ver os termos mais usados em todo corpus, ou dentro de um único documento quando apenas um é selecionado. Nesta imagem é visto todo o corpus textual.

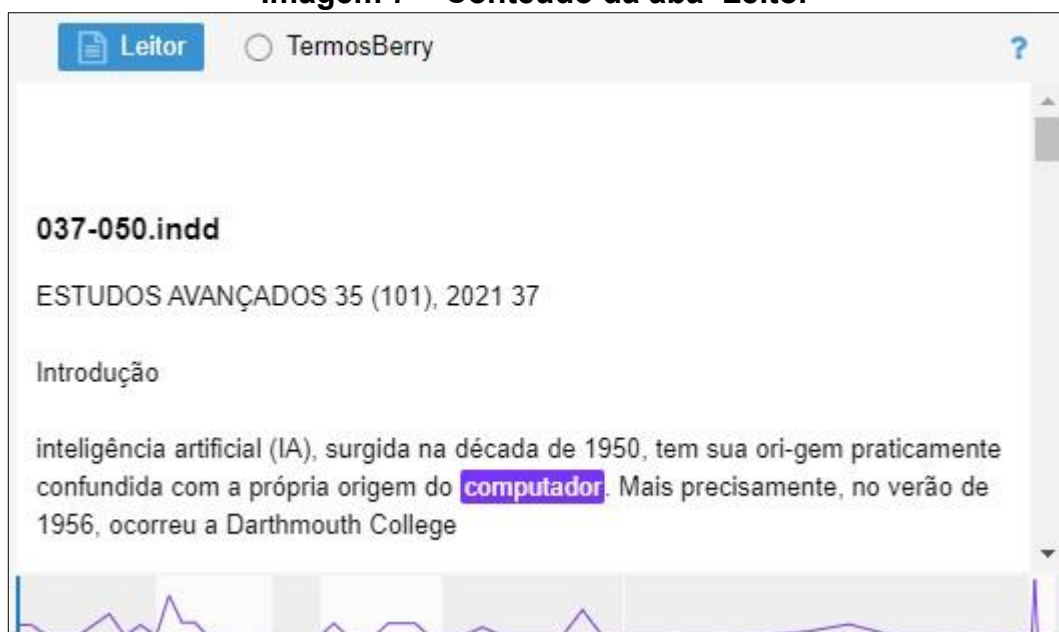
Imagem 6 – Gráfico de Tendências



Fonte: Criada pela autora 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: Criada pela autora através da plataforma Voyant Tools (2023).

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.

### Imagem 8 - Sumário (parte 1)

Este corpus possui 9 documentos 40,438 formas únicas de palavras. Criado há 2 horas.

Extensão do documento: 

- Mais longo: **Interação homem máquina** (15400); **Inteligencia Artificila...** (6898); **037-050.indd** (5432); **Algoritmos para construçã...** (4726); **085-094.indd** (4451)
- Mais curto: **Inteligência Artificial...** (154); **Jornal O Dia** (407); **Jornal Do Brasil** (1067); **Algoritmos e IA** (1903); **085-094.indd** (4451)

Densidade vocabular: 

- Mais alto: **Inteligência Artificial...** (0.662); **Jornal O Dia** (0.548); **Jornal Do Brasil** (0.473); **Algoritmos e IA** (0.412); **037-050.indd** (0.362)
- Mais baixo: **Interação homem máquina** (0.241); **Algoritmos para construçã...** (0.287); **Inteligencia Artificila...** (0.287); **085-094.indd** (0.333); **037-050.indd** (0.362)

Média de palavras por frase: 

- Mais alto: **Inteligência Artificial...** (38.5); **Jornal O Dia** (29.1); **Inteligencia Artificila...** (26.9); **Interação homem máquina** (24.6); **Jornal Do Brasil** (23.2)
- Mais baixo: **085-094.indd** (18.5); **Algoritmos e IA** (21.1); **Algoritmos para construçã...** (21.8); **037-050.indd** (22.2); **Jornal Do Brasil** (23.2)

Readability Index: 

- Mais alto: **Inteligência Artificial...** (17.228); **Algoritmos e IA** (16.988); **Algoritmos para construçã...** (16.118); **037-050.indd** (15.122); **085-094.indd** (14.728)
- Mais baixo: **Jornal Do Brasil** (12.099); **Jornal O Dia** (13.512); **Interação homem máquina** (13.807); **Inteligencia Artificila...** (14.037); **085-094.indd** (14.728)

Criada pela autora 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:

- artificial** (208); **ia** (206); **inteligência** (192); **dados** (142); **máquina** (140)

Palavras distintas (comparadas com o restante do corpus):

- 037-050.indd**: **ia** (59), **agents** (12), **agent** (12), **sichman** (10), **101** (13).
- 085-094.indd**: **ia** (46), **sucesso** (12), **fev** (8), **am** (8), **2021** (16).
- Algoritmos e IA**: **frazão** (9), **ana** (9), **jota** (7), **constituicao** (7), **colunas** (7).
- Algoritmos para construçã...**: **2014** (44), **especialista** (19), **paciente** (29), **prudente** (14), **vol** (13).
- Inteligencia Artificila...**: **2018** (50), **133** (15), **univ** (14), **privacidade** (28), **jan** (14).
- Inteligência Artificial...**: **concretas** (2), **vislumbrem** (1), **vislumbra** (1), **sofisticacão** (1), **resguardando** (1).
- Interação homem máquina**: **2022** (40), **teccogsn** (35), **patamar** (35), **interacção** (42), **jul** (35).
- Jornal Do Brasil**: **pessoa** (12), **testamento** (4), **morte** (4), **edyanne** (4), **direitos** (8).
- Jornal O Dia**: **colégio** (8), **escolar** (4), **medidas** (5), **santo** (3), **nota** (3).

Criada pela autora 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.

## **CONCLUSÃO**

A ferramenta utilizada neste trabalho se mostrou bastante eficaz quando a análise de dados textuais e mineração de textos, mostrando a importância deste *software* em interpretar e explorar um grande volume de dados. Este trabalho não apenas demonstrou a eficácia do Voyant Tools, mas também mostrou a importância das abordagens utilizadas na busca de conhecimento em meio a informações que chegam muito rápido e em grande quantidade.

A capacidade da ferramenta em fornecer *insights* profundos e visualizações intuitivas destaca a sua importância como uma ferramenta importante para lidar com a crescente complexidade dos dados textuais. Em um cenário onde a velocidade, o volume e a veracidade de dados são desafios constantes, a ferramenta Voyant Tools não apenas simplificou a análise textual, mas também enfatizou a necessidade de estratégias para a compreensão e extração de significado de grandes conjuntos de dados. Ao explorar a fundo as funcionalidades desta ferramenta, este trabalho destacou a relevância das abordagens adotadas de maneira eficaz o grande volume expressivo de informações presentes no contexto atual.

## **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 de set. de 2023.

BEPPLER, M; FERNANDES, A. Aplicação de text mining para a extração de conhecimento jurisprudencial. In: PRIMEIRO CONGRESSO SUL CATARINENSE DE Educação, 2005. Acesso em: 07 de nov. de 2023.

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: 04 de nov. de 2023.

ENCONTROGRAFIA. Ebook: Análise Textual Discursiva - Mosaico de Pesquisas Autorais. Disponível em: [https://encontrografia.com/wp-content/uploads/2023/03/Ebook\\_Analise-Textual-Discursiva-mosaico-de-pesquisas-autorais.pdf](https://encontrografia.com/wp-content/uploads/2023/03/Ebook_Analise-Textual-Discursiva-mosaico-de-pesquisas-autorais.pdf). Acesso em: 19 de set. de 2023.

GARTNER, Inc. (2022). The Future of Data and Analytics: Trends and Predictions 2022-2023. Stamford, CT: Gartner. Acesso em: 07 de nov. de 2023.

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-anos-no-mundo/>. Acesso em: 07 de nov. de 2023.

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

LARHUD. Voyant Tools. Disponível em: [http://www.larhud.ibict.br/index.php?title=Voyant\\_Tools](http://www.larhud.ibict.br/index.php?title=Voyant_Tools). Acesso em: 21 de set. de 2023.

LIDDY, E. D. Natural Language Processing. In: Encyclopedia of Library and Information Science, 2nd ed. New York: Marcel Decker, Inc., 2003. Acesso em 09 de nov. de 2023.

MORAIS, R. V.; AMBRÓSIO, L. A. Mineração de textos: uma abordagem baseada em conceitos de inteligência artificial. São Paulo: Editora UNESP, 2007. Acesso em: 21 de set. de 2023.

O que é processamento de linguagem natural? Disponível em: [https://www.sas.com/pt\\_br/insights/analytics/processamento-de-linguagem-natural.html](https://www.sas.com/pt_br/insights/analytics/processamento-de-linguagem-natural.html). Acesso em: 05 de nov. de 2023.

PEZZINI, Marco. Mineração de textos: conceitos e aplicações. São Paulo: Saraiva, 2016. 182 p. 1. ed. Acesso em: 21 de set. de 2023.

SALTON, G.; MCGILL, M. J. Introduction to Modern Information Retrieval. New York: John Wiley & Sons, 1983. Acesso em: 31 de out. de 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 de set. de 2023.

SILVA, L. A.; PERES, S. M.; BOSCARIOLI, C. Introdução à Mineração de Dados: com aplicações em R. Rio de Janeiro: Elsevier, 2016. Acesso em: 31 de out. de 2023.

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

TIBCO. What is Text Analytics? Disponível em: <https://www.tibco.com/pt-br/reference-center/what-is-text-analytics#:~:text=A%20an%C3%A1lise%20de%20texto%20combina,para%20derivar%20percep%C3%A7%C3%B5es%20e%20padr%C3%B5es>. Acesso em: 19 de set. de 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\\_Be\\_nicio\\_2020.pdf](https://repositorio.ufrn.br/bitstream/123456789/31417/1/Aplicacaomineracaotexto_Be_nicio_2020.pdf). Acesso em: 21 de set. de 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*, 129, 87. Acesso em 05 de nov. de 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 05 de nov. de 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-protacao-de-direitos.html>. Acesso em 04 de nov. de 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: 05 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.professoraanafraza.com.br/files/publicacoes/2018-05-16-Algoritmos\\_e\\_inteligencia\\_artificial.pdf](https://www.professoraanafraza.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/fundosinvestimentosmachinelearning.pdf>. Acesso em: 05 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: 05 de 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: 05 de nov. de 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: 05 nov. 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-usam-inteligencia-artificial-para-criar-imagens-de-colegas-nuas.html>. Acesso em: 04 de nov. de 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 ARTIFICIAL INTELLIGENCE AND DATA PROTECTION: ABOUT INFORMATIVE AUDODETERMINATION AND INFORMATIONAL MANIPULATION BY MACHINE LEARNING. v. 26, p. 1809–1628, [s.d.]. Acesso em: 05 de 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: 05 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\\_modelodemachinelearning.pdf](https://ric.cps.sp.gov.br/bitstream/123456789/6472/1/bigdata_2021_2_andrew_modelodemachinelearning.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 05 de nov. de 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%20APLICADA%20NA%20%C3%81REA%20DA%20SA%C3%9ADE.pdf>. Acesso em: 05 de novembro de 2023.

VALLE, B. DE M. Tecnologia da informação no contexto organizacional. Ciência da Informação, v. 25, n. 1, 1996. Acesso em: 05 de nov. de 2023.  
[Morais/publication/268289851\\_A\\_TECNOLOGIA\\_DA\\_INFORMACAO\\_NA\\_CONTABILIDADE/links/574b918108ae2e0dd301a872/A-TECNOLOGIA-DA-INFORMACAO-NA-CONTABILIDADE.pdf](https://www.unoeste.br/revista/publication/268289851_A_TECNOLOGIA_DA_INFORMACAO_NA_CONTABILIDADE/links/574b918108ae2e0dd301a872/A-TECNOLOGIA-DA-INFORMACAO-NA-CONTABILIDADE.pdf) acesso em 05 de nov. de 2023.

