

DELINÉIA: sistema de apoio ao delineamento de escopo temático no ensino superior

Rafael Antunes dos Santos

<https://orcid.org/0000-0002-1529-9063>
rderafa@gmail.com
Universidade Federal do Rio Grande do
Sul (UFRGS)

Eliseo Berni Reategui

<https://orcid.org/0000-0002-5025-9710>
eliseoreategui@gmail.com
Universidade Federal do Rio Grande do
Sul (UFRGS)

RESUMO:

Este trabalho apresenta o Delinéia, um sistema educacional de código aberto que integra análise de coocorrência de palavras-chave com inteligência artificial generativa para apoiar estudantes do ensino superior no delineamento do escopo de seus projetos de pesquisa. O problema investigado é: como a avaliação automatizada de definições preliminares de um projeto (tema, questão de pesquisa e palavras-chave) pode apoiar estudantes no delineamento mais preciso do escopo de suas propostas? O sistema coleta metadados da base OpenAlex via API, processa tópicos indexados para construção de grafos de coocorrência com NetworkX e utiliza o modelo Gemini 2.5 Pro para gerar feedback personalizado. A avaliação segue abordagem mista, com instrumento baseado nos modelos TAM3 e UTAUT. Resultados preliminares indicam potencial para auxiliar na delimitação temática e no desenvolvimento de habilidades em buscas acadêmicas.

Palavras-chave: Análise de coocorrência de palavras. Inteligência artificial generativa. Delineamento de pesquisa. Educação superior. OpenAlex.

EIXO TEMÁTICO:

Técnicas, Ferramentas e
Infraestruturas

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.952

COMO CITAR ESTE ARTIGO:

SANTOS, Rafael Antunes dos; REATEGUI, Eliseo Berni. Delinéia: sistema de apoio ao delineamento de escopo temático no ensino superior. *In*: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. **Anais [...]**. Curitiba, 2026. DOI: 10.22477/x.ebbc.952. Disponível em: <https://doi.org/10.22477/x.ebbc.952>.



1 INTRODUÇÃO

O delineamento do escopo de pesquisa representa um desafio significativo para estudantes do ensino superior, especialmente nos estágios iniciais de seus projetos de pesquisa. A delimitação do escopo de pesquisa envolve decisões deliberadas sobre quais aspectos incluir ou excluir do projeto, considerando restrições de tempo, recursos e viabilidade (Cooksey; McDonald, 2019), incluindo a identificação de palavras-chave adequadas. Tais decisões são importantes para uma delimitação temática eficaz e para o desenvolvimento de estratégias de buscas bibliográficas precisas. Ferramentas bibliométricas tradicionais, como *VOSviewer* e *Bibliometrix*, embora robustas para análise de produção científica consolidada, não foram concebidas com foco educacional ou para apoiar o processo iterativo de refinamento de projetos de pesquisa em desenvolvimento. Simultaneamente, o surgimento de modelos de Inteligência Artificial Generativa (IAG) criou novas possibilidades para *feedback* personalizado e assistência na escrita acadêmica, porém frequentemente desconectadas de evidências bibliométricas.

Este trabalho tem como objetivo apresentar a arquitetura metodológica e os procedimentos técnicos do sistema Delinéia¹, um sistema educacional que integra análise de coocorrência de palavras com IAG para apoiar estudantes no delineamento de seus projetos de pesquisa. O sistema combina dados da base *OpenAlex* com o modelo *Gemini 2.5 Pro* e oferece uma trilha de aprendizagem que valoriza o processo reflexivo do estudante na construção do conhecimento sobre o campo de estudo escolhido. A pesquisa explora a seguinte questão: como a avaliação automatizada de definições preliminares de um projeto, como tema, questão de pesquisa e palavras-chave, pode apoiar estudantes no delineamento mais preciso do escopo de suas propostas?

¹ Disponível em: <https://rafaelantunes123-delineia.hf.space>. Acesso em 9 jan. 2026.

2 FUNDAMENTAÇÃO

A análise de coocorrência de palavras constitui uma técnica bibliométrica que examina a frequência com que termos aparecem conjuntamente em publicações científicas, revelando estruturas conceituais de campos de conhecimento (Callon *et al.*, 1983). O modelo permite identificar conexões semânticas entre conceitos, evidenciando temas emergentes, consolidados ou declinantes (Aria; Cuccurullo, 2017). Já o *OpenAlex*, lançado em 2022, representa uma iniciativa de ciência aberta que disponibiliza metadados de mais de 250 milhões de trabalhos acadêmicos². Sua API permite acesso gratuito a conceitos hierarquizados, facilitando análises bibliométricas (Priem; Piwowar; Orr, 2022). Por sua vez, os modelos de linguagem de grande porte (LLMs), como o *Gemini*, demonstram capacidade para gerar texto coerente, oferecer explicações contextualizadas e auxiliar em tarefas de síntese informacional (Lei *et al.*, 2024). No contexto educacional, esses modelos podem funcionar como assistentes que fornecem *feedback* personalizado, respeitando as necessidades individuais dos estudantes (Chang *et al.*, 2024).

O comportamento de uso da informação compreende os atos físicos e mentais envolvidos na incorporação da informação encontrada à base de conhecimento preexistente do indivíduo (Wilson, 1999). Esse estágio é definido como uma atividade construtiva de busca por sentido, na qual o sujeito busca encontrar significados que se ajustem ao seu quadro de referência pessoal para estender o conhecimento sobre um problema ou tópico específico (Kuhlthau, 1991). O uso efetivo se concretiza na transformação da informação em conhecimento, culminando em uma nova compreensão ou solução que pode ser apresentada e compartilhada.

O grafo de coocorrência atua como “organizador prévio” (Ausubel, 2000), apresentando a estrutura semântica antes que o estudante refine sua busca, ancorando

² Disponível em: <https://openalex.org>. Acesso em 9 jan. 2026.

novos termos em tópicos já conhecidos. A teoria implica que recursos educacionais devem considerar os conhecimentos prévios de quem aprende, permitindo que novos conteúdos se ancorem em estruturas cognitivas preexistentes. Sistemas que personalizam o *feedback* conforme o perfil do usuário potencializam essa ancoragem ao apresentar informações contextualizadas às necessidades específicas de cada estudante.

3 METODOLOGIA

O *Delinéia* foi desenvolvido como uma aplicação *web* utilizando o *framework Streamlit* em *Python* e hospedagem no *HuggingFace Spaces*. A arquitetura integra três componentes principais: (1) módulo de coleta de dados via API do *OpenAlex*; (2) módulo de processamento e análise de coocorrência com *NetworkX*; e (3) módulo de geração de conteúdo via API do *Gemini 2.5 Pro*. A coleta de dados é realizada via *endpoint /works* da API do *OpenAlex*, com paginação de até 500 resultados. Os campos recuperados são: identificador, título (*display_name*), ano e data de publicação, tópicos indexados (*/topics*), autorias, localização primária, tipo de documento, contagem de citações, DOI e índice invertido do resumo. Não são aplicados filtros de ano, idioma ou país na busca, somente para tipo de documento (*article* e *review*). A filtragem ocorre sobre os tópicos indexados. São considerados tópicos com *score* de relevância igual ou superior a 0,35 e nível hierárquico igual ou superior a zero. Caso o resumo esteja ausente, o *prompt* do *Gemini* utiliza apenas o título e os tópicos de alta relevância, garantindo que a análise não seja interrompida. O grafo de coocorrência é construído com *NetworkX* a partir das listas de tópicos por artigo. Os 9 termos mais centrais são exibidos ao estudante, com limite baseado no princípio cognitivo de Miller (1956), 7 ± 2 .

A geração de conteúdo pelo Gemini 2.5 Pro é estruturada por meio de *prompts* instrucionais com papel definido (orientador acadêmico), contexto do projeto do estudante e diretrizes de tom e formato da resposta. Dado o volume e a especificidade dos *prompts* utilizados, à título ilustrativo, a avaliação das palavras-chave segue o seguinte esquema: (1) atribuição de papel (“Você é um professor universitário orientando um aluno de pesquisa”); (2) fornecimento do contexto do projeto (tema, questão, palavras-chave e nível de segurança fornecidos pelo estudante); (3) instrução estruturada para dois parágrafos avaliativos (um sobre as palavras-chave e outro sobre a questão de pesquisa) com início e encerramento prescritos; e (4) diretrizes de tom (direto, honesto, sem clichês). Os *prompts* estão disponíveis no repositório público do sistema³.

A precisão das respostas geradas foi avaliada de forma não sistemática ao longo do desenvolvimento, com inúmeros testes realizados pelos autores, e também em estudo preliminar publicado. (Santos; Reategui, 2025). Em testes internos, observou-se que o Delinéia reduz alucinações se comparado ao uso isolado do LLM, pois restringe o espaço de busca aos tópicos do *OpenAlex*.

O fluxo de uso compreende uma trilha de cinco etapas: (1) apresentação do problema de pesquisa pelo estudante; (2) visualização do grafo e do glossário de tópicos; (3) seleção de tópicos; (4) *feedback* personalizado com sugestão de palavras-chave e chaves de busca; e (5) avaliação. Cada etapa foi concebida como um momento de aprendizagem.

A interface incorpora elementos de gamificação, com distintivos que reconhecem o progresso do estudante: explorador, pesquisador, seletor, delineador e avaliador. O sistema *Delinéia* oferece visualizações interativas incluindo construtor de chaves de busca, histórico para comparar delineamentos e painel para exibição

³ Disponível em: <https://github.com/rderafa-rgb/Delineia>.

de métricas de artigos, de tópicos, de pares de tópicos, mapa temático de posicionamento estratégico de *clusters* e exportação em diferentes formatos de arquivos, operando princípios de ciência aberta para reprodutibilidade dos dados em outros artefatos de análise bibliométrica. Após o preenchimento do formulário inicial, o usuário recebe o relatório de delineamento, conforme a Figura 1:

Figura 1 - Relatório de delineamento
Delineamento do escopo da pesquisa

Estrutura do Relatório de Feedback

Dados do Projeto

Tema de pesquisa:
Questão de pesquisa:
Palavras-chave iniciais:

Nível de confiança inicial:
Distintivos: ■ Explorador | ■ Pesquisador | ■ Seletor | ■ Delineador

A

Avaliação do Projeto

- Parecer sobre palavras-chave originais e sugestões de refinamento
- Análise da questão de pesquisa (formulação, viabilidade, operacionalização)
- Comparação entre busca espontânea e palavras-chave formais
- Recomendação de consulta ao orientador

B

Métricas da Análise Bibliométrica e Grafo de Coocorrência

Métricas	
Artigos analisados:	500
Tópicos identificados:	735
Coocorrências:	8891

Tópicos selecionados: Publishing, Political science, Scholarly communication, Computer science...

Grafo de Coocorrência (OpenAlex)

exemplo de visualização

C

Glossário de Tópicos

- Computer science (Ciência da Computação): definição contextualizada ao tema...
- Knowledge management (Gestão do Conhecimento): definição contextualizada...
- tópico_n: definição...

D

Interpretação Personalizada do Grafo

- Análise dos conceitos centrais e estrutura do campo
- Identificação de clusters temáticos e suas conexões
- Lacunas promissoras e oportunidades de pesquisa
- Recomendações para posicionamento do trabalho

E

Sugestões de Palavras-chave

- palavra-chave₁ (termo em inglês): justificativa e relevância para o projeto
- palavra-chave₂ (termo em inglês): justificativa...
- palavra-chave_n: ...

F

Chaves de Busca Sugeridas

- Chave ampla (exploratória): ("termo:" OR "termo:" OR "termo:")
- Chave focada (conceitos centrais): "termo:" AND "termo:" AND "termo:"
- Transparência: chave original usada na consulta ao OpenAlex

Objetivo: garantir reprodutibilidade da análise

G

Fonte: Estrutura baseada no relatório de *feedback* do sistema *Delinéia*.

Os dados são usados em análises sobre utilidade percebida, facilidade de uso, qualidade da informação, intenção de uso, confiança no sistema, experiência do usuário, satisfação geral (NPS), autoavaliação e perfil do respondente. Os itens do instrumento são baseados nos modelos *Technology Acceptance Model* (TAM3) e *Unified Theory of Acceptance and Use of Technology* (UTAUT), de Venkatesh *et al.* (2003) e Venkatesh e Bala (2008), além de questões abertas. A pesquisa ainda oferece opções complementares de continuidade mediadas por videoconferência para coleta de dados qualitativos, incluindo entrevistas individuais semiestruturadas, protocolo *think aloud* e grupos focais a distância, com estudantes de graduação e pós-graduação de instituições públicas de ensino superior, em diferentes fases de desenvolvimento de projetos acadêmicos (TCC, dissertações e teses). Estes dados, juntamente com os resultados das questões abertas, são analisados à luz da análise de conteúdo (Bardin, 2011) com categorias definidas *a posteriori*.

4 RESULTADOS E DISCUSSÃO

O sistema encontra-se em fase de coleta de dados. Resultados preliminares de testes piloto com estudantes de graduação e de pós-graduação indicam que a abordagem do sistema oferece apoio relevante em três dimensões: (a) delimitação temática, ao tornar visíveis, por meio do grafo de coocorrência, os tópicos mais centrais do campo e as relações entre eles, permitindo ao estudante identificar o núcleo conceitual da área e distingui-lo de temas periféricos ou emergentes; (b) descoberta de termos, ao sugerir tópicos relacionados não considerados inicialmente; e (c) fundamentação de escolhas, ao apresentar métricas como frequência de tópicos, centralidade de grau e densidade de coocorrência que oferecem ao estudante evidências objetivas para decidir quais termos incluir ou excluir do escopo do projeto.

A integração entre análise bibliométrica e IAG apresenta complementaridade. Enquanto os dados do *OpenAlex* fornecem evidências objetivas sobre a estrutura do campo, o modelo *Gemini* contextualiza essas informações em linguagem acessível e sugere recomendações adaptadas ao problema específico do estudante. Como limitações, identificam-se a dependência da qualidade dos metadados do *OpenAlex*, a necessidade de conectividade para acesso às APIs, o viés inerente aos modelos de linguagem, riscos sociais e comportamentais (Chang *et al.*, 2024) e potencial para problemas com alucinações e segurança científica (Lei *et al.*, 2024).

5 CONSIDERAÇÕES FINAIS

O *Delinéia* representa uma interseção entre bibliometria, IAG e educação superior, sendo ferramenta que oferece suporte para a formação acadêmica. O diferencial reside na ênfase educacional: a trilha de aprendizagem promove a reflexão do estudante sobre o processo de delineamento, favorecendo a construção ativa do conhecimento a partir do que o aprendiz já sabe. O código aberto e o uso de uma base aberta favorecem a reprodutibilidade e a adaptação pela comunidade acadêmica.

REFERÊNCIAS

ARIA, M.; CUCCURULLO, C. bibliometrix: An R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, v.11, n.4, p.959-975, 2017. DOI: <https://doi.org/10.1016/j.joi.2017.08.007>. Acesso em: 7 maio 2026.

AUSUBEL, D. P. **The acquisition and retention of knowledge**: a cognitive view. Dordrecht: Springer Science+Business Media, 2000. DOI: <https://doi.org/10.1007/978-94-015-9454-7>. Acesso em: 7 maio 2026.

BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2011.

CALLON, M; COURTIAL, J.-P.; TURNER, W. A.; BAUIN, S. From translations to problematic networks: an introduction to co-word analysis. **Social Science Information**, v. 22, n. 2, p. 191-235, 1983. DOI: <https://doi.org/10.1177/053901883022002003>. Acesso em: 7 maio 2026.

CHANG, Y.; WANG, X.; WANG, J.; WU, Y.; YANG, L.; ZHU, K.; CHEN, H.; YI, X.; WANG, C.; WANG, Y.; YE, W.; ZHANG, Y.; CHANG, Y.; YU, P. S.; YANG, Q.; XIE, X. A survey on evaluation of large language models. **ACM Transactions on Intelligent Systems and Technology**, v.15, n.3, 2024. DOI: <https://doi.org/10.1145/3641289>. Acesso em: 7 maio 2026.

COOKSEY, R; MCDONALD, G. How do I scope, shape and configure my research project. In: **SURVIVING and thriving in postgraduate research**. Singapore: Springer, 2019. p. 417-501. DOI: https://doi.org/10.1007/978-981-13-7747-1_12. Acesso em: 7 maio 2026.

KUHLTHAU, C. C. Inside the search process: information seeking from the user's perspective. **Journal of the American Society for Information Science**, v.42, n.5, p.361-371, 1991. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5<361::AI-D-ASL6>3.0.CO;2-%23](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5<361::AI-D-ASL6>3.0.CO;2-%23). Acesso em: 7 maio 2026.

LEI, G.; DOCHERTY, R.; COOPER, S. J. Materials science in the era of large language models: a perspective. **Digital Discovery**, v.3, n.7, p.1257, 2024. DOI: <https://doi.org/10.1039/d4dd00074a>. Acesso em: 7 maio 2026.

MILLER, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. **Psychological Review**, Washington, v. 63, n. 2, p. 81-97, 1956. DOI: <https://doi.org/10.1037/h0043158>. Acesso em: 7 maio 2026.

PRIEM, J.; PIWOWAR, H.; ORR, R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. **ArXiv**, 2022. DOI: <https://doi.org/10.48550/arXiv.2205.01833>. Acesso em: 9 jan. 2026.

SANTOS, R. A; REATEGUI, E. B. Uso de inteligência artificial generativa e análise de palavras-chave para apoiar o planejamento de projetos de pesquisa no ensino superior. **RELATEC: Revista Latinoamericana de Tecnología Educativa**, v.24, n.2, p.87-104, 2025. DOI: <https://doi.org/10.17398/1695-288X.24.2.87>.

VENKATESH, V.; BALA, H. Technology Acceptance Model 3 and a research agenda on interventions. **Decision Sciences**, v. 39, n. 2, p. 273-315, 2008. DOI: <https://doi.org/10.1111/j.1540-5915.2008.00192.x>. Acesso em: 9 jan. 2026.

VENKATESH, V.; MORRIS, M. G.; DAVIS, G. B.; DAVIS, F. D. User acceptance of information technology: toward a unified view. **MIS Quarterly**, v. 27, n. 3, p. 425-478, 2003. DOI: <https://doi.org/10.2307/30036540>. Acesso em: 9 jan. 2026.



10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



WILSON, T. D. Models of information behavior research. **Journal of Documentation**, v. 55, n. 3, p. 249-270, 1999. DOI: <https://doi.org/10.1108/EUM0000000007145>. Acesso em: 9 jan. 2026.