

10. REVOLUCIONANDO A PESQUISA CIENTÍFICA COM A PLATAFORMA KNIME ANALYTICS

Fernanda Farinelli

10.1 INTRODUÇÃO

A pesquisa em Ciências Sociais Aplicadas (CSA) se concentra em compreender e analisar os fenômenos sociais, culturais e econômicos que afetam a sociedade. Essa área abrange um vasto campo de estudo, englobando disciplinas como Ciência da Informação, Economia, Sociologia, Psicologia, Administração, entre outras. Seu principal foco reside na compreensão das complexas dinâmicas sociais, econômicas e culturais que moldam nosso mundo. Ela desempenha um papel fundamental na busca por soluções para os desafios contemporâneos que enfrentamos, como a desigualdade social, a sustentabilidade ambiental, o desenvolvimento econômico e a melhoria da qualidade de vida (Gil, 2008; Minayo, 2001).

Nos últimos anos, a pesquisa no campo das CSA tem passado por uma transformação significativa com o uso crescente da Tecnologia da Informação (TI). A TI, incluindo a *internet*, dispositivos móveis e inteligência artificial trouxeram novas perspectivas e ferramentas para os pesquisadores dessa área. O casamento entre tal campo e a TI tem se revelado uma aliança poderosa, transformando a maneira como conduzimos pesquisas e compreendemos a sociedade.

O processo de pesquisa envolve diversas etapas essenciais para a produção de conhecimento. Inclui, dentre outras etapas, a coleta de dados, que pode ser conduzida por meio de métodos quantitativos ou qualitativos, seguida pelo tratamento ou elaboração dos dados para, então, passar pela etapa de análise dos dados com o intuito de identificar padrões e responder às questões de pesquisa. Posteriormente, os resultados são interpretados e as conclusões são comunicadas por meio de relatórios ou publicações acadêmicas, desempenhando um papel fundamental no

avanço do conhecimento em uma determinada área (Gil, 2002, 2008; Lakatos; Marconi, 2010; Marconi; Lakatos, 2002).

As tecnologias digitais têm se tornado recursos essenciais para os pesquisadores nos diversos campos de pesquisa. Observa-se que elas têm impactado a pesquisa científica ao trazer alternativas para a coleta e análise de dados em grande escala, assim como oferecer melhores possibilidades de tratamento, organização e visualização dos resultados (Swiech; Francisco; Lima, 2016).

As novas possibilidades da TI também revolucionaram a forma como os cientistas se comunicam, colaboram e compartilham suas pesquisas (Muel-ler, 1994). A TI também tem desempenhado um papel fundamental na disseminação de pesquisas. As publicações acadêmicas, agora, são facilmente encontradas *on-line*, tornando o conhecimento mais acessível a uma audiência global. Ademais, as mídias sociais e as plataformas de compartilhamento de informações permitem que os pesquisadores alcancem um público mais amplo e envolvam a sociedade em debates relevantes.

Além disso, a análise de *big data* permite aos pesquisadores examinarem grandes conjuntos de informações de forma mais eficiente e detalhada, abrindo novas possibilidades de pesquisa. As vastas quantidades de dados disponíveis *on-line* abriram um mundo de novas possibilidades para a pesquisa. Pesquisadores de diversas disciplinas podem explorar uma variedade de fontes de dados, desde os demográficos e econômicos até informações sobre comportamento humano e opiniões públicas (Leonelli, 2022). As redes sociais, por exemplo, fornecem uma rica fonte de dados sobre comportamentos sociais e interações humanas, que podem ser analisados para obter *insights* sobre tendências culturais e comportamentais.

Essa disponibilidade de dados, muitas vezes em tempo real, permite que os pesquisadores realizem estudos longitudinais abrangentes, analisem tendências globais e até mesmo investiguem fenômenos em escala macro, proporcionando uma compreensão mais profunda e holística dos desafios e oportunidades em seus campos de estudo (Leonelli, 2022; Moura; Amorim, 2014). Além disso, a capacidade de acessar dados de fontes diversas e geograficamente dispersas promove a colaboração interdisciplinar e internacional, enriquecendo ainda mais a pesquisa e impulsionando a inovação científica (Garcia; Vieira; Vivacqua; França; Dias, 2020).

Essa transformação não é apenas uma resposta às demandas da era digital, mas também uma oportunidade para impulsionar o rigor, a eficiência e a amplitude da pesquisa. Nesse contexto, este capítulo tem como objetivo explorar a influência e o impacto da TI na pesquisa e metodologia científica, destacando a *plataforma KNIME* (Konstanz Information Miner), *Knime Analytics Platform*, como uma ferramenta multifacetada e versátil que atende a uma variedade de etapas e necessidades na pesquisa científica sobretudo nas CSA.

À medida que se avança nessa jornada pela interseção entre a pesquisa em CSA e TI, são exploradas as diversas formas pelas quais a ferramenta *KNIME* pode enriquecer os estudos, identificando as etapas metodológicas em que sua aplicação é pertinente. A expectativa é que este capítulo possa capacitar os pesquisadores a adotarem esse tipo de tecnologia para o aprimoramento de suas pesquisas.

10.2 PLATAFORMA KNIME ANALYTICS

A análise de dados evoluiu de um campo especializado para uma ferramenta fundamental em quase todos os setores da sociedade. À medida que a era da informação avança, a quantidade de dados disponíveis se torna cada vez mais impressionante. Com isso, surge a necessidade de ferramentas que possam extrair significado e *insights* valiosos desse oceano de informações.

É nesse contexto que o *KNIME* se destaca como uma poderosa ferramenta de código aberto para análise de dados, processamento de informações e criação de fluxos de trabalho analíticos. A ferramenta *KNIME* é muito mais do que uma simples ferramenta, é uma plataforma completa e acessível que emerge como uma solução poderosa e versátil para lidar com os desafios da coleta, preparação, análise e visualização de dados. Atualmente, a ferramenta está na versão 5.1 com uma interface gráfica moderna e aprimorada, contribuindo para redução na curva de aprendizado para novos usuários (KNIME, 2022a).

10.2.1 DOWNLOAD E INSTALAÇÃO

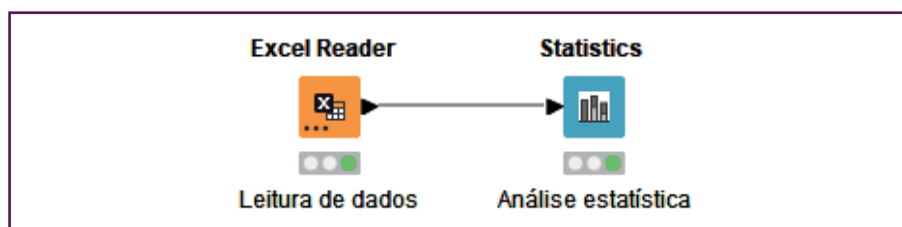
A versão mais recente da *Plataforma de Análise KNIME*, a versão 5.1.1, está disponível para *download* no *site* oficial¹⁰³, e é compatível com os sistemas operacionais para *Windows*, *Linux* e *macOS*. Para detalhes sobre a instalação consulte o guia oficial (KNIME, 2023b) fornecido pelo fabricante.

10.2.2 CONCEITOS BÁSICO DA FERRAMENTA

A *plataforma KNIME* se fundamenta na ideia do paradigma da programação visual na qual implica que um programa seja criado usando um grafo como um fluxo de dados (Khodnenko; Ivanov; PROKOFIEV; Lantseva, 2020; KNIME, 2023c).

Em seu ambiente de programação visual, adota-se o conceito de fluxos de dados, também conhecidos como *pipelines*. Isso permite que os usuários construam fluxos de trabalho (Figura 1) de forma intuitiva e eficiente, combinando diversos componentes e etapas para processar dados. Um fluxo de trabalho (workflow) é uma representação gráfica de uma sequência de etapas ou operações que descrevem todo o processo de análise ou processamento de dados (Berthold *et al.*, 2006; Berthold *et al.*, 2009; Hayasaka; Silipo, 2023; KNIME, 2022a; 2023c). É uma maneira visual de organizar e executar tarefas de análise de dados em um ambiente interativo.

figura 1 - Exemplo de fluxo de trabalho visual do KNIME



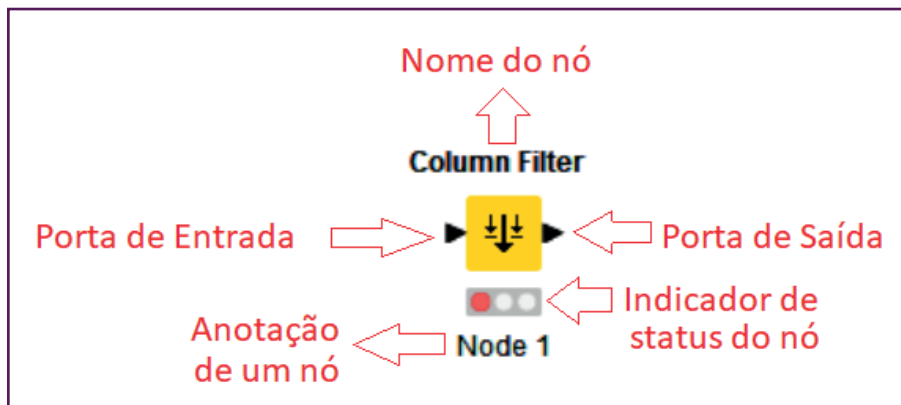
Fonte: Elaborado pela autora (2023).

103 Disponível em: <https://www.knime.com/downloads>. Acesso em: 17 set. 2023.

Nesse contexto, os usuários podem criar fluxos de trabalho compostos por nós que manipulam os dados, e esses dados são transportados entre os nós por meio de conexões. Um “nó” refere-se a uma unidade de processamento individual. Cada nó desempenha uma função específica em um fluxo de trabalho, e pode realizar ações como leitura de dados, pré-processamento, modelagem estatística, visualização, entre outros.

Na Figura 2, que se segue, apresenta-se uma minuciosa análise da anatomia de um nó, destacando seus componentes e características essenciais. Cada nó geralmente possui portas que servem como entrada e saída de dados respectivamente à esquerda e à direita do nó. Cada nó do repositório de nós tem um nome único que o identifica no contexto geral do *KNIME*. O “nome do nó” refere-se simplesmente à designação ou rótulo que é atribuído a um nó específico. Esses nomes são frequentemente escolhidos de forma a descrever sucintamente a função ou o propósito do nó, o que facilita a compreensão do usuário da ferramenta (KNIME, 2023c).

Figura 2 - Anatomia de um nó

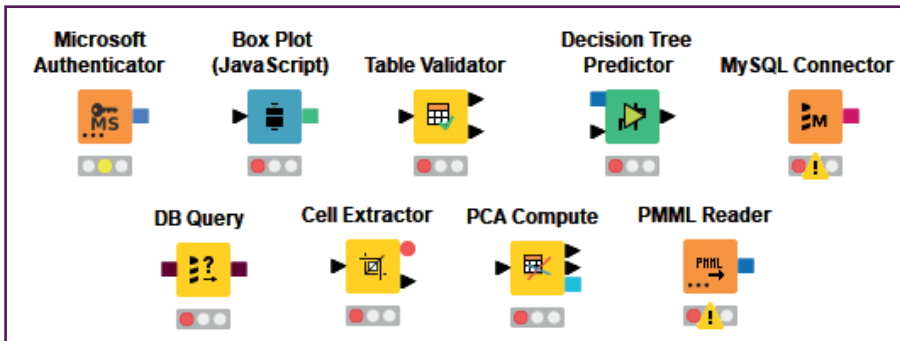


Fonte: Elaborado pela autora (2023).

Já a “anotação de um nó” é uma funcionalidade que permite aos usuários adicionarem informações textuais ou notas personalizadas a um nó específico em um fluxo de trabalho. Essas anotações são úteis para documentar e explicar o propósito ou a função de um nó, fornecendo detalhes adicionais que auxiliem na compreensão do fluxo de trabalho (KNIME, 2023c).

Existem diferentes tipos de portas de entrada e saída, normalmente, são representadas por diferentes símbolos (Figura 3). Os nós são diferenciados por cores, as quais refletem suas categorias. Como exemplo, todos os nós na tonalidade amarela destinam-se à manipulação de dados.

Figura 3 - Exemplo das diferentes portas de entrada e saída existentes



Fonte: Elaborado pela autora (2023).

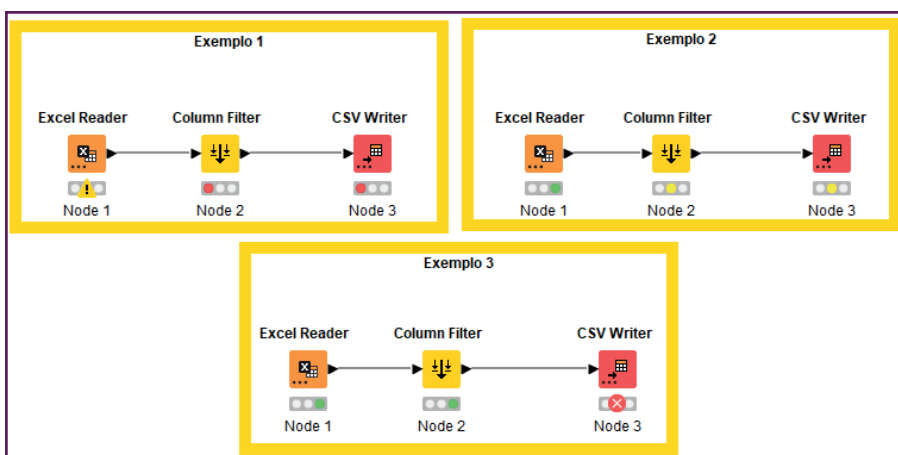
É importante ter em mente a compatibilidade entre essas portas ou, mais precisamente, entre os tipos de dados que elas manipulam. A regra fundamental é que uma porta de saída específica deve, obrigatoriamente, compartilhar o mesmo símbolo e cor com a porta de entrada correspondente para que sejam consideradas compatíveis.

A Figura 4 apresenta exemplos de três diferentes nós disponíveis no *KNIME*. Os nós "*Excel Reader*" e "*Column Filter*" possuem um triângulo preto à sua direita que representa as portas de saída. O nó "*Column Filter*" e o nó "*CSV Writer*" possuem um triângulo preto à sua esquerda que representa as portas de entrada de dados.

Um nó recebe um conjunto de dados como entrada, o processa e o torna disponível em sua porta de saída (Berthold *et al.*, 2009; Hayasaka; Silipo, 2023; KNIME, 2023c). Cada nó pode assumir quatro estados distintos, que seguem um esquema de cores semelhante a um semáforo de trânsito. Esses estados são descritos a seguir e ilustrados na Figura 4 (KNIME, 2023c):

- Inativo e não configurado -> Luz vermelha;
- Configurado, mas não executado -> Luz amarela;
- Executado com sucesso -> Luz verde;
- Executado com erros -> Luz vermelha com um "X".

Figura 4 - Exemplo dos diferentes estados dos nós



Fonte: Elaborado pela autora (2023).

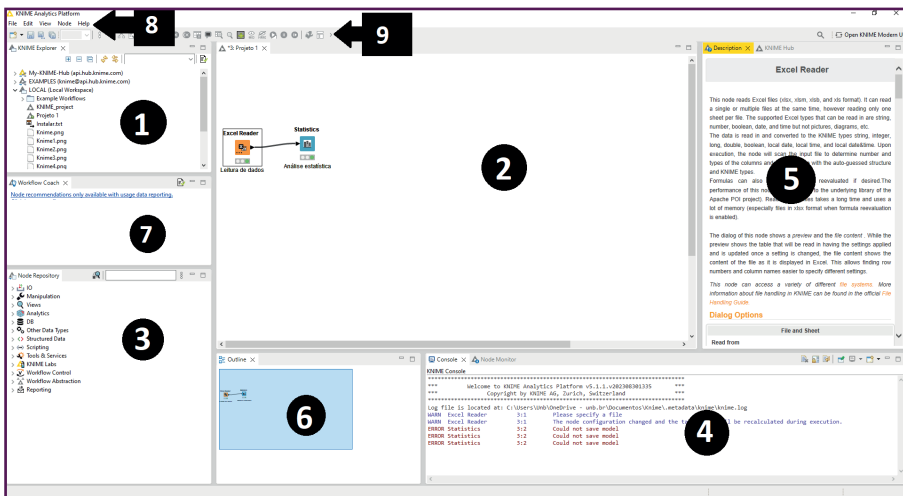
Na Figura 4, no exemplo 1, o primeiro nó "Excel Reader" apresenta um símbolo amarelo com um sinal de exclamação, que indica um alerta que o nó deve ser configurado, esse sinal sobrepõe o sinalizador do meio que está em amarelo. No Exemplo 1, ainda pode-se notar que os nós "Column Filter" e "CSV Writer" apresentam a *status* vermelha, indicando que estão inativos e não configurados. A configuração do primeiro nó nesse fluxo e sua respectiva execução alteram o *status* desse nó para executado com sucesso, conforme ilustrado no Exemplo 2. Automaticamente, pode-se ver que o segundo nó passa para o estado "configurado, mas não executado", pois esse nó não demanda nenhuma configuração especial. Já o terceiro nó deve ser configurado para que fique com a luz amarela. Por fim, no Exemplo 3, nota-se que os dois primeiros nós foram executados com sucesso e o terceiro nó foi executado com erro.

Na subseção seguinte, explora-se em detalhes a interface de trabalho do *KNIME*, examinando as características gerais da interface a fim de se obter uma compreensão completa da ferramenta.

10.2.3 INTERFACE DE TRABALHO

A ferramenta *KNIME* disponibiliza uma interface gráfica intuitiva (Figura 5) que confere à ferramenta acessibilidade, permitindo sua utilização por profissionais de diversas disciplinas, abrangendo não apenas programadores, mas também indivíduos de diferentes especializações (KNIME, 2023c).

Figura 5 - Interface gráfica do KNIME até a versão 4.7



Fonte: Elaborado pela autora (2023).

A interface gráfica ou área de trabalho do *KNIME* é composta pelos seguintes componentes, numerados de 1 a 9 na Figura 6 acima e explicados a seguir no Quadro 1.

Quadro 1 - Detalhes da interface gráfica da plataforma KNIME

N°	Nome	Descrição
1	Explorador (KNIME Explorer)	Uma visão geral dos fluxos de trabalho disponíveis e grupos de fluxos de trabalho existentes nas áreas de trabalho que você tem ativo, ou seja, no seu espaço de trabalho local e nos servidores KNIME. O espaço de trabalho é onde os fluxos de trabalho são criados e organizados.
2	Editor de Fluxo de Trabalho (Workflow Editor)	Uma área de edição visual para o fluxo de trabalho ou workflow atualmente ativo. É aqui que você cria, edita e configura os fluxos de trabalho. Você pode ter múltiplas abas de edição abertas, uma para cada workflow que está trabalhando.
3	Repositório de Nós (Node Repository)	Todos os nós disponíveis na Plataforma KNIME que são instalados junto com a instalação padrão e os nós das extensões que você instalou estão listados aqui. Os nós são organizados por categorias, mas você também pode usar a caixa de pesquisa na parte superior do repositório de nós para encontrar nós. Os usuários podem pesquisar e arrastar nós do navegador para a área de trabalho para adicioná-los ao fluxo de trabalho.
4	Console	É a área onde são exibidas as mensagens de execução que indicam o que está acontecendo nos bastidores durante a execução de um fluxo de trabalho. Isso pode incluir informações sobre o progresso da execução, erros encontrados, resultados produzidos e outras mensagens relevantes que auxiliam os usuários no monitoramento e na depuração de seus fluxos de trabalho.

Nº	Nome	Descrição
5	Descrição do Nó (Node Description)	Área onde é apresentada a descrição de um nó selecionado (no Editor de Fluxo de Trabalho ou no Repositório de Nós) ou do fluxo de trabalho em si (caso nenhum nó do fluxo seja selecionado).
6	Visão Geral (Outline)	Essa área apresenta uma visão geral do fluxo de trabalho atualmente ativo. Seu uso é interessante quando um workflow fica grande, com diversos nós, e não é facilmente visível na área de edição.
7	Assistente de Fluxo de Trabalho (Workflow Coach)	Essa área é destinada à lista recomendações de nós com base nos fluxos de trabalho criados pela comunidade de usuários. Ela fica inativa se você não permitir que o KNIME colete estatísticas de uso. Essa janela pode ser fechada e, quando desejado, ser reaberta.
8	Barra de Menu	Contém uma série de menus suspensos que fornecem acesso a uma ampla gama de comandos e funcionalidades.
9	Barra de Ferramentas	Contém botões para ações comuns, como salvar, abrir, executar e interromper fluxos de trabalho.

Fonte: Elaborado pela autora a partir de (KNIME, 2023c).

Com um entendimento sólido da interface, pode-se explorar como essa ferramenta permite construir fluxos de trabalho personalizados e executar análises de dados eficazes.

10.2.4 CONSTRUÇÃO DE UM FLUXO DE TRABALHO

O processo de construção de um fluxo de trabalho normalmente é iniciado por um nó de origem, que, em geral, tem a finalidade de coletar ou extrair dados provenientes de uma fonte de dados específica, por exemplo, um

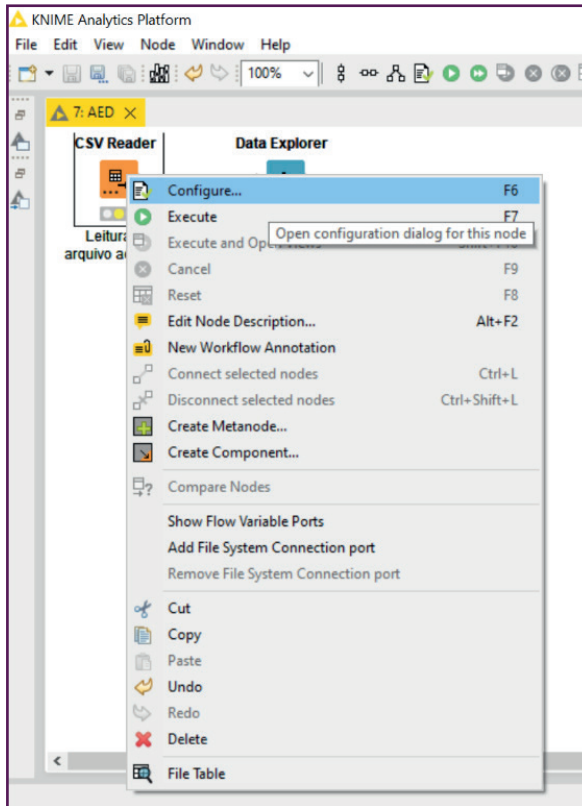
arquivo, uma *página web*, um banco de dados, dentre outras fontes. Os dados coletados são armazenados em um formato interno baseado em tabelas, no qual cada coluna possui um tipo de dado específico. Na sequência, essas tabelas de dados são encaminhadas a outros elementos por meio de conexões, possibilitando que esses elementos realizem uma série de operações, tais como modificações, transformações, modelagem ou geração de representações visuais fundamentadas nos dados previamente processados.

Um fluxo de trabalho da *plataforma KNIME* é criado simplesmente arrastando e soltando os nós existentes no repositório de nós para o editor de fluxo de trabalho ou clicando duas vezes sobre os nós. Dessa forma, à medida que os nós são adicionados sequencialmente, o *pipeline* é rapidamente construído. Cada nó deve ser configurado conforme necessário e, em seguida, o fluxo de trabalho pode ser executado para processar os dados.

10.2.4.1 CONFIGURAÇÃO DE UM NÓ

Até a versão 5.1 do *KNIME*, para configurar um nó basta clicar duas vezes sobre ele ou pressionar a tecla F6 com o nó selecionado (Figura 6).

Figura 6 - Acionando a configuração de um nó

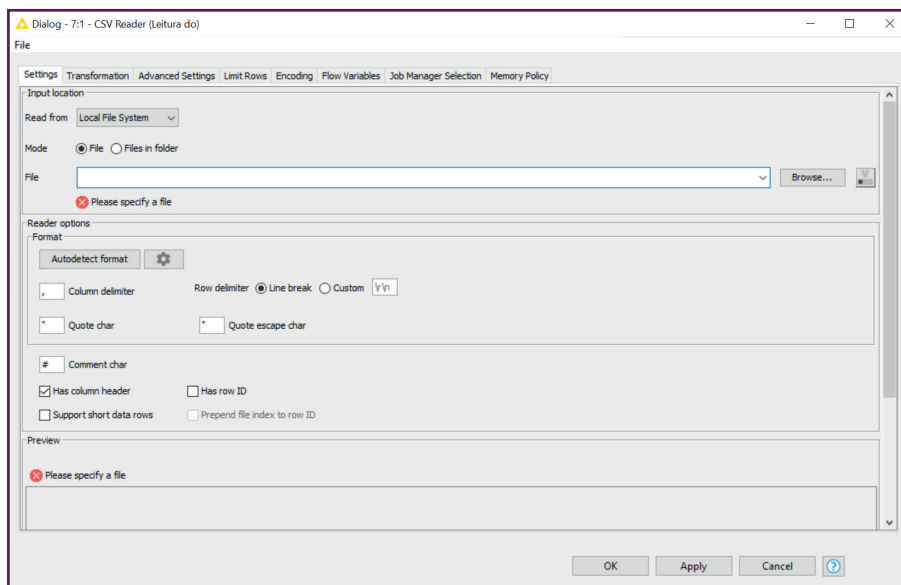


Fonte: Elaborado pela autora (2023).

Dentro da janela de configurações, tem-se acesso a um conjunto de opções e parâmetros que podem ser adaptados de acordo com as exigências do projeto. A diversidade dessas opções está vinculada ao tipo de nó em questão e às funcionalidades específicas que ele proporciona. Por exemplo, é possível definir as entradas de dados, configurar os parâmetros de processamento, especificar as ações a serem executadas e outras configurações pertinentes ao funcionamento do nó.

Para ilustrar a configuração de um nó, a Figura 7, a seguir, apresenta a janela de configurações do nó "CSV Reader". Esse nó é usado para ler dados de um arquivo CSV (Comma-Separated Values) (KNIME, 2022b).

Figura 7 - Janela de configuração do nó CSV Reader



Fonte: Elaborado pela autora (2023).

Existem várias opções de configuração para especificar como os dados do arquivo CSV devem ser lidos e interpretados. Conforme a Figura 7, alguns exemplos de configuração para esse nó são (KNIME, 2022b):

- No campo "File" informe o nome e caminho completo do arquivo a ser lido;
- Escolha o caractere ou sequência de caracteres que delimita os campos no arquivo CSV. O caractere padrão é a vírgula (,), mas pode ser configurado para outros caracteres, como ponto e vírgula (;) ou tabulação (t);
- Para adivinhar automaticamente a estrutura do arquivo, clique no botão "Autodetect format";
- Se a primeira linha do arquivo CSV contiver nomes de coluna, você pode marcar a opção "Has column header".

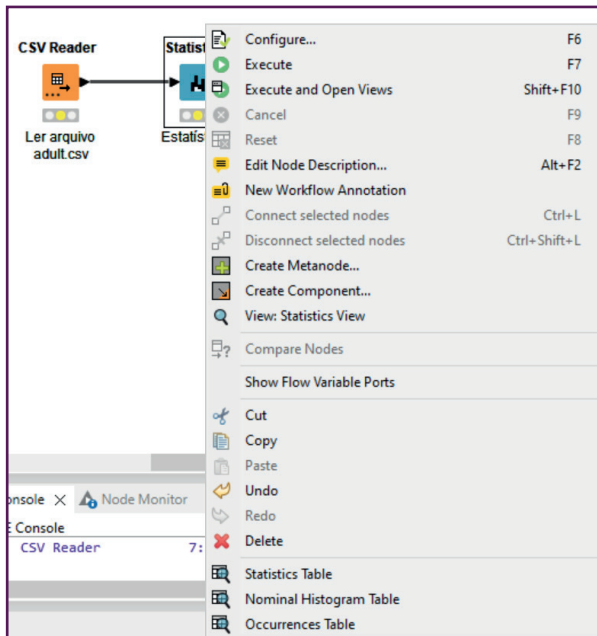
Informações detalhadas sobre cada nó podem ser encontradas na janela "Description". Ao consultar essa seção, os pesquisadores podem obter orientações detalhadas sobre como configurar cada nó. Dessa forma, para

cada nó, explore as opções de configuração disponíveis e faça os ajustes necessários para configurar o nó de acordo com os requisitos do seu fluxo de trabalho. Certifique-se de entender o impacto de cada configuração antes de aplicá-la.

10.2.4.2 EXECUÇÃO DO NÓ OU DO WORKFLOW

Após configurar os nós do fluxo de trabalho, o próximo passo essencial é executar o nó ou até mesmo o *workflow* completo. Para isso, basta clicar com o botão direito do mouse sobre o nó desejado e o mesmo menu exibido no ato da configuração será aberto. As opções “*Execute*” e “*Execute and Open Views*” estão ativas (Figura 8).

Figura 8 - Detalhe do menu para acesso a configuração e execução do nó



Fonte: Elaborado pela autora (2023).

Ao selecionar “*Execute*”, o nó será processado de acordo com as configurações previamente definidas. Durante sua execução será possível observar o progresso da execução. É importante observar qualquer mensagem de erro

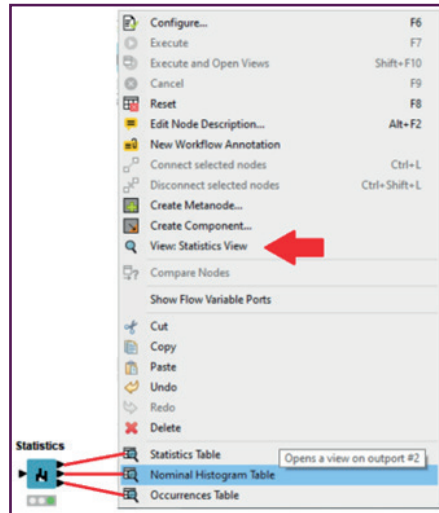
ou alerta, pois isso pode indicar problemas nos dados ou na configuração do nó. Ao utilizar a opção "*Execute and Open Views*", tão logo encerrada a execução, se os dados foram processados corretamente, será exibida uma janela com os resultados imediatamente após o processamento dos dados.

10.2.4.3 VERIFICAÇÃO DE RESULTADOS

Durante a construção e execução do *workflow* é importante proceder a uma inspeção dos resultados intermediários após a execução de cada nó. Essa prática ajuda a entender a tarefa encapsulada em cada nó e ajuda a assegurar a precisão e validade das operações realizadas. Essa verificação é realizada por meio da análise das portas de saída associadas a cada nó no fluxo de trabalho. Cada nó apresenta portas de saída que encapsulam os dados processados por ele.

Para visualizar os resultados, clicar sobre o nó com o botão direito do mouse, no menu que é exibido, assim é possível acessar o resultado de cada porta ao selecionar a respectiva opção localizada no inferior desse menu (Figura 9). Quando um nó possui múltiplas portas, a lista de resultados segue a mesma ordem que as portas, basta selecionar a porta de saída relevante e examinar os dados apresentados. O nó "*Statistics*" possui três portas de saída conforme indicado na Figura 9, e as três opções inferiores do menu equivalem aos resultados da execução desse nó, cada tabela é a saída de uma das portas. Adicionalmente, conforme destacado pela seta vermelha, em geral, nós que têm a opção de "*Execute and Open Views*" possuem a opção de visão interativa dos resultados, geralmente identificado pelo ícone de lupa.

Figura 9 - Exemplo da relação de postas e os respectivos resultados



Fonte: Elaborado pela autora (2023).

Dessa forma, é possível realizar uma validação dos resultados em relação às expectativas previamente estabelecidas. Saber interpretar e validar esses resultados é fundamental para garantir que as análises estejam corretas e confiáveis.

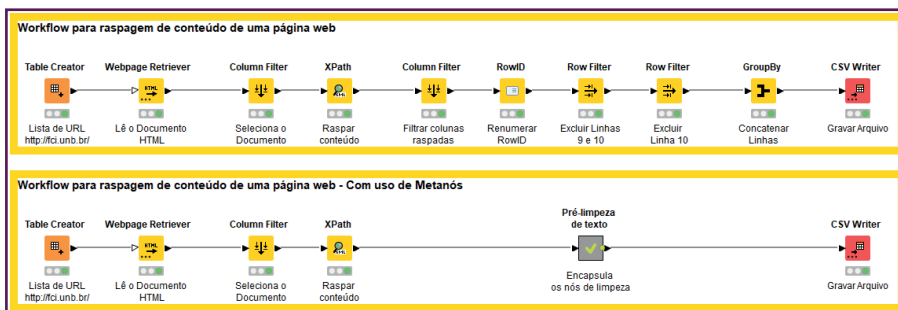
10.2.5 COMPONENTES E METANÓS

No *KNIME*, Componentes (em inglês *Components*) e Metanós (em inglês *Metanodes*) são recursos que permitem a organização, reutilização e modularização de partes de seus *workflows*, tornando o processo de construção e gerenciamento mais eficiente e organizado. Eles são particularmente úteis quando se deseja criar módulos de análise personalizados que podem ser aplicados em várias partes do seu processo de análise (KNIME, 2023f).

O exemplo da Figura 10 apresenta dois *workflows* que fazem a coleta de dados disponíveis em uma *página web*. No primeiro, existem cinco nós que são responsáveis por realizar uma pré-limpeza do conteúdo raspado, são os nós de "Column Filter" até o "GroupBy". Já no segundo *workflow*,

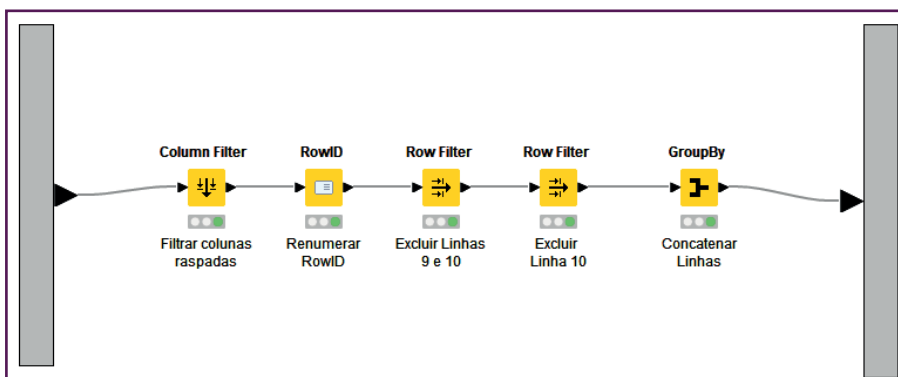
tem-se apenas o metanó identificado por “*Pré-limpeza de texto*”, encapsulando esses nós conforme pode ser visualizado na Figura 11.

Figura 10 - Exemplo de workflow com e sem metanó



Fonte: Elaborado pela autora (2023).

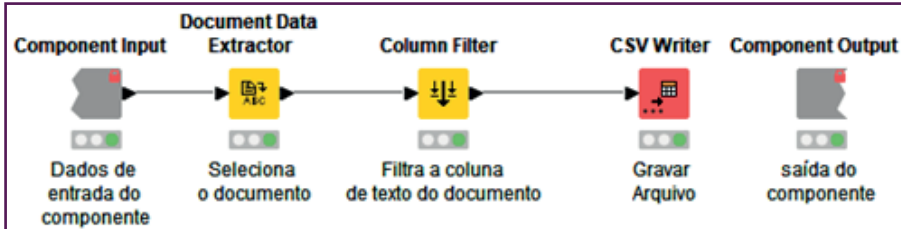
Figura 11 - Workflow encapsulado no metanó



Fonte: Elaborado pela autora (2023).

Na Figura 12 é apresentado um exemplo de *workflow* encapsulado em um componente. Nota-se a distinção em relação a Figura 11 pela existência de dois nós específicos: o “*Component Input*”, que representa a entrada de dados no componente, e o “*Component Output*”, que representa a saída dos dados processados pelo componente. Esses pontos de entrada e saída funcionam como pontos de integração entre o componente e os *workflows* que o utilizam, facilitando a comunicação eficaz de dados entre diferentes partes da solução.

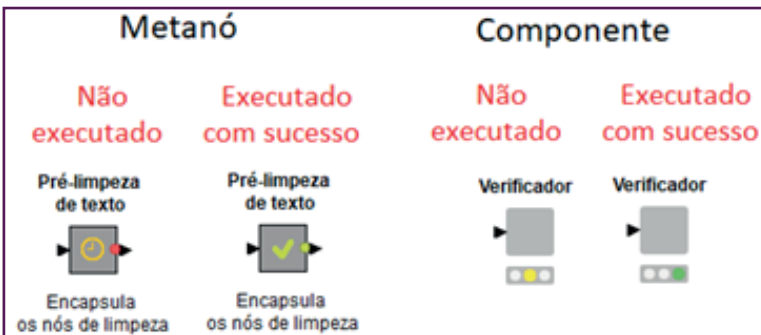
Figura 12 - Exemplo de um *workflow* encapsulado em um componente



Fonte: Elaborado pela autora (2023).

Os componentes e metanós encapsulam *workflows* ou partes específicas, com o objetivo de agrupar um conjunto de nós que, quando combinados, representam uma operação lógica específica. No entanto, existem diferenças significativas entre eles. Por exemplo, diferente dos metanós, os componentes têm a capacidade de possuir uma janela de configuração personalizada, com uma interface de diálogo dedicada e podem ainda apresentar saídas de visualizações interativas e complexas. Além disso, os componentes podem ser reutilizáveis no mesmo fluxo ou até mesmo em fluxos de trabalho separados. Eles podem ser compartilhados com outros usuários por meio do *KNIME Community Hub* (KNIME, 2022a, 2023e, 2023f). A Figura 13 ilustra a distinção visual entre metanós e componentes, ambos em sua representação gráfica antes da execução e após uma execução bem-sucedida.

Figura 13 - Exemplo de metanó e componente



Fonte: Elaborado pela autora (2023).

10.2.6 EXTENSÕES E INTEGRAÇÕES

As extensões e integrações são recursos adicionais que podem ser incorporados ao ambiente *KNIME* para estender suas funcionalidades e integrá-lo com outras tecnologias (KNIME, 2023a, 2023c, 2023e).

As extensões são pacotes de *plugins*¹⁰⁴ que adicionam funcionalidades específicas à plataforma. Esses *plugins* podem ser desenvolvidos pela comunidade ou pela equipe própria *KNIME AG*. Podem incluir nós (nodos) adicionais para processamento de dados, *algoritmos* de aprendizado de máquina, conectores para diferentes fontes de dados, visualizações personalizadas e outras funcionalidades (KNIME, 2023a). As extensões desenvolvidas e mantidas pela *KNIME AG* são de código aberto e disponíveis na página de extensões¹⁰⁵.

Já as integrações referem-se à capacidade da plataforma se integrar com outras ferramentas, tecnologias, linguagens de programação ou serviços. Por exemplo, o *KNIME* pode ser usado na preparação e processamento dados e, em seguida, enviar dados a uma ferramenta de visualização como o *Power BI*¹⁰⁶ para criar gráficos interativos e painéis informativos. As integrações também são desenvolvidas e mantidas pela *KNIME AG*, ou pelos fornecedores das ferramentas que elas integram. As integrações podem ser visualizadas e consultadas na página de integrações¹⁰⁷.

104 *Plugins* são componentes de software modulares e extensíveis que adicionam funcionalidades específicas a uma aplicação ou sistema maior. Eles são desenvolvidos separadamente e podem ser integrados facilmente, permitindo a extensão ou personalização da funcionalidade principal de um *software* sem a necessidade de alterar seu código-fonte fundamental. Em essência, *plugins* são peças de software que expandem as capacidades de um programa principal.

105 Disponível em: <https://www.knime.com/knime-extensions>. Acesso em: 20 ago. 2023.

106 Para mais detalhes consulte: <https://powerbi.microsoft.com/pt-br/>. Acesso em: 20 ago. 2023.

107 Disponível em: <https://www.knime.com/knime-integrations>. Acesso em: 20 ago. 2023.

10.3 INTEGRAÇÃO DO KNIME NA METODOLOGIA DE PESQUISA

A necessidade da adoção de TI para apoiar as várias etapas metodológicas do processo de pesquisa tem se tornado cada vez mais evidente. Na era atual da pesquisa científica, na qual grandes volumes de dados são gerados e necessitam ser analisados, ferramentas avançadas são essenciais para extrair respostas significativas e impulsionar ainda mais o desenvolvimento da ciência. Na primeira parte deste capítulo, foi apresentada a *Plataforma KNIME Analytics*, discutindo suas principais capacidades e funcionalidades. Nesta seção, será explorada como a plataforma pode ser habilmente incorporada na pesquisa científica, destacando suas capacidades.

10.3.1 ETAPAS GERAL DA PESQUISA CIENTÍFICA

A condução de uma pesquisa científica envolve várias etapas cruciais para garantir que os dados coletados sejam confiáveis, relevantes e úteis para responder às perguntas de pesquisa. A seguir foram sintetizadas quatro etapas envolvidas no processo de pesquisa: aquisição ou coleta de dados, preparação ou tratamento de dados, análise de dados e visualização de dados.

10.3.1.1 AQUISIÇÃO/COLETA DE DADOS

A etapa inicial da pesquisa científica engloba a aquisição ou coleta de dados, sendo uma etapa fundamental para reunir informações que servirão como alicerces para todo o processo de pesquisa. Independentemente dos métodos ou técnicas empregados, toda pesquisa demanda a obtenção de dados de uma variedade de fontes. Nessa etapa, os pesquisadores obtêm as informações essenciais necessárias para responder às questões de pesquisa formuladas e para testar as hipóteses estabelecidas. É importante que a coleta de dados seja precisa e bem planejada, uma vez que essa precisão é vital para assegurar a validade e a confiabilidade dos resultados alcançados durante a pesquisa (Gil, 2002, 2008; Lakatos; Marconi, 2010; Marconi; Lakatos, 2002).

A coleta de dados em pesquisas pode variar consideravelmente de acordo com o tipo de pesquisa e as circunstâncias envolvidas. Dentre os procedimentos adotados para essa coleta, destacam-se:

- **Entrevistas:** Entrevistas estruturadas ou semiestruturadas podem ser conduzidas para obter informações de participantes ou especialistas no campo de estudo. Após a transcrição das respostas, os dados são organizados em planilhas, arquivos ou bancos de dados, podem ser codificados e categorizados para identificar padrões e temas;
- **Questionários:** O uso de questionários padronizados é comum para coletar dados de grandes grupos de pessoas de maneira consistente. Esses questionários incluem perguntas fechadas, nas quais os participantes escolhem entre opções predefinidas e perguntas abertas, que permitem respostas em suas próprias palavras. As respostas obtidas são frequentemente registradas em planilhas e bancos de dados para análise posterior;
- **Observação:** Os pesquisadores podem observar eventos, comportamentos ou fenômenos diretamente, registrando suas observações de forma sistemática. Após a coleta, os dados são organizados de forma sistemática descrevendo detalhes do contexto, comportamentos ou eventos relevantes. Os pesquisadores, muitas vezes, utilizam códigos ou símbolos para categorizar as observações. Esses dados podem ser registrados em formato textual, em bancos de dados e planilhas.

É importante ressaltar que as coletas de dados podem ser realizadas em diversas fontes de informação acadêmica, tais como artigos científicos, teses e dissertações. Essas fontes desempenham um papel crucial ao fornecer fundamentação teórica e prática para a pesquisa científica. Além disso, é comum conduzir pesquisas como estudos métricos da informação e revisões sistemáticas de literatura, visando essencialmente examinar a produção científica existente nessas fontes de informação. Tais fontes são categorizadas em três tipos distintos: primárias, secundárias e terciárias, cada uma com suas características específicas. A escolha pelo tipo de fonte é determinada pelos objetivos da pesquisa.

As fontes primárias são aquelas que apresentam informações originais e inéditas, como artigos científicos, teses e dissertações. As fontes secundárias são aquelas que compilam informações de diversas fontes primárias, como bases de dados, revisões sistemáticas e meta-análises. Já as fontes terciárias são aquelas que fornecem informações resumidas e

de fácil acesso, como *catálogos on-line*, manuais, guias, enciclopédias e dicionários (Souza *et al.*, 2022).

Em estudos métricos da informação e revisões sistemáticas de literatura, a coleta e aquisição de dados envolvem métodos específicos para analisar a produção científica existente em uma área de estudo.

10.3.1.2 PREPARAÇÃO E/OU TRATAMENTO DE DADOS

Após a coleta dos dados, os dados devem ser organizados e tratados para posterior análise e interpretação. A preparação ou elaboração dos dados devem seguir os seguintes passos: seleção, codificação, tabulação (Lakatos; Marconi, 2010; Marconi; Lakatos, 1990).

Durante a seleção, são escolhidos os dados pertinentes para a pesquisa, enquanto na codificação, eles são transformados em categorias ou códigos, facilitando sua interpretação quantitativa. Posteriormente, na tabulação, os dados são organizados de forma estruturada em tabelas, preparando-os para análises detalhadas e interpretações significativas. Além disso, a tabulação facilita a representação gráfica dos dados, possibilitando uma compreensão e interpretação mais rápidas.

Em pesquisas que envolvem estudos métricos da informação e revisões sistemáticas de literatura, a etapa de elaboração ou preparação dos dados é fundamental para garantir a confiabilidade e validade das análises. Nessa fase, os pesquisadores realizam uma seleção criteriosa de estudos relevantes e os organizam. Para essa organização, os pesquisadores precisam extrair sistematicamente as informações pertinentes de cada estudo. Isso pode incluir a criação de uma planilha ou banco de dados eletrônico contendo detalhes de cada estudo, como título, autores, instituições afiliadas, palavras-chave, local e ano de publicação, métodos, resultados e conclusões. Durante o processo de preparação dos dados, os pesquisadores também podem realizar análises da qualidade metodológica dos estudos incluídos, corrigindo dados conforme necessário, garantindo, assim, a integridade dos dados utilizados na pesquisa métrica ou revisão sistemática de literatura (Biolchini *et al.*, 2005; Galvão; Ricarte, 2019; Grácio *et al.*, 2020; Kitchenham, 2004).

De acordo com a natureza da pesquisa em questão e a tipologia dos dados coletados, a fase de preparação de dados pode implicar na execução de diversas atividades, tais como:

- **Limpeza de dados:** Identificar, corrigir e remover erros ou valores inconsistentes nos conjuntos de dados para garantir que sejam precisos, confiáveis e adequados para análise estatística. Por exemplo, podem ser tratados erros de erros de digitação e valores extremos (outliers);
- **Tratamento de Valores Ausentes:** Caso haja dados ausentes, os pesquisadores podem decidir preencher esses valores usando técnicas como a imputação de dados (substituindo os valores ausentes por estimativas baseadas em dados existentes) ou remover as entradas com dados ausentes, dependendo da natureza do estudo;
- **Transformação de dados:** Os dados podem ser reformatados, agregados ou transformados para torná-los adequados para análise. Por exemplo, em alguns casos, variáveis contínuas podem ser discretizadas, transformando-as em variáveis categóricas ou ordinais. Ou ainda a inclusão de novas variáveis derivadas dos dados originais para capturar características específicas ou padrões que não são diretamente observáveis nos dados brutos;
- **Padronização:** Garantir que os dados estejam em um formato uniforme e compatível para análise. Por exemplo, as datas podem ser formatadas de maneira consistente (dd/mm/aaaa ou mm/dd/aaaa) e as categorias podem ser padronizadas para evitar redundâncias ou sobreposições;
- **Codificação:** Atribuir códigos ou categorias a dados qualitativos para facilitar sua análise e interpretação. Essa tarefa é especialmente relevante quando se lida com respostas abertas de questionários, entrevistas ou outras formas de dados qualitativos. Por exemplo, a codificação intervalar é uma técnica de codificação de dados em pesquisas qualitativas que envolve atribuir códigos a segmentos específicos de dados que estão dentro de intervalos pré-definidos.

10.3.1.3 ANÁLISE DE DADOS

A etapa de análise de dados na pesquisa científica é um processo complexo que envolve a aplicação de métodos estatísticos, análises qualitativas e técnicas específicas para extrair *insights* significativos dos dados coletados. Na pesquisa científica, a análise e interpretação dos dados representam o núcleo central do processo. A importância dos dados reside não apenas em si mesmos, mas na capacidade de fornecer respostas às investigações propostas (Gil, 2002, 2008; Lakatos; Marconi, 2010; Marconi; Lakatos, 2002).

É crucial definir a técnica de análise apropriada para o desenvolvimento do seu trabalho de acordo com o tipo de pesquisa que está sendo conduzida, seja ela quantitativa ou qualitativa. A pesquisa quantitativa lida principalmente com dados numéricos, assim, as técnicas de análise quantitativa envolvem o uso de métodos estatísticos para analisar padrões e relações nos dados. Já a pesquisa qualitativa, em geral, trabalha com dados categóricos (não numéricos) e emprega técnicas de análise que envolvem interpretar significados, padrões e contextos dos dados (Prodanov; Freitas, 2013).

Nessa fase, os pesquisadores exploram uma diversidade de técnicas tanto qualitativas quanto quantitativas para extrair conhecimento substancial dos dados coletados. Lembre-se de que em alguns estudos, uma abordagem mista, combinando elementos quantitativos e qualitativos, pode ser adotada para obter uma compreensão mais abrangente do fenômeno em estudo.

Em pesquisa de caráter quantitativo, os dados são tratados numericamente para discernir padrões e relações significativas. Nesse sentido é comum a adoção de técnicas estatísticas como: i) análise descritiva ou Análise Exploratória de Dados (AED) para resumir características fundamentais dos dados, como medidas de posição e medidas de dispersão; ii) análise de regressão para examinar relações entre variáveis; iii) e análise de séries temporais para identificar tendências ao longo do tempo.

Já em pesquisas de caráter qualitativo o foco está na compreensão aprofundada e na interpretação dos dados. Algumas das técnicas empregadas são: i) análise de conteúdo que busca significados subjacentes em dados textuais; ii) a modelagem de tópicos que revela temas latentes em grandes conjuntos de documentos ou textos em geral; iii) a análise de rede social

que e examina a estrutura das relações sociais entre indivíduos, grupos ou organizações; iv) análise de sentimentos que é usada para determinar a tonalidade emocional do texto, seja positiva, negativa ou neutra e; v) a análise de discurso que examina o uso da linguagem para entender como as pessoas constroem significado e representam o mundo ao seu redor.

Estudos métricos da informação e revisões sistemáticas de literatura são tipos de pesquisas especializadas que demandam uma análise aprofundada da produção científica em uma área ou assunto de estudo.

A análise de dados em estudos métricos da informação lida com indicadores de produção, de ligação e de citação como variáveis quantitativas e/ou qualitativas. Esses indicadores contribuem para entender não apenas a produção científica, mas também as relações entre pesquisadores, instituições, países e áreas de conhecimento (Prado; Castanha, 2020). Os indicadores são analisados por meio de várias técnicas e métodos estatísticos.

Em pesquisas dedicadas à revisão sistemática de literatura, os pesquisadores têm como objetivo gerar novos conhecimentos a partir dos estudos selecionados para revisão. Esse processo implica em uma leitura sintópica, na qual relações entre os textos são estabelecidas e os dados são transformados em resumos significativos, buscando responder à questão que motivou a revisão (Brizola; Fantin, 2016). Algumas técnicas de análise qualitativa como a modelagem de tópicos e a análise de conteúdo podem ser empregadas para analisar esses estudos. A mineração de texto¹⁰⁸ também pode ser uma técnica automatizada para contribuir na geração dos resultados esperados.

10.3.1.4 VISUALIZAÇÃO DE DADOS

A etapa de visualização de dados em uma pesquisa científica refere-se ao processo de representar informações complexas e volumosas por meio

108 Mineração de textos é uma área de processo de descoberta de conhecimento que fornece técnicas efetivas de descoberta de conhecimento em bases de dados não estruturados, como textos. A mineração de textos é responsável por identificar informações úteis e implícitas nos textos, permitindo a extração de conhecimento e *insights* valiosos (Morais; Ambrósio, 2007).

de gráficos, tabelas, mapas ou outras formas visuais. É uma etapa que interliga intimamente com a análise de dados e o processo de apresentação no relatório de pesquisa.

Uma representação ideal dos resultados da pesquisa envolve a descrição dos dados, geralmente realizada por artefatos gráficos como tabelas, quadros e gráficos, acompanhados por textos explicativos. A adoção de técnicas de visuais de apresentação de resultados proporciona ao leitor uma compreensão e interpretação rápidas da grande quantidade de informações (Gil, 2008; Lakatos; Marconi, 2010).

Gráficos, quadros, tabelas, mapas conceituais, mapas de calor, *Bag of Words* (BoW), diagramas e nuvens de palavras são algumas das técnicas frequentemente usadas para transformar dados brutos em informações significativas e facilmente interpretáveis. Essas técnicas não apenas facilitam a compreensão dos resultados, mas também fornecem evidências visuais que sustentam as conclusões e os argumentos apresentados no relatório.

10.3.1.5 EXECUÇÃO DA PESQUISA CIENTÍFICA EM SÍNTESE

A execução da pesquisa científica é um processo complexo que segue várias etapas essenciais para assegurar a confiabilidade e relevância dos resultados da pesquisa. A pesquisa é executada por meio de quatro etapas cruciais no processo de pesquisa: aquisição de dados, preparação ou tratamento de dados, análise de dados e visualização de dados.

Começando com a coleta precisa de informações de fontes variadas, seguindo para a preparação e tratamento dos dados, envolvendo desde a limpeza até a padronização, e sua subsequente análise quantitativa ou qualitativa. Finalmente, a visualização dos dados, por meio de gráficos, tabelas e outras formas visuais, desempenha um papel crucial na apresentação clara e envolvente dos resultados. Essas etapas, combinadas, transformam dados brutos em conhecimento sólido, contribuindo significativamente para o avanço do conhecimento científico em diversas áreas de estudo.

Todas essas etapas do processo de pesquisa científica podem ser aprimoradas e otimizadas por meio do uso de ferramentas de TI. Com o auxílio de *softwares* especializados, pesquisadores podem coletar dados de forma

mais eficiente, organizar e preparar informações de maneira automatizada, realizar análises estatísticas complexas com precisão e criar visualizações de dados interativas e informativas. Além disso, o uso de ferramentas de TI permite a gestão eficaz de grandes volumes de dados, facilitando a identificação de padrões e tendências significativas. Essas tecnologias desempenham um papel fundamental na aceleração do processo de pesquisa, proporcionando aos cientistas recursos valiosos para conduzir estudos mais detalhados e avançados em diversas áreas do conhecimento.

10.3.2 USO DO *KNIME* EM CADA ETAPAS DA METODOLOGIA DE PESQUISA

A integração de ferramentas de TI tornou-se indispensável para pesquisadores em diversas áreas, especialmente em CSA. O avanço tecnológico revolucionou a maneira como os dados são coletados, analisados e interpretados, proporcionando aos pesquisadores uma gama de recursos poderosos para explorar fenômenos complexos e extrair *insights* valiosos. Nas CSA, em que a compreensão profunda de comportamentos humanos, padrões sociais e dinâmicas organizacionais é fundamental, as ferramentas de TI oferecem uma vantagem significativa.

Nesse contexto, tanto a *plataforma KNIME* quanto ferramentas similares, como o *OpenRefine* e o *Orange*, destacam-se como soluções multifuncionais essenciais para pesquisadores. Nesta seção, explora-se como o *KNIME* se encaixa de forma precisa em cada etapa da metodologia de pesquisa científica. Para ilustrar sua aplicação prática e impacto no campo, apresenta-se exemplos de *workflows* que podem ser usados em casos reais de estudos em CSA.

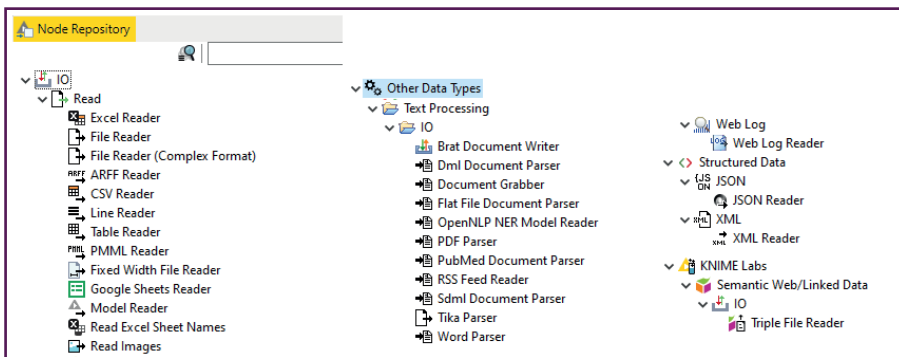
10.3.2.1 COLETANDO DADOS COM O *KNIME*

O *KNIME* se destaca por sua capacidade de coletar dados de diversas fontes, adaptando-se a diferentes formatos e origens. É possível coletar dados de arquivos localizados na mesma máquina em que a ferramenta está instalada, como arquivos disponíveis remotamente em servidores ou na *web*, por exemplo, arquivos disponíveis no Portal Brasileiro de Dados Abertos. Para arquivos, os pesquisadores podem utilizar nós específicos

para formatos como *CSV*, *Excel*, *XML*, *PDF* e *JSON*, garantindo a integração perfeita de dados de documentos variados.

A Figura 14 apresenta uma lista de nós que podem ser usados para realizar a coleta de dados em arquivos de diferentes formatos. A maior parte dos nós de entrada e saída (*IO* do inglês *In Out*) já são instalados em conjunto com a instalação básica. Entretanto, alguns nós só estarão disponíveis após a instalação de suas respectivas extensões ou integrações. Por exemplo, os nós "*PDF Parser*" e "*Tika Parser*" que fazem a leitura de arquivos em formato *PDF* necessitam da instalação da extensão "*KNIME Textprocessing*"; o nó "*XML Reader*" para ler arquivos em formato *XML* e o nó "*JSON Reader*" só estarão disponíveis após a instalação das extensões "*KNIME XML-Processing*" e "*KNIME JSON-Processing*" respectivamente.

Figura 14 - Exemplos de nós para coleta de dados em arquivos



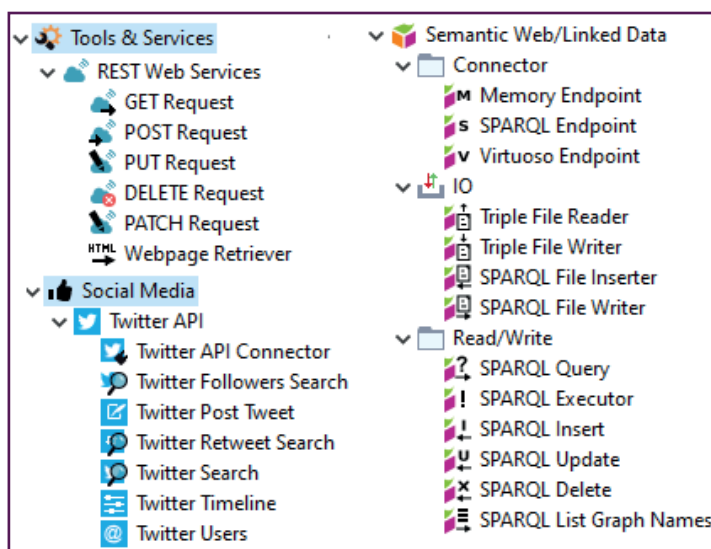
Fonte: Elaborado pela autora (2023).

Em relação a bancos de dados, a extensão "*KNIME Database*" proporciona uma gama de nós para bancos de dados relacionais, como *MySQL*, *PostgreSQL* e *SQLite*, permitindo a extração eficaz de dados estruturados. Essa extensão permite conectar-se a bancos de dados compatíveis com JDBC (Java Database Connectivity). Não é necessário instalar nenhuma extensão adicional, pois essa extensão já é instalada em conjunto com a instalação básica. Esses nós estão localizados na categoria "*DB*" no repositório de nós, na qual você pode encontrar vários nós para acesso, manipulação e escrita de banco de dados (KNIME, 2023g).

Para lidar com bancos de dados não estruturados, existem extensões e integrações específicas. Por exemplo, a extensão “*KNIME Big Data Extensions*” integra o *Apache Spark* e o ecossistema *Apache Hadoop*. Ela oferece aos usuários a capacidade de acessar uma variedade de bancos de dados, incluindo *Amazon Redshift*, *H2*, *Hive*, *Impala*, entre outros (KNIME, 2023d). Essa extensão pode ser instalada em sua totalidade ou apenas em partes conforme necessidade do pesquisador.

O *KNIME* permite ainda que os pesquisadores raspem dados de *páginas web* por meio de nós específicos para raspagem da *web* (web scraping) ou de redes sociais como *Twitter*. No geral, esses nós são fornecidos por extensões e integrações específicas que não fazem parte da instalação básica. A Figura 15 apresenta uma síntese de nós que podem ser usados para coletar dados disponíveis na *web*.

Figura 15 - Exemplos de nós para coleta de dados na web e redes sociais



Fonte: Elaborado pela autora (2023).

Na coluna da esquerda da Figura 15, são visualizados nós das extensões “*KNIME REST Client Extension*” e “*KNIME Twitter Connectors*”. A primeira, oferece uma variedade de nós que possibilitam aos pesquisadores interagir de forma direta com serviços *web* que seguem o padrão *REST*. Esses nós

permitem realizar solicitações *HTTP*, enviar parâmetros, receber dados de resposta e manipular informações provenientes de *APIs RESTful*. Por meio do nó "*Webpage Retriever*" é possível recuperar *páginas da web*, emitindo solicitações *HTTP GET* e analisando as *páginas da web HTML* solicitadas. Por padrão, a tabela de saída conterá uma coluna com o *HTML* analisado convertido em *XHTML*. Ao conectar essa saída ao nó *Xpath* do *KNIME*, o pesquisador pode configurar os dados que deseja extrair do documento *XHTML*. Isso é útil quando você precisa acessar dados específicos dentro de documentos *XML* ou *XHTML*. Um exemplo de *workflows* com esses nós foi apresentado na Figura 10.

Já a extensão "*KNIME Twitter Connectors*" oferece aos pesquisadores uma maneira direta de interagir com a *API* (Application Programming Interface) do *Twitter* em seus projetos de pesquisa. Ao utilizar essa extensão, os pesquisadores podem extrair dados do *Twitter*, como *tweets*, informações de perfil de usuário, tendências e muito mais. Os nós disponíveis nessa extensão permitem pesquisas específicas e coleta de dados em tempo real. Nota-se que essa extensão está sujeita às limitações das políticas da plataforma *Twitter*, as quais devem ser observadas pelos usuários. Essas limitações podem incluir restrições no volume de dados que pode ser acessado, bem como nos tipos de dados que podem ser coletados, conforme as diretrizes e políticas de uso do *Twitter*¹⁰⁹. Tal extensão é bem útil para pesquisas que focam em análise de sentimentos ou mineração da opinião na rede social *Twitter*.

Na coluna da direita da Figura 15, é listado um conjunto de nós específicos para trabalhar *web semântica* e dados interligados (linked data) da extensão "*KNIME Semantic Web*". Com ela, os pesquisadores terão a capacidade de acessar e manipular recursos da *web semântica*, incluindo *endpoints SPARQL*. Assim, os pesquisadores podem explorar dados e informações em ontologias, triplas *RDF*¹¹⁰ (Resource Description Framework) e outros recursos semânticos na *web*. Isso amplia significativamente as

109 Detalhes no site oficial do *Twitter*: <https://developer.twitter.com/en>. Acesso em: 25 set. 2023.

110 Triplas *RDF*: são estruturas de dados fundamentais compostas por três partes: o sujeito (representando o recurso), o predicado (descrevendo a relação) e o objeto (indicando o valor ou recurso), usadas na *web semântica* para a integração e troca de informações semânticas.

possibilidades de pesquisa, permitindo uma análise mais profunda e detalhada de dados relacionados à *web semântica*.

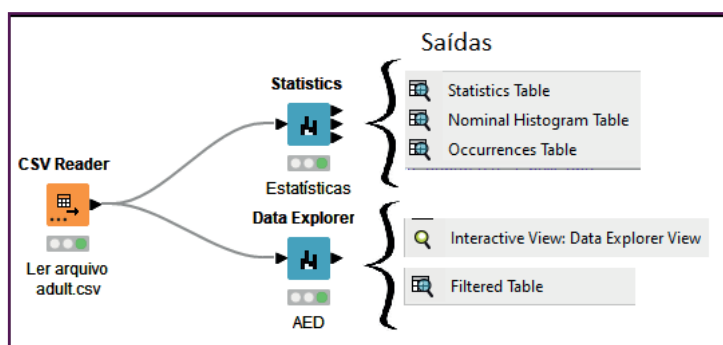
Existem outras extensões e integrações que possuem nós destinados a coletar dados em diferentes fontes de dados, veja a lista completa de extensões¹¹¹ disponíveis no *KNIME Community Hub*.

10.3.2.2 PREPARAÇÃO DE DADOS COM KNIME

A preparação de dados é uma etapa fundamental em qualquer projeto de pesquisa científica, pois dados bem-preparados são essenciais para análises precisas e resultados confiáveis. A *plataforma KNIME* oferece diversos nós e extensões que atendem as técnicas para limpar, transformar e organizar dados de acordo com as necessidades dos pesquisadores. Para cada tarefa de preparação pode-se visualizar diversas funcionalidades conforme apresentadas a seguir.

Na **identificação de erros**, é possível realizar uma análise descritiva ou exploratória dos dados por meio da utilização dos nós “*Statistics*” e “*Data Explorer*”, conforme exemplo de *workflow* da Figura 16.

Figura 16 - Exemplo de workflow para identificação de erros e suas saídas



Fonte: Elaborado pela autora (2023).

111 Veja as extensões disponíveis no *Knime Community Hub* em: <https://hub.knime.com/knime/extensions>.

Esses nós permitem calcular estatísticas descritivas em colunas numéricas, incluindo valores como contagem, mínimo, máximo, média, mediana, desvio padrão e variância, e medidas de distribuição como assimetria (Skewness) e Curtose (Kurtosis). Ambos os nós ajudam a identificar valores ausentes em colunas nominais e numéricas, fornecendo informações sobre o número de valores ausentes (missing values) em cada variável e ainda para valores numéricos o número de ocorrências do valor 0 (zero). Esses nós ainda calculam o número de valores únicos e a contagem de cada valor nominal presente nas variáveis categóricas. Ele oferece uma visão quantitativa das variáveis nominais, indicando quantas vezes cada categoria aparece no conjunto de dados por meio de gráficos do tipo histograma. O nó "*Statistics*" ainda gera uma tabela de saída de ocorrências indicando para cada valor de cada variável quantas vezes ele ocorre, proporcionando uma compreensão abrangente da distribuição dos dados e ajudando na identificação de possíveis valores discrepantes ou *outliers*. O nó "*Data Explorer*" permite que os usuários interajam com os gráficos. Por exemplo, os usuários podem clicar em uma categoria no gráfico de barras para filtrar os dados e observar como outras variáveis numéricas se comportam em relação a essa categoria específica.

Na tarefa de **Correção de Erros**, diversas estratégias podem ser empregadas de acordo com as necessidades específicas do processo de preparação de dados. No repositório de nós da pasta "*Manipulation*", encontram-se diversas ferramentas designadas para a manipulação eficaz dos dados. Destacam-se alguns nós particularmente úteis nesse contexto, tais como "*Row Filter*", "*Missing Value*", "*Rule Engine*", "*String Replacer*", "*Math Formula*", "*String Manipulation*", "*Duplicate Row Filter*", entre outros que se encontram disponíveis.

Esses nós fornecem uma variedade de funcionalidades, permitindo desde a filtragem de linhas com base em critérios específicos ("*Row Filter*" e "*Duplicate Row Filter*"), a substituição de padrões de texto ("*String Replacer*" e "*String Manipulation*"). Por exemplo, o nó "*Row Filter*" pode ser empregado para excluir as linhas em que o valor de uma variável específica está ausente ou possui um valor específico.

O nó "*Missing Value*" é um nó mais avançado para lidar com valores ausentes encontrados nas células da tabela de entrada. Ele oferece opções de tratamento padrão para todas as colunas de um tipo específico ou opções

para tratamento individualizado de cada coluna. Por exemplo, imagine um conjunto de dados com uma coluna de idade e algumas entradas estão em branco. É possível preencher os valores ausentes com a média arredondada para valor inteiro das idades existentes.

O nó "*Math Formula*" avalia uma expressão matemática com base nos valores em uma linha. Os resultados computados podem ser adicionados como uma nova coluna ou utilizados para substituir uma coluna de entrada. Por exemplo, em um conjunto de dados com colunas "Preço" e "Quantidade", pode ser criada uma nova coluna chamada "Total" com a fórmula: Preço x Quantidade. Ou ainda, em um conjunto de dados que tenha uma coluna chamada "Data de Nascimento" e deseje calcular a idade das pessoas com base nessa informação.

O nó "*Rule Engine*" permite que o usuário forneça uma lista de regras personalizadas e as aplique a cada linha na tabela de entrada. Quando uma regra é correspondida, o valor do seu resultado é adicionado em uma nova coluna ou substitui o valor resultante na própria coluna da tabela. Por exemplo, em um conjunto de dados de alunos, pode-se criar uma regra para atribuir um status "Aprovado" aos alunos com notas superiores a 60 e um status "Reprovado" aos alunos com notas inferiores ou iguais a 60.

O *KNIME* também oferece uma variedade de nós especializados para facilitar a tarefa de **integração de dados**, por exemplo os nós "*Joiner*", "*Concatenate*", "*GroupBy*" e "*Column Combiner*". O nó "*Joiner*" é instrumental na combinação de dados de duas ou mais tabelas com base em chaves específicas. Os dados resultantes são agregados de acordo com as chaves de junção, proporcionando uma visão dos dados combinados. O nó "*Concatenate*" mescla dados verticalmente, empilhando linhas de várias tabelas ou conjuntos de dados. Isso é valioso quando é necessário combinar dados de fontes semelhantes, pois mantém as colunas correspondentes das tabelas de entrada, criando um conjunto de dados contínuo.

O nó "*Column Combiner*" é utilizado para combinar duas ou mais colunas em uma única coluna. Ele é útil quando há necessidade de criar uma nova variável que seja uma combinação ou concatenação de informações de diferentes colunas. Por exemplo, é possível combinar o nome e o sobrenome de uma pessoa em uma única coluna de nome completo. O nó "*GroupBy*" é fundamental para análises agregadas. Ele agrupa os dados

com base em uma ou mais colunas de referência e permite calcular estatísticas ou realizar operações em grupos específicos de dados. Por exemplo, pode-se agrupar dados por região geográfica e calcular a média de vendas em cada região.

Na tarefa de **padronização e normalização**, o *KNIME* oferece um conjunto de opções, além dos nós já apresentados para correção de erros que também podem desempenhar o papel de padronização de dados. Nós como o "*Number To String*" e o "*Category To Number*" realizam a transformação de dados, permitindo a conversão entre tipos de dados diferentes. O nó "*Normalizer*" pode ser utilizado para ajustar os dados a uma escala específica, como a *escala Z* ou a *escala mín-máx*, garantindo que diferentes variáveis estejam na mesma escala para análises justas e equitativas. Em contrapartida, o nó "*Denormalizer*" pode reverter as transformações aplicadas aos dados normalizados.

Esses são apenas alguns exemplos dos nós disponíveis que podem favorecer a pesquisa científica apoiando a tarefa de preparação e elaboração de dados. Seja para lidar com valores ausentes, aplicar transformações complexas, filtrar dados específicos ou combinar diferentes colunas, existe um nó específico para cada cenário. Para encontrar os nós que se adequem às características de suas pesquisas, sugere-se explorar o *KNIME Hub* e o repositório de nós. Dessa forma, pesquisadores podem escolher e utilizar os nós mais adequados para atender às suas necessidades específicas de preparação de dados antes de iniciar suas análises, garantindo, assim, a integridade e a precisão dos dados durante todo o processo de pesquisa.

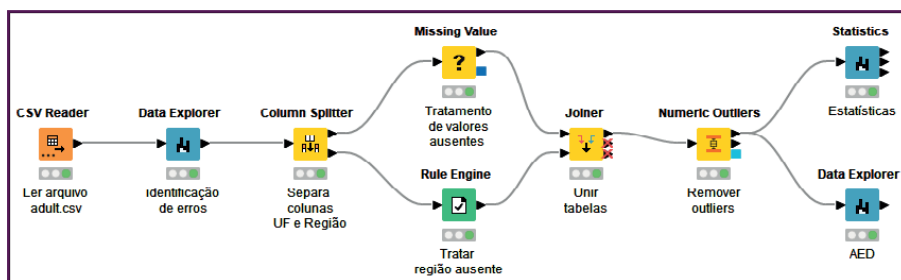
10.3.2.3 ANÁLISE DE DADOS COM *KNIME*

Na tarefa de análise de dados, o *KNIME* oferece funcionalidades que cobrem uma variedade de abordagens, desde métodos qualitativos que exploram nuances subjetivas até métodos quantitativos que se baseiam em dados numéricos e estatísticas precisas.

No contexto da AED, pode-se trabalhar com um *workflow* simples como já foi apresentado anteriormente e ilustrado na Figura 16. Os nós "*Statistics*" e "*Data Explorer*" não só podem ser usados para compreender os dados, mas também podem ser usados para tratar uma abordagem quantitativa

na pesquisa científica. A Figura 17 ilustra um exemplo no qual o nó “*Data Explorer*” primeiro foi usado para identificação de erros e necessidade de tratamento nos dados e, ao final, após a preparação de dados, ele é usado novamente e paralelo aos nós “*Statistics*” com o intuito de gerar uma análise estatística dos dados.

Figura 17 - Exemplo de workflow de AED



Fonte: Elaborado pela autora (2023).

O nó “*Data Explorer*” gera como saída um sumário com os resultados separados para as variáveis numéricas e as variáveis nominais ou categóricas. A saída numérica (Figura 18) mostra as principais propriedades estatísticas dos dados numéricos, como mínimo, máximo, mediana (opcional), desvio padrão, variância, assimetria, curtose, soma total, número de valores zero, número de valores ausentes.

Figura 18 - Exemplo de resultado do nó “*Data Explorer*” para variáveis numéricas

Column	Exclude Column	Minimum	Maximum	Mean	Median	Standard Deviation	Variance	Skewness	Kurtosis	Overall Sum	No. zeros	No. missings
idade	<input type="checkbox"/>	8	75	37.876	37	12.972	168.262	0.457	-0.500	362434	0	0
anos_estudo	<input type="checkbox"/>	5	16	10.327	10	2.223	4.941	0.272	-0.174	98816	0	0

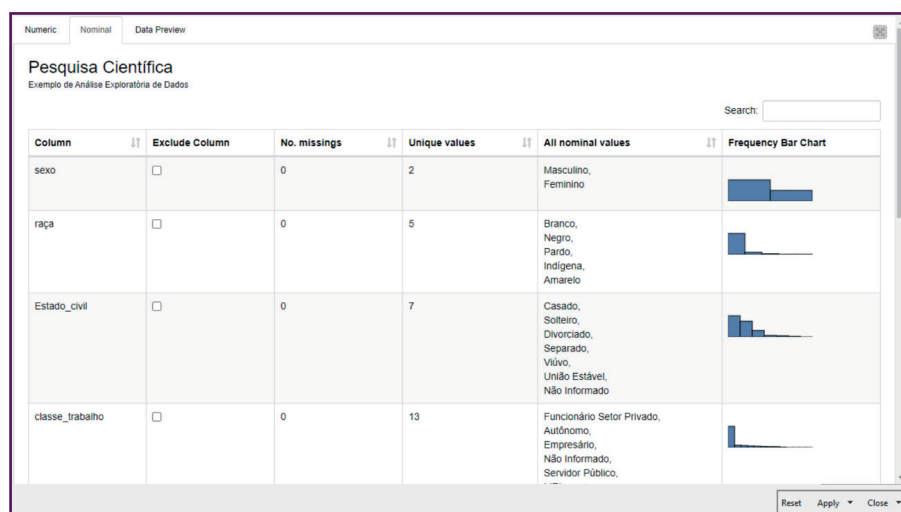
Showing 1 to 2 of 2 entries

Reset Apply Close

Fonte: Elaborado pela autora (2023).

A saída nominal (Figura 19) lista as principais propriedades dos valores nominais, como o número de valores ausentes, valores únicos e os *n* valores mais frequentes e menos frequentes. Para cada coluna, o nó também calcula um histograma mostrando a distribuição dos valores numéricos ou as frequências com que os valores nominais ocorrem.

Figura 19 - Exemplo de resultado do nó "Data Explorer" para variáveis nominais

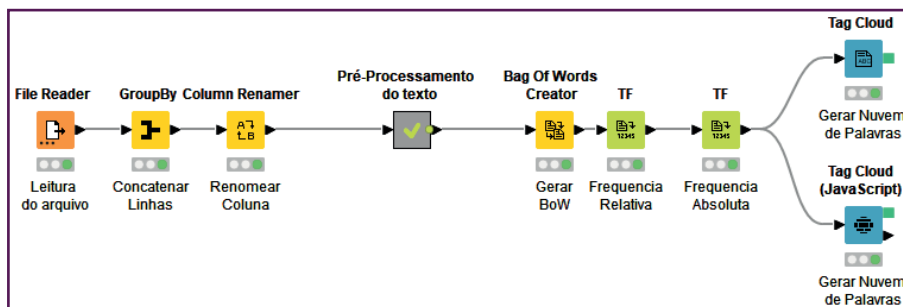


Fonte: Elaborado pela autora (2023).

Além de análises estatísticas, o *KNIME* oferece uma variedade de técnicas de análise de dados por meio de nós específicos. Desde análises descritivas até modelos estatísticos avançados, os pesquisadores podem explorar padrões e relacionamentos em seus dados. Alguns nós e extensões úteis para o pesquisador são apresentadas a seguir.

A extensão "*KNIME Textprocessing*" permite aos pesquisadores realizarem análises de texto dentro do ambiente do *KNIME*. Ela oferece um conjunto de nós para geração de *BoW*, cálculo de frequência de palavras, identificação de entidades nomeadas e etiquetas *Part-of-Speech tags* (POS) e geração de nuvem de palavras, entre outros, conforme exemplo apresentado na Figura 20.

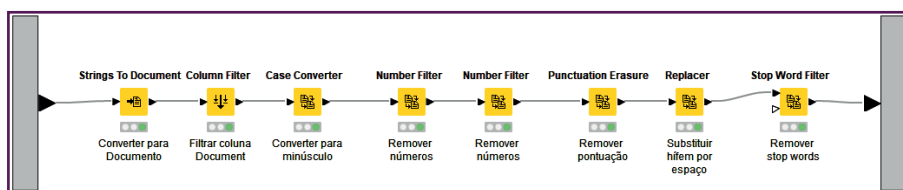
Figura 20 - Exemplo de workflow para análise de texto



Fonte: Elaborado pela autora (2023).

Essa extensão também apresenta um conjunto de nós para o pré-processamento do texto que permite realizar tarefas como *tokenização* (dividir o texto em palavras ou frases), remoção de números, pontuação e *stop words*, *stemming* (reduzir palavras à sua forma raiz) e *lematização* (transformar palavras em suas formas base ou lemas). No exemplo da Figura 20, alguns desses nós foram incluídos no metanó “Pré-Processamento do texto” que apresenta sua visualização expandida na Figura 21.

Figura 21 - Exemplo de metanó “Pré-Processamento do texto” expandido



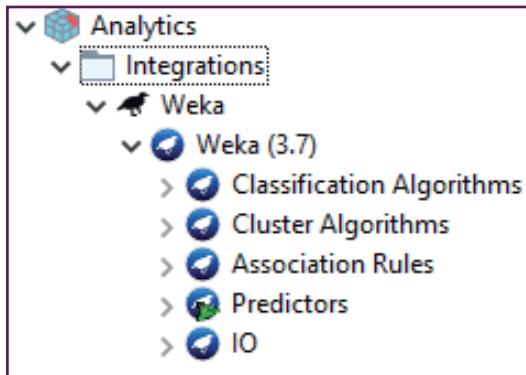
Fonte: Elaborado pela autora (2023).

Essa extensão é especialmente valiosa quando a pesquisa se trata de lidar com dados textuais em análises de texto, mineração de texto, Processamento de Linguagem Natural (PNL) e tarefas relacionadas.

Outra opção interessante é a integração “*KNIME Weka Data Mining Integration*” que incorpora as capacidades do *framework* de mineração de dados *Weka* adicionando ainda mais possibilidades de análise de dados. Essa integração permite aos pesquisadores explorarem uma gama ainda mais ampla de técnicas de mineração de dados e aprendizado de máquina

para descobrir *insights* valiosos para suas pesquisas. Após sua instalação, seus nós estarão disponíveis no repositório de nós na categoria “*Analytics* → *Integrations*” conforme ilustrado na Figura 22.

Figura 22 -Exemplo de nós da integra com a ferramenta Weka

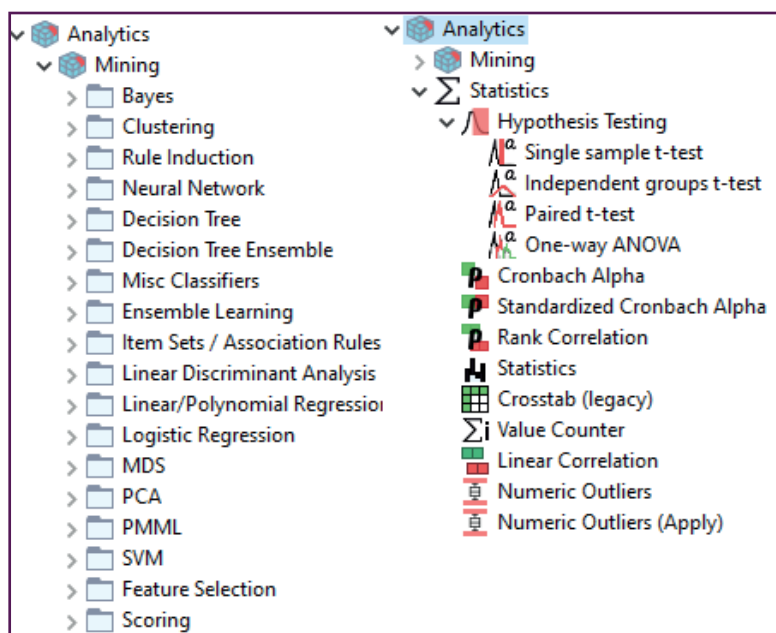


Fonte: Elaborado pela autora (2023).

No repositório de nós (Figura 23), na categoria “*Analytics* → *Mining*”, encontra-se diversas opções de nós para a análise de dados conforme a necessidade da pesquisa. Esses nós fornecem opções para explorar padrões, fazer previsões e outras possibilidades de análise nos dados. Para um entendimento completo e detalhado de cada nó é recomendável explorar a aba de descrição fornecida pela própria plataforma.

Na categoria “*Analytics* → *Statistics*” do repositório de nós, encontra-se um conjunto de nós da extensão “*KNIME Statistics Nodes*” que oferece uma variedade de nós estatísticos para análise de dados (Figura 23). Esses nós possibilitam aos cientistas calcularem estatísticas descritivas precisas, realizar testes de hipóteses significativos e explorar correlações complexas entre variáveis.

Figura 23 - Exemplo de nós para análise de dados



Fonte: Elaborado pela autora (2023).

Além do nó "*Statistics*", observa-se o nó "*Linear Correlation*" para calcular a correlação linear entre duas variáveis contínuas, medindo a força e a direção da relação entre essas variáveis. O nó "*Rank Correlation*" calcula a correlação entre duas variáveis ordinais. O nó "*Cronbach Alpha*" calcula o coeficiente alfa de *Cronbach*, uma medida de consistência interna de um conjunto de itens ou perguntas em um questionário ou escala. Ele é amplamente utilizado na área de pesquisa para avaliar a confiabilidade de um teste ou instrumento de medição. A extensão engloba alguns nós para testes de hipóteses, como teste *t* e Análise de Variância (ANOVA) que ajudam a fazer inferências sobre populações com base em amostras de dados, e nós para análise de regressão linear, permitindo prever variáveis dependentes com base em variáveis independentes.

Para uma compreensão aprofundada de como utilizar o *KNIME* para análise de texto e aplicar essa ferramenta em pesquisas envolvendo análises métricas da informação e revisão sistemática de literatura, explore os *workflows* disponíveis no *KNIME Hub*, por exemplo, os *workflows* "*PubMed Network Analysis*", "*PubMed Literature Search*", "*Creating a Corpus of*

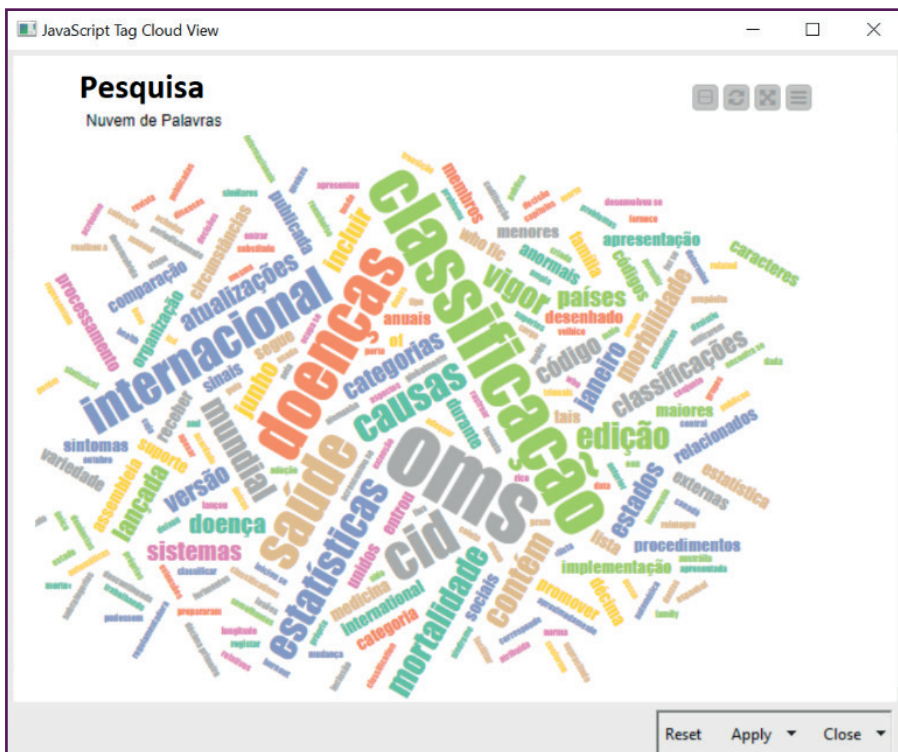
Documents" e "*Text mining techniques in life sciences literature*". Esses recursos são úteis para pesquisadores que necessitam analisar conjuntos de dados textuais, como artigos científicos e literatura acadêmica. Ao estudar esses *workflows*, os cientistas podem aprender técnicas avançadas de mineração de texto, análise de redes e criação de corpora de documentos, necessárias para realizar análises métricas detalhadas da informação e revisões sistemáticas de literatura de forma eficaz.

Pesquisadores que desejam realizar análises de modelagem de tópicos, por exemplo usando a técnica *Latent Dirichlet Allocation* (LDA), os *workflows* "*Topic Modeling on Biomedical Literature*," "*Topic Modeling with LDA*," "*Identify and visually represent a topic model*," "*Topic Detection LDA: Summarizing Romeo & Juliet or cataloging News*" e "*IS Literature Mining with Topic Detection*" são úteis. Para uma introdução prática e detalhada sobre como realizar análises de sentimentos e trabalhar com dados do *Twitter* e análise de sentimentos, existem diversos modelos de fluxos de trabalho disponíveis no *KNIME Hub*, por exemplo: "*Sentiment Analysis*," "*Twitter Data Collection*" e "*Twitter Data Analysis*".

10.3.2.4 VISUALIZAÇÃO DE DADOS COM KNIME

Na tarefa de visualização de dados, a *plataforma KNIME* oferece uma ampla gama de nós para visualização com gráficos, tabelas e outros. No *workflow* da Figura 20, observa-se o nó "*Tag Cloud (Java Script)*". Como visto na Figura 24, ele cria nuvens de palavras, nas quais o tamanho e a cor das palavras são determinados pela frequência dos termos no conjunto de dados.

Figura 24 - Exemplo de visualização de nuvem de palavras

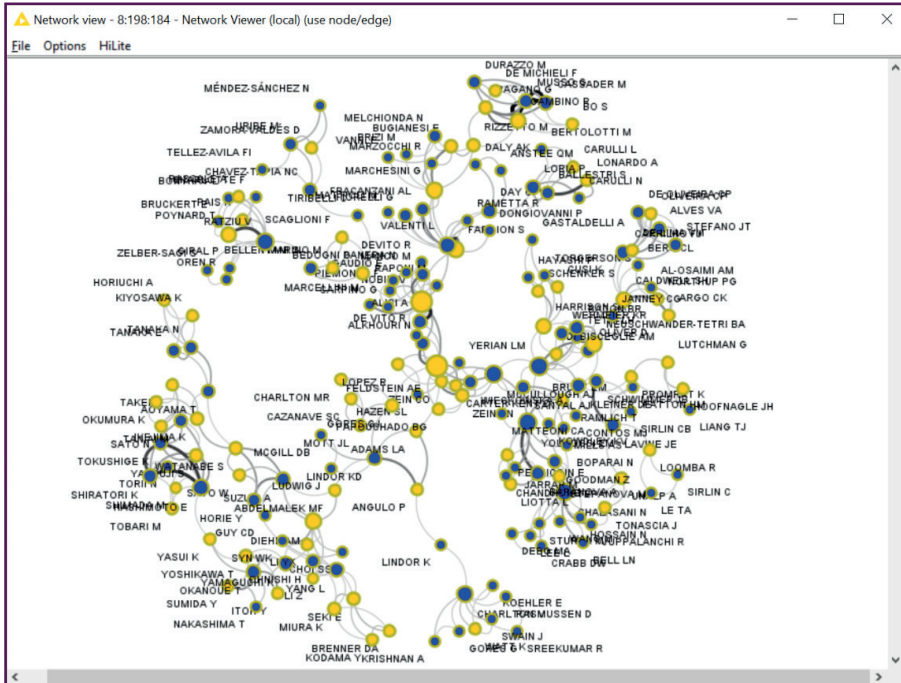


Fonte: Elaborado pela autora (2023).

Outra visualização muito usada em pesquisas que necessitam construir *gráficos de rede* para mostrar conexões entre diferentes entidades em um conjunto de dados é criada com o nó "Network Viewer (local)". Ele permite a criação desse tipo de visualização, conforme ilustrado na Figura 25, um exemplo de uma rede de colaboração de autores extraída do *workflow "PubMed Network Analysis"* ¹¹².

112 Disponível em: https://hub.knime.com/knime/spaces/Examples/latest/08_Other_Analytics_Types/05_Network_Mining/07_Pubmed_Author_Network_Analysis-wWLJp6ofl-qo9VnUq. Acesso em: 20 set. 2023.

Figura 25 - Exemplo de um gráfico de rede de colaboração

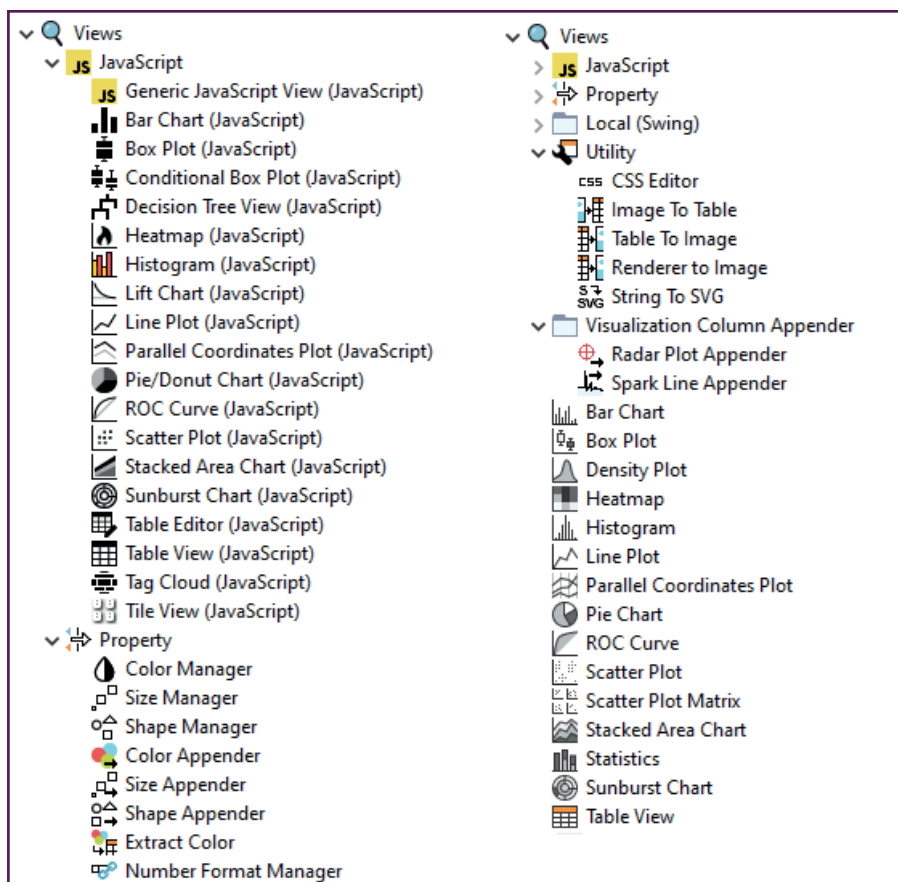


Fonte: Elaborado pela autora (2023).

Na categoria “Views” (Visualizações) do repositório de nós, existe uma variedade de nós dedicados à visualização de dados (Figura 26). Esses nós são projetados para criar diferentes tipos de gráficos, tabelas e outras representações visuais dos dados em seus fluxos de trabalho. Eles oferecem opções para personalizar a aparência e a interatividade das visualizações.

O nó “Table View (JavaScript)” permite exibir dados em formato de tabela. Ele é útil para uma inspeção detalhada dos dados tabulares, exibindo as informações em um formato claro e organizado, como as tabelas de saída do nós “Data Explores” da Figura 18 e Figura 19.

Figura 26 - Exemplo de nós para análise de dados



Fonte: Elaborado pela autora (2023).

Essa lista de nós permite criar diferentes tipos de gráficos, alguns deles que podem ser gerados no *KNIME* e os respectivos nós disponíveis incluem:

- Gráficos de Barras: Permitem comparar a frequência ou proporção de diferentes categorias de dados usando barras verticais ou horizontais. Podem ser usados os nós " *Bar Chart*";
- Gráficos de Linhas: São ideais para mostrar tendências ao longo do tempo ou em uma série de dados sequenciais. Podem ser criados pelos nós " *Line Plot*";

- Gráficos de Dispersão: Mostram a relação entre duas variáveis, sendo úteis para identificar padrões ou correlações nos dados e gerados pelos nós "*Scatter Plot*";
- Gráficos de Pizza (Pie) e de Rosca (Donut): *Gráficos de pizza* representam partes de um todo usando setores circulares, usados para mostrar proporções percentuais. Os *gráficos de Rosca* são semelhantes aos de pizza, mas com um buraco no centro, permitindo mostrar várias categorias e subcategorias. Podem ser usados os nós "*Pie Chart*" ou "*Pie/Donut Chart*" e, em sua janela de configuração, deve-se optar por uso o tipo pizza ou rosca;
- Histogramas: Exibem a distribuição de uma variável contínua por meio de barras verticais, mostrando a frequência de valores em intervalos específicos, criados pelos nós "*Histogram*";
- Gráficos de Caixa: Mostram a distribuição estatística dos dados, incluindo a mediana, quartis e possíveis *outliers*. Podem ser usados os nós "*Box Plot*";
- Mapas de Calor: Representam dados em uma matriz de cores, sendo úteis para visualizar padrões em grandes conjuntos de dados. Os nós "*Heatmaps*" são responsáveis por gerar essa visualização;
- Gráficos de Função de Densidade: Exibem a distribuição de dados contínuos, mostrando onde os dados são mais densos. Pode ser criado usando os nós "*Density Plot*";
- Gráfico de mosaico (treemap): Nesse gráfico, a hierarquia dos dados é representada por retângulos aninhados e cada retângulo equivale a uma categoria específica, sendo o tamanho do retângulo proporcional à quantidade ou ao valor numérico associado a essa categoria. O nó "*Tile View (JavaScript)*" pode gerar um gráfico desse tipo;
- Gráfico de Área Empilhada: Esse tipo de gráfico é útil para mostrar a composição e a tendência de múltiplas séries de dados ao longo do tempo ou de outras variáveis. Cada série é representada como uma área colorida no gráfico, empilhada sobre a série anterior. O nó "*Stacked Area Chart*" permite criar visualizações desse tipo.

Além dessas opções, algumas integrações como “*KNIME Power BI Integration*” e “*KNIME Tableau Integration*” são úteis para quem deseja usar o *KNIME* apenas para coleta, preparação e análise, e deixar as visualizações para outras ferramentas que permitem uma apresentação ainda mais rica e dinâmica.

A visualização de dados desempenha um papel importante na interpretação e comunicação de resultados de pesquisa em diversos contextos acadêmicos, desde artigos científicos a dissertações e teses. Antes de usar a ferramenta, é essencial que os pesquisadores compreendam como seus dados devem ser representados visualmente para transmitir seus resultados de pesquisa de forma clara. Após definir os requisitos visuais específicos para seu relatório de pesquisa, recomenda-se explorar o repositório *KNIME Hub* para encontrar recursos que atendam às suas necessidades específicas.

10.3.3 *KNIME COMMUNITY HUB*

O *KNIME Community Hub* ou apenas *KNIME Hub* é um repositório on-line integrado a *KNIME*, dedicado a reunir e compartilhar conhecimentos e recursos criados pela comunidade global de usuários da ferramenta. Esse repositório serve como um depósito centralizado, no qual pesquisadores e cientistas de dados podem acessar uma variedade de fluxos de trabalho, nós, componentes e extensões (KNIME, 2023e).

Para os pesquisadores, o *KNIME Community Hub*¹¹³ oferece cerca de 18250 fluxos de trabalho, 4504 nós, 1503 componentes e 239 extensões. Eles podem encontrar fluxos de trabalho prontos para uso, desenvolvidos por especialistas em diferentes áreas de pesquisa. Além disso, há uma variedade de nós, componentes e extensões que podem ser incorporados diretamente nos fluxos de trabalho dos pesquisadores, permitindo-lhes criar soluções personalizadas para suas pesquisas científicas.

113 Disponível em: <https://hub.knime.com/>. Acesso em: 20 set. 2023.

10.4 CONSIDERAÇÕES FINAIS

O presente capítulo mergulhou nas possibilidades oferecidas pela *Plataforma KNIME Analytics*, destacando-a como uma ferramenta de valor para ser usada nas pesquisas científicas, sobretudo na área de CSA. Em um cenário no qual as tecnologias de informação e comunicação desempenham um papel central, esse estudo demonstrou de forma clara o papel dessa plataforma na coleta, preparação, análise e visualização de dados durante as etapas metodológicas de pesquisas acadêmicas.

Em resumo, o *KNIME* se sobressai por sua capacidade de integração, flexibilidade e variedade de extensões, tornando-o uma boa escolha para pesquisadores em diversos campos de pesquisa. Sua flexibilidade também é uma vantagem fundamental, permitindo que os pesquisadores adaptem suas metodologias de acordo com as necessidades específicas de seus projetos. Ao oferecer suporte em todas as fases da pesquisa científica, a plataforma possibilita uma abordagem mais eficiente e completa para a análise de dados, economizando tempo valioso que pode ser direcionado para a análise e interpretação dos resultados e para descobertas significativas.

Entretanto, enquanto celebramos as vantagens da plataforma, é imperativo reconhecer os desafios éticos e metodológicos associados ao uso da TI em pesquisas em CSA. A preservação da privacidade dos dados, a avaliação rigorosa da confiabilidade das fontes *on-line* e a interpretação cuidadosa dos resultados são questões que devem permanecer no centro das discussões dos pesquisadores. A tecnologia é uma aliada poderosa, mas, somente quando associada a uma abordagem ética e crítica, ela pode verdadeiramente impulsionar a pesquisa científica para novos horizontes.

REFERÊNCIAS

- BERTHOLD, Michael R *et al.* KNIME: the konstanz information miner. In: **4th Annual Industrial Simulation Conference (ISC)**, Palermo, Itália: CentAUR, p. 58-61, 2006. Disponível em: https://centaur.reading.ac.uk/6139/1/2006_DiFatta06-MASS-ISC.pdf. Acesso em: 10 ago. 2023.
- BERTHOLD, M. R. *et al.* KNIME - the Konstanz information miner: version 2.0 and beyond. **Acm Sigkdd Explorations Newsletter**, [s. l.], v. 11, n. 1, p. 26-31, 16 nov. 2009. Association for Computing Machinery (ACM). DOI: <https://doi.org/10.1145/1656274.1656280>. Disponível em: <https://dl.acm.org/doi/10.1145/1656274.1656280>. Acesso em: 10 out. 2023.
- BIOLCHINI, J. *et al.* Systematic review in software engineering. **Technical report**, Rio de Janeiro, v. 679, n. 05, p. 45, May 2005. Disponível em: <https://www.cos.ufrj.br/uploadfile/es67905.pdf>. Acesso em: 10 out. 2023.
- BRIZOLA, Jairo; FANTIN, Nádia. Revisão da literatura e revisão sistemática da literatura. **Revista de Educação do Vale do Arinos**, Juara, v. 3, n. 2, p. 23-39, jul./dez. 2016. Disponível em: <https://periodicos.unemat.br/index.php/relva/article/view/1738/1630>. Acesso em: 10 ago. 2023.
- GALVÃO, Maria Cristiane Barbosa; RICARTE, Ivan Luiz Marques. Revisão sistemática da literatura: conceituação, produção e publicação. **Logeion: Filosofia da Informação**, [s. l.], v. 6, n. 1, p. 57-73, 15 set. 2019. Disponível em: <https://revista.ibict.br/fiinf/article/view/4835/4187>. Acesso em: 10 out. 2023.
- GARCIA, Ana Cristina B. *et al.* Groupware 4.0: avanços e desafios da computação social. **Jornada de Atualização em Informática**. Porto Alegre: SBC, 2020, v. 39, p. 142-186. Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/57/252/492-1>. Acesso em: 11 out. 2023.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.
- GRÁCIO, M. C. C. *et al.* **Tópicos da bibliometria para bibliotecas universitárias**. São Paulo: UNESP, 2020.

HAYASAKA, S.; SILIPO, R. **KNIME Analytics Platform for Beginners**. Zurich, Switzerland: KNIME Press, 2023.

KHODNENKO, Ivan *et al.* **A Lightweight Visual Programming tool for Machine Learning and Data Manipulation**. Las Vegas: IEEE, 2020, p. 981-985. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9458229>. Acesso em: 10 out. 2023.

KITCHENHAM, Barbara. **Procedures for Performing Systematic Reviews**. Keele: Keele University, 2004, v. 33, 28 p. Disponível em: <https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf>. Acesso em: 11 out. 2023.

KNIME. **KNIME Best Practices Guide**. Version 5.1. Zurich, Switzerland: KNIME AG, 2022a, p. 20. Disponível em: https://docs.knime.com/latest/analytics_platform_best_practices_guide/analytics_platform_best_practices_guide.pdf. Acesso em: 11 out. 2023.

KNIME. **KNIME File Handling Guide**. Version 5.1. Zurich, Switzerland: KNIME AG, 2022b, p. 41. Disponível em: https://docs.knime.com/latest/analytics_platform_file_handling_guide/analytics_platform_file_handling_guide.pdf. Acesso em: 11 out. 2023.

KNIME. **Extensions and Integrations Guide**. Version 5.1. Zurich, Switzerland: KNIME AG, 2023a, p. 4. Disponível em: https://docs.knime.com/latest/analytics_platform_extensions_and_integrations/analytics_platform_extensions_and_integrations.pdf. Acesso em: 11 out. 2023.

KNIME. **KNIME Analytics Platform Installation Guide**. Version 5.1. Zurich, Switzerland: KNIME AG, 2023b, p. 29. Disponível em: https://docs.knime.com/latest/analytics_platform_installation_guide/analytics_platform_installation_guide.pdf. Acesso em: 11 out. 2023.

KNIME. **KNIME Analytics Platform User Guide**. Version 5.1. Zurich, Switzerland: KNIME AG, 2023c, p. 50. Disponível em: https://docs.knime.com/latest/analytics_platform_user_guide/analytics_platform_user_guide.pdf. Acesso em: 11 out. 2023.

KNIME. **KNIME Big Data Extensions User Guide**. Version 5.1. Zurich, Switzerland: KNIME AG, 2023d, p. 19. Disponível em: https://docs.knime.com/latest/analytics_platform_extensions_and_integrations/big_data_extensions_user_guide.pdf.

knime.com/latest/bigdata_extensions_user_guide/bigdata_extensions_user_guide.pdf. Acesso em: 11 out 2023.

KNIME. **KNIME Community Hub User Guide**. Version 1.6. Zurich, Switzerland: KNIME AG, 2023e, p. 35. Disponível em: https://docs.knime.com/2023-07/hub_user_guide/index.pdf. Acesso em: 11 out. 2023.

KNIME. **KNIME Components Guide**. Version 5.1. Zurich, Switzerland: KNIME AG, 2023f, p. 59. Disponível em: https://docs.knime.com/latest/analytics_platform_components_guide/analytics_platform_components_guide.pdf. Acesso em: 11 out.2023.

KNIME. **KNIME Database Extension Guide**. Version 5.1. Zurich, Switzerland: KNIME AG, 2023g, p. 89. Disponível em: https://docs.knime.com/latest/db_extension_guide/db_extension_guide.pdf. Acesso em: 11 out. 2023.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos da metodologia científica**. 7. ed. São Paulo: Atlas, 2010.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Metodologia do trabalho científico**: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos. 3. ed. São Paulo: Atlas, 1990.

LEONELLI, S. **A pesquisa científica na era do Big Data**: cinco maneiras que mostram como o big data prejudica a ciência, e como podemos salvá-la. Rio de Janeiro: FIOCRUZ, 2022.

MARCONI, Marina de. Andrade; LAKATOS, Eva Maria. **Técnicas de pesquisa**. 5. ed. São Paulo: Atlas, 2002.

MINAYO, M. C. S. (org.). **Pesquisa Social**: teoria,método e criatividade. 21. ed. Petrópolis: Vozes, 2002.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. Mineração de textos. **Relatório Técnico-Instituto de Informática (UFG)**, 2007. Disponível em: https://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf. Acesso em: 11 out. 2023.

MOURA, A. L. T.; AMORIM, D. G. Big Data: o impacto e sua funcionalidade na sociedade tecnológica. **Revista Opara**, Petrolina,, v. 4, n. 1, p. 53-64, 2014.

MUELLER, S. P. M. O impacto das tecnologias de informação na geração do artigo científico: tópicos para estudo. **Ciência da Informação**, Brasília, DF, v. 23, n. 3, p. 309-317, set./dez. 1994. DOI: <https://doi.org/10.18225/ci.inf.v23i3.528>. Disponível em: <https://revista.ibict.br/ciinf/article/view/528>. Acesso em: 11 out. 2023.

PRADO, M. A. R. do; CASTANHA, R. C. G. Indicadores: conceitos fundamentais e importância em CT&I. In: GRÁCIO, M. C. C. (org.). **Tópicos da bibliometria para bibliotecas universitárias**, São Paulo: Cultura Acadêmica, 2020, p. 50-71.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013.

SOUZA, A. D. *et al.* A tipologia das fontes de informação em saúde: suporte à tomada de decisão. **Asklepion**: Informação em Saúde, Rio de Janeiro, v. 2, n. 1, p. 51-74, 28 jun. 2022. Disponível em: <https://brapci.inf.br/index.php/res/v/198107>. Acesso em: 11 out. 2023.

SWIECH, C.; FRANCISCO, A. C. A.; LIMA, S. A. A tecnologia da informação e comunicação transformando e inovando a prática da pesquisa científica. **Revista ESPACIOS**, [s. l.], v. 37, n. 11, p. 14, 30. jan. 2016. Disponível em: <https://www.revistaespacios.com/a16v37n11/16371115.html>. Acesso em: 11 out. 2023.

DADOS DA AUTORA:

Fernanda Farinelli



Fernanda Farinelli é Professora Adjunta na Faculdade de Ciência da Informação da UnB. Doutora em Gestão e Organização do Conhecimento pela Escola de Ciência da Informação da UFMG pesquisando o tema ontologias formais realistas como solução de integração semântica de dados. Ontologista responsável pelo projeto da OntONEo (Ontologia do domínio obstétrico e neonatal). Pesquisadora visitante no Departamento de Filosofia e no Departamento de Informática Biomédica da Universidade Estadual de Nova York em Buffalo entre 05/2015 e 04/2017. Mestre em Administração de Empresas com ênfase em Gestão estratégica da informação (Fundação Pedro Leopoldo/MG). Especialista em Banco de Dados (UNI-BH). Bacharel em Ciência da Computação (PUC-MG). Possui mais de 15 anos de experiência em Gestão de Dados atuando com administração de banco de dados, arquitetura e administração de dados e implantação de governança de dados em grandes empresas como Unisys Brasil, Cedro Têxtil, Prodemge. Atua há cerca de 15 anos como docente em cursos de graduação e pós-graduação em renomadas instituições de ensino no estado de Minas Gerais como PUC-MG, IEC, Fundação Pedro Leopoldo, Universidade de Itaúna, Faculdade Cotemig, Unipac e IGTI. Possui as certificações CDMP, CBIP, CDP e OCP.

<https://orcid.org/0000-0003-2338-8872>

fernanda.farinelli@unb.br

Como referenciar o capítulo 10:

FARINELLI, Fernanda. Revolucionando a pesquisa científica com a plataforma KNIME Analytics. *In*: SHINTAKU, Milton; MACÊDO, Diego José; MARIN, Luciano Heitor Gallegos (org.). **Tecnologias utilizadas em pesquisas acadêmicas em Ciências Sociais Aplicadas**. Brasília, DF: Ibict, 2023. cap. 10. p. 275-326. ISBN 978-65-89167-94-5. DOI: <http://doi.org/10.22477/9786589167938cap10>.