



10º EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



PLATAFORMA NBViz PARA ANÁLISE E VISUALIZAÇÃO DE DADOS A PARTIR DA FUSÃO DE BASES DE DADOS CIENTÍFICAS

Leonardo Neves Bolfarini

<https://orcid.org/0009-0002-4769-0750> | leonardobolfarini@gmail.com

Universidade de Marília (UNIMAR)

Rafael Gutierrez Castanha

<http://orcid.org/0000-0002-3117-1780> | r.castanha@gmail.com

Universidade de Marília (UNIMAR)

Resumo:

Tem por objetivo apresentar o desenvolvimento da NBViz, plataforma capaz de fundir metadados das bases *Web Of Science* e *Scopus*, bem como, proporcionar indicadores bibliométricos e recursos de visualização de dados a partir da produção científica combinada destas bases. A metodologia envolveu análise estrutural dos metadados, padronização de campos, deduplicação de registros e desenvolvimento de uma API e interface web para processamento e exploração dos dados. A plataforma permite fusão das bases, geração e construção de redes de coautoria e coocorrência de palavras-chave, e análises estatísticas. Constatou-se que a NBViz possui alto desempenho computacional além de boa usabilidade.

Palavras-Chave: Bases de dados científicas. Bibliometria. Visualização de dados.

EIXO TEMÁTICO:

Técnicas, Ferramentas e Infraestruturas

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.1042

COMO CITAR ESTE ARTIGO:

BOLFARINI, Leonardo Neves; CASTANHA, Rafael Gutierrez. Plataforma NBViz para análise e visualização de dados a partir da fusão de bases de dados científicas. In: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. **Anais [...]**. Curitiba, 2026. DOI: 10.22477/x.ebbc.1042. Disponível em: <https://doi.org/10.22477/x.ebbc.1042>.



1 INTRODUÇÃO

A expansão das publicações científicas nas últimas décadas tem intensificado a necessidade de ferramentas para integrar, tratar e visualizar dados bibliográficos de forma eficiente. Entre as principais fontes de dados científicos estão as bases de dados *Web of Science* e *Scopus*, que, embora não sejam de acesso aberto, são amplamente utilizadas para formulação de políticas científicas, avaliações e estudos bibliométricos. Contudo, ambas apresentam estruturas distintas de metadados, dificultando junção de ambas para consolidação e análise das informações. Com isso, pesquisadores enfrentam desafios para realizar estudos perante a necessidade de utilizar essas bases de forma combinada para construção e análise de indicadores científicos diversos.

De acordo com Vidal *et al.* (2023), os indicadores científicos contemplam múltiplas dimensões da produção e da comunidade científica, permitindo analisar padrões, dinâmicas e impactos do fazer científico. Essas métricas são fundamentais para compreender o progresso da ciência e monitorar seu desempenho ao longo do tempo. Nessa perspectiva, Mugnaini, Jannuzzi e Quoniam (2004) argumentam que o desenvolvimento científico e tecnológico de um país depende da capacidade de produzir indicadores confiáveis, capazes de subsidiar políticas públicas, programas de avaliação e alocação de recursos.

Historicamente, a avaliação científica e pesquisas cientométricas se utilizam das bases *Web of Science* e *Scopus*, sendo que estudos comparativos demonstram que, enquanto a primeira mantém maior rigor e seletividade em sua indexação, a *Scopus* oferece uma cobertura de periódicos mais abrangente (Singh *et al.*, 2021). No contexto nacional, é sabido que instituições de avaliação e fomento como por exemplo, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CA-

PES), utilizam-se de dados oriundos dessas bases para formulação de políticas de avaliação científica.

Atualmente, a *Web of Science* e *Scopus* destacam-se como as principais bases de dados bibliográficas internacionais. A primeira reúne mais de 99 milhões de registros de periódicos, anais de eventos e livros, enquanto a segunda ultrapassa 100 milhões, oferecendo ampla cobertura temática e atualização contínua. A integração dessas bases configura oportunidade estratégica para análises mais robustas e representativas da produção científica indexada nessas bases

Diante do exposto, esta pesquisa busca responder ao seguinte problema: Qual é a contribuição de uma plataforma pública, denominada NBViz, voltada à integração, padronização e visualização de dados oriundos das bases *Web of Science* e *Scopus* para o fortalecimento de análises bibliométricas mais abrangentes e consistentes? Tendo por objetivo, apresentar o desenvolvimento e testagem da NBViz, uma plataforma capaz de fundir metadados das bases *Web Of Science* e *Scopus*, bem como, proporcionar indicadores bibliométricos e recursos de visualização de dados a partir da produção científica combinada destas bases. Tradicionalmente, combinar e padronizar dados de diferentes bases científicas é uma tarefa complexa devido às diferenças na estrutura de metadados e na nomenclatura dos campos.

Mesmo que tenham outras formas de combinar bases de dados, essas abordagens podem demandar alguns processos manuais ou um conhecimento técnico mais apurado, enquanto a NBViz (<https://nbviz.labs.unimar.br/>) auxilia desenvolve esse processo de forma automatizada, replicável e baseada em código aberto, que consolida os dados e oferece visualizações exploráveis por meio da exportação de gráficos, grafos e tabelas a partir de uma plataforma de interface amigável

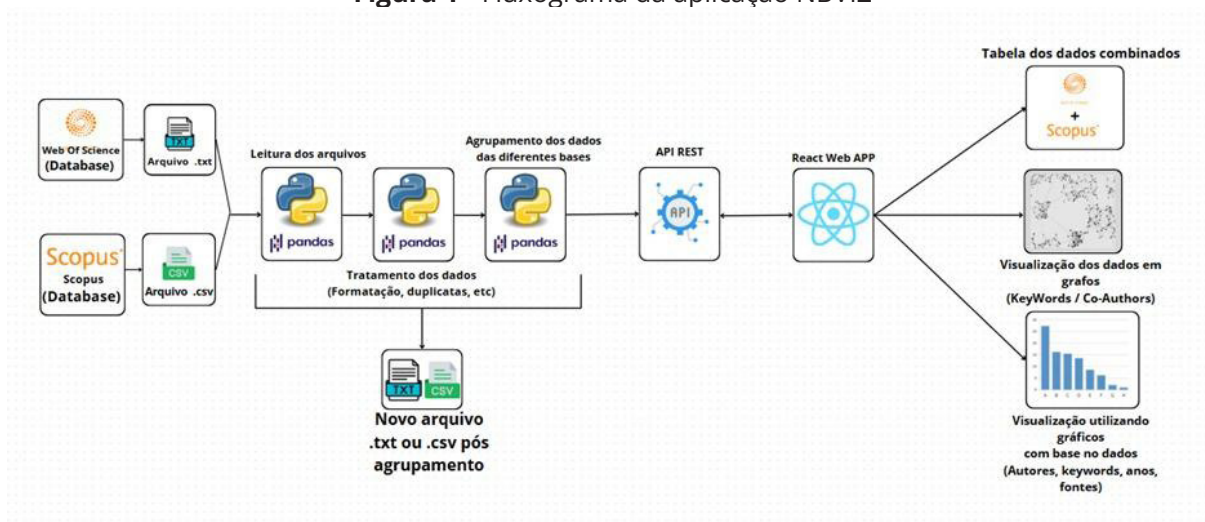
e de fácil utilização, ampliando seu acesso. Diferencia-se de outras soluções ao integrar, em um único ambiente, funcionalidades de fusão, padronização e análise, posicionando-se como uma alternativa para suprir essa lacuna no contexto das ferramentas bibliométricas. Ao integrar duas das principais bases de dados científicas do mundo, pesquisadores passam a ter acesso a uma representação mais completa da produção científica para suas análises.

2 METODOLOGIA

A metodologia foi estruturada em quatro etapas: coleta e análise dos dados, tratamento e transformação dos dados, desenvolvimento da API e da interface web. O fluxograma apresentado na Figura 1 sintetiza a arquitetura da aplicação, evidenciando o fluxo que se inicia na extração dos dados brutos e culmina na disponibilização de visualizações interativas e arquivos consolidados ao usuário.

A partir das bases de dados, exportam-se os registros da Scopus e Web of Science, considerando os formatos (.csv e .txt). A etapa inicial consistiu na análise estrutural dos arquivos, identificando diferenças de nomenclatura, organização e delimitadores dos campos (*field tags*). Implementou-se a padronização dos metadados: autores, título, ano, fonte, DOI, resumo e palavras-chave, com ajustes estruturais e transformações textuais. Destacam-se: i) conversão dos textos para minúsculas; ii) remoção de inconsistências de capitalização; iii) uniformização de delimitadores; iv) reorganização dos campos em estrutura comum. Essa etapa reduz ruídos e garante consistência, sobretudo na comparação de registros de sistemas distintos.

Figura 1 - Fluxograma da aplicação NBVIZ



Fonte: Elaboração dos autores

Após a padronização, realizou-se a deduplicação dos registros usando DOI e título. O DOI foi considerado identificador primário, enquanto o título foi usado como critério complementar em casos de ausência ou inconsistência. Outros campos não foram usados diretamente na eliminação de duplicatas, mas sua padronização aumenta a confiabilidade final. Com o algoritmo consolidado, procedeu-se à encapsulação em uma API REST, abstraindo a complexidade do usuário. A API recebe arquivos (.txt e .csv), executa padronização e deduplicação e retorna versões consolidadas, tornando o processo acessível e reproduzível, sem manipulação direta do algoritmo.

O usuário pode enviar os arquivos das bases e baixar um único arquivo “all_in_one”, contendo autores, título, ano, fonte, DOI, resumo e palavras-chave, em dois formatos: i) .txt com a nomenclatura da *Web of Science* (AU, TI, PY, SO, DI, AB, DE); ii) .csv com a nomenclatura da Scopus (*Authors, Title, Year, Source title, DOI, Abstract, Author Keywords*). Esse processamento permite o uso do arquivo em outros softwares de análise bibliométrica.

Além da combinação de arquivos, foram implementadas rotas para análise ex-

ploratória, incluindo cálculo de frequências de autores, palavras-chave e fontes, e séries temporais da produção científica. Também foram desenvolvidas análises relacionais para construção de redes de coautoria e coocorrência de palavras-chave. Essas redes usam a teoria dos grafos, em que cada autor ou termo é um vértice e as relações de colaboração ou coocorrência são arestas. Essa modelagem identifica padrões de colaboração e agrupamentos temáticos.

Com a API estruturada, iniciou-se o desenvolvimento da aplicação *web*, construída com [Next.js](#) e React. A interface modular separa funcionalidades de integração das bases, visualização estatística e análise de redes. A NBViz permite o envio de arquivos originais ou já tratados, permitindo gerar redes e análises diretamente de arquivos oriundos das bases. Também é possível baixar os dados consolidados. As visualizações de gráficos usam *Recharts*, componentes gráficos integrados ao *React*.

Para renderizar as redes foi usada a biblioteca *Sigma.js*, especializada em visualização interativa de grafos no *web*. *Sigma.js* usa *WebGL*, permitindo grafos de forma otimizada devido ao processamento acelerado por GPU. Para exportar os dados, implementou-se download em formato *Pajek (.net)*, permitindo análise em softwares especializados. Para demonstrar a funcionalidade, apresenta-se a interface gráfica nas Figuras 1a, 1b e 1c, além de testes de desempenho de arquivos, geração de redes e análise estatística.

3 RESULTADOS

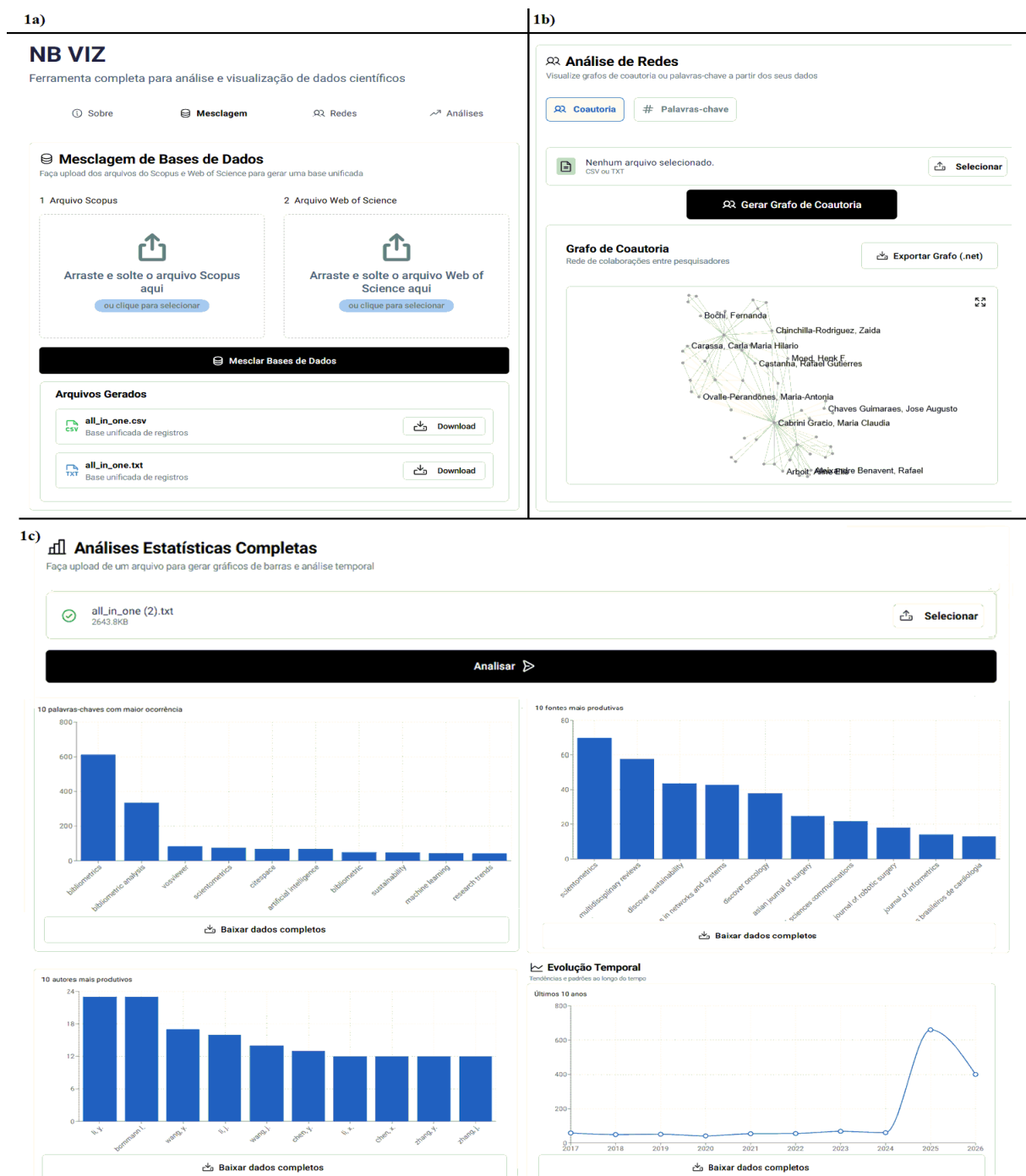
Com intuito de demonstrar a funcionalidade da NBViz, apresenta-se na Figura 1a a interface gráfica na qual o usuário insere os arquivos oriundos da *Scopus* e *Web of Science* e, ao clicar em “Mesclagem”, ocorre a geração de um arquivo único, passível de download em duas extensões: .txt com os cabeçalhos da Web of Science

e .csv com os cabeçalhos da *Scopus*, ambos intitulados "*all_in_one*". A partir desses arquivos, é possível gerar redes de coautoria ou palavras-chave, conforme a Figura 1b.

Tanto na rede de coautorias quanto na de palavras-chave, o usuário pode utilizar a funcionalidade de zoom para ampliar a rede e, ao selecionar um nó ou aresta, o sistema destaca em vermelho o valor da ponderação da aresta com valores de frequência, seja de coautorias ou ocorrências entre palavras-chave.

Além disso, é possível realizar o *download* da rede em extensão .net utilizando o botão "Exportar grafo (.net)". Já a Figura 1c explicita quatro análises estatísticas de distribuição de frequência em gráficos de barras: i) 10 palavras-chave de maior frequência; ii) 10 autores mais produtivos; iii) 10 fontes que mais publicaram no *corpus* analisado. Como última análise, apresenta-se a distribuição temporal dos documentos por ano de publicação em um gráfico de linhas, considerando os últimos 10 anos de produção. Mesmo que a interface exiba até 10 resultados por gráfico, o usuário pode baixar os dados completos de cada análise.

Figura 1 - Interface Gráfica das funcionalidades da NBviz



Fonte: Elaboração dos autores. Disponível em: <https://nbviz.labs.unimar.br/>

Em termos de desempenho, realizou-se um teste com os primeiros 10 mil registros obtidos a partir da busca em título, resumo e palavras-chave por "bibliometrics" em cada uma das bases (Web of Science e Scopus). A fusão entre as bases identificou 18696 registros únicos e, no processamento *back-end* (local), durou

2,80 segundos. Ao observar a geração de redes, notaram-se 41880 nós e 175637 relações (vértices) de coautoria, além de 14589 palavras-chave e 76499 relações (vértices) de coocorrência. A geração do arquivo de grafos durou 0,96 segundos para coautorias e 0,76 segundos para palavras-chave, e a visualização da rede de coautoria 0,48 segundos e 0,13 segundos para palavras-chave. Além disso, não foram encontradas perdas de dados após a combinação de dados perante os dados originais oriundos dos registros exportados das bases, atestando validade do processo. Em testes *online*, diversos fatores podem influenciar o desempenho da NBViz, pois o processamento de dados é feito no lado do servidor, assim, aspectos como capacidade computacional, memória disponível e carga de requisições podem impactar diretamente no tempo de execução.

4 CONSIDERAÇÕES FINAIS

Esta pesquisa apresentou e demonstrou o uso da plataforma NBViz, solução de código aberto para processamento de dados científicos, tendo como principais funcionalidades a fusão de arquivos de registros da *Web of Science* e *Scopus*, geração e visualização de redes de coautoria e coocorrência de palavras-chave, além de análises de distribuição de frequência. Todos os resultados são passíveis de exportação. Por se tratar de uma aplicação online, a NBViz demonstrou capacidade de processar grande quantidade de dados, superando o processamento de *softwares* bibliométricos como, por exemplo, o VosViewer, que não apresenta bom desempenho na visualização de volumes muito altos de dados (em torno de 100 mil nós).

Destacam-se como limitações a quantidade de bases que a plataforma combina, além de carecer de outras análises relacionais, como cocitação e acoplamento bibliográfico. No que se refere ao uso das bases de dados, sabe-se que bases como *Dimensions* (também de acesso restrito) também poderiam ser integradas.

Além disso, a NBViz não contempla a base de dados *OpenAlex*, que fornece uma capacidade de exploração (de forma gratuita) muito grande, visto que possui API pública e disponibiliza a seus usuários a possibilidade de exportar um alto volume de dados.

Como trabalhos futuros, pretende-se transformar a NBViz em biblioteca da linguagem *Python*, compará-la, em termos de funcionalidades com outras ferramentas já existentes, além de implementar soluções de análise de redes utilizando as referências dos documentos, possibilitando análises como cocitação e acoplamento bibliográfico entre documentos e autores. Espera-se também fornecer a possibilidade de fusão de outras bases, sobretudo de acesso aberto, como a *OpenAlex*, *BRAPCI*, *BrCris* *OasisBR*, além da supracitada *Dimensions*. Por fim, a NBViz representa um avanço nas ferramentas bibliométricas gratuitas, de acesso aberto, com alto desempenho e fácil uso.

REFERÊNCIAS

MUGNAINI, Rogério; JANNUZZI, Paulo de Martino; QUONIAM, Luc. Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal. **Ciência da informação**, Brasília. v. 33, p. 123-131, 2004. DOI: <https://doi.org/10.1590/S0100-19652004000200013>. Acesso em: 04 mar. 2026.

SINGH, Vivek Kumar; SINGH, Prashasti; KARMAKAR, Mousumi; LETA, Jacqueline; MAY, Philipp. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. **Scientometrics**. Budapest, Hungria. 2021. DOI: <https://doi.org/10.1007/s11192-021-03948-5>

VIDAL, Lucia Helena Cunha; SEGUNDO, Washington Luís Ribeiro de Carvalho; MENA CHALCO, Jesús Pascual; GABRIEL JUNIOR, Gabriel Junior, Rene Faustino. BrCris: visualização de indicadores científicos. In: Encontro Nacional de Pesquisa em Ciência da Informação, 2023, Aracaju. **Anais [...]** Aracaju: [S. l.], 2023. Disponível em: https://enancib.ancib.org/enancib/pt_BR/article/view/1343. Acesso em: 04 mar. 2026.