



Recebido: 28 de abril de 2025
Revisado: 1 de julho de 2025
Aceito: 17 de julho de 2025

Endereço dos autores:

¹ Programa de Pós-Graduação em
Informática na Educação (PPGIE) da
Universidade Federal do Rio
Grande do Sul (UFRGS). Av. Paulo
Gama s/n – Farroupilha, Porto
Alegre- RS, 90040-060 (Brasil)

E-mail / ORCID

rafael.antunes@ufrgs.br

 <https://orcid.org/0000-0002-1529-9063>

eliseoreategui@gmail.com

 <https://orcid.org/0000-0002-5025-9710>

ARTIGO / ARTICLE

Uso de inteligência artificial generativa e análise de palavras-chave para apoiar o planejamento de projetos de pesquisa no ensino superior

Using Generative Artificial Intelligence and Keyword Analysis to Support the Design of Research Projects in Higher Education

Rafael Antunes dos Santos¹ e Eliseo Berni Reategui²

Resumo: Em diferentes níveis da educação universitária, estudantes enfrentam desafios para identificar informações alinhadas a seus projetos de pesquisa, exigindo equilíbrio entre especificidade e abrangência na identificação de informações relevantes. Estudos mostram que habilidades metacognitivas, como o refinamento de palavras-chave e a navegação por categorias organizadas, favorecem a eficiência e a qualidade das buscas realizadas. Diante dessas dificuldades, este artigo investiga o uso combinado da análise de coocorrência de palavras e da Inteligência Artificial Generativa (IAGen) como estratégia de apoio ao delineamento do escopo de projetos de pesquisa no ensino superior. O estudo adotou uma metodologia qualitativa, envolvendo 23 estudantes de graduação e pós-graduação, que responderam a dois formulários intercalados pela entrega de um relatório personalizado. Esse relatório foi produzido a partir de buscas em bases bibliográficas, integradas à análise com IAGen, e abordou aspectos como sugestões de palavras-chave, estratégias de busca e um glossário. Os resultados indicaram que o procedimento instigou processos reflexivos nos estudantes, sendo que a maioria considerou a intervenção útil, especialmente para refinar a definição de palavras-chave. Também ficou evidente que a eficácia do método depende da clareza das informações iniciais fornecidas pelos participantes e do nível de maturidade dos projetos em desenvolvimento.

Palavras-chave: Análise de coocorrência de palavras, Inteligência artificial generativa, Escopo de pesquisa, Projeto de Pesquisa, Ensino superior, Ciência aberta.

Abstract: At different levels of higher education, students face challenges in identifying information aligned with their research projects, requiring a balance between specificity and breadth in locating relevant sources. Studies show that metacognitive skills, such as refining keywords and navigating through organized categories, enhance the efficiency and quality of information searches. In light of these challenges, this article investigates the combined use of word co-occurrence analysis and Generative Artificial Intelligence (GenAI) as a strategy to support the definition of research project scopes in higher education. The study employed a qualitative methodology, involving 23 undergraduate and graduate students who completed two questionnaires, interspersed with the delivery of a personalized report. This report was generated based on searches in bibliographic databases, integrated with GenAI analysis, and included keyword suggestions, search strategies, and a glossary. The results indicated that the procedure stimulated reflective processes in the students, with most considering the intervention useful, particularly for refining keyword definitions. It was also evident that the method's effectiveness depends on the clarity of the initial information provided by participants and the maturity level of their projects.

Keywords: Co-word analysis, Generative artificial intelligence, Research scope, Research project, Higher education, Open science.

1. Introdução

Em diferentes níveis de educação universitária, o delineamento do escopo de um projeto de pesquisa é uma atividade em que estudantes de graduação e de pós-graduação encontram inúmeros desafios para buscar e identificar informação bibliográfica relacionada aos seus projetos. Estes estudantes precisam equilibrar a harmonia entre a especificidade dos seus projetos de pesquisa e a sensibilidade para encontrar informações úteis que dialoguem com as suas proposições científicas. Estudos afirmam que alunos com melhores habilidades no refinamento de palavras-chave tendem a alcançar resultados de busca mais bem sucedidos em suas tarefas (Tu; Shih; Tsai, 2008). Práticas como navegação hierárquica baseada em localização e o uso de tags sociais são exemplos que também podem melhorar a eficácia das buscas (Liu et al., 2011). Tais estudos são significativos na circunstância em que consideram a importância das palavras-chave para o contexto de aprendizagem dos alunos, mesmo que estes resultados tenham sido alcançados com alunos de ensino médio em buscas na web.

Ferramentas tecnológicas para o delineamento do escopo de projetos tornam-se mais críticas com o aumento do volume e complexidade da literatura científica. Uma delas, a análise de coocorrência de palavras, identifica termos que frequentemente aparecem juntos, revelando associações conceituais (Callon et al., 1983). Esse método permite compreender relações semânticas e oferece uma base sólida para geração de palavras-chave e definição do escopo da pesquisa (Klarin, 2024). Já os Grandes Modelos de Linguagem (Large Language Models, LLMs) representam um avanço significativo, capazes de processar, gerar e estruturar informações complexas devido ao treinamento em vastos volumes textuais. No contexto científico, demonstram potencial para ampliar perspectivas analíticas e estimular pesquisadores ao sugerir conexões com áreas análogas (Peres et al., 2023). A combinação dessas tecnologias proporciona maior clareza conceitual, aprofundando o entendimento e ampliando possibilidades metodológicas nas pesquisas acadêmicas contemporâneas.

Diante desse contexto, este trabalho tem como objetivo investigar de que forma a avaliação automatizada de projetos de pesquisa acadêmica pode auxiliar estudantes na compreensão do escopo de suas propostas e na definição de palavras-chave que representem com precisão o campo de estudo. O crescente volume e a complexidade cada vez maior da literatura científica demandam abordagens que auxiliem os estudantes em uma experiência de aprendizado ativa. Nesse cenário, é essencial que eles se sintam capazes de relacionar conceitos, identificar padrões e explorar possibilidades de pesquisa de forma autônoma e crítica. Assim, o artigo propõe duas questões de investigação:

QP1. A análise de palavras-chave, associada à visualização gráfica de coocorrências, contribuiu para a reformulação da lista e para o aumento da segurança na escolha das palavras-chave?

QP2. Qual a percepção dos estudantes sobre a utilidade de repensar o conjunto de palavras-chave como forma de refinar o escopo de projetos de pesquisa?

Juntas, essas questões compõem uma estrutura analítica para compreender como a combinação da análise de coocorrência de palavras e da Inteligência Artificial

Generativa (IAGen) pode auxiliar os estudantes no refinamento iterativo de seus projetos de pesquisa. Tais dinâmicas podem ser analisadas à luz da Teoria do Comportamento de Busca e Uso da Informação que descreve como os usuários interagem com sistemas informacionais, selecionam e refinam termos de busca para encontrar conteúdos relevantes (Wilson, 1999).

Nesse processo, a formulação e o ajuste iterativo das consultas desempenham um papel essencial na construção do conhecimento, permitindo que os estudantes ampliem ou ajustem o escopo de sua investigação com base em novas descobertas e relações semânticas. O método aqui apresentado avalia se a integração da IAGen a dados extraídos da literatura científica podem, de fato, se consolidar como um recurso didático para o delineamento do escopo da pesquisa. Assim, pretende-se contribuir para a qualidade dos trabalhos e para a autonomia do processo de investigação. Como trabalhos correlatos ao apoio à projetos de pesquisa, cita-se as teses de Simone de Oliveira (2017) e de Ederson Bastiani (2022), apesar de não concentrarem-se na análise de palavras-chave.

As próximas seções trazem os fundamentos teóricos do delineamento, da análise de coocorrência de palavras e da IAGen, necessários ao entendimento do objetivo da pesquisa.

1.1. Delineamento do escopo de um projeto de pesquisa

A pesquisa acadêmica é essencial ao ensino superior, formando profissionais capacitados e contribuindo para inovação, avanço do conhecimento e desenvolvimento social (Schwartzman, 2022). Uma definição estruturada do escopo de pesquisa no contexto do comportamento informacional refere-se à “[...] delimitação sistemática dos limites conceituais, metodológicos e práticos de uma investigação, estabelecendo o que será incluído e excluído do estudo” (Wilson, 1999, p. 256-257). Kuhlthau (2004) complementa, destacando que o escopo evolui durante a busca informacional, sobretudo na fase de formulação, marcada por sentimentos iniciais de dúvida e confusão que progridem gradualmente para maior confiança conforme a informação é assimilada.

A definição adequada do escopo e a configuração de um projeto de pesquisa são essenciais, especialmente em níveis de pós-graduação. Os orientadores desempenham o papel de ajudar estudantes a moldar e ajustar seus projetos para que sejam factíveis e relevantes; no entanto, explorar abordagens metodológicas pode ser benéfico para encontrar um equilíbrio ideal (Cooksey; McDonald, 2019). Assim, propiciam-se oportunidades ao estudante, como definir os limites e a extensão do seu estudo (Mbuagbaw et al., 2020; Peters et al., 2020), identificar as variáveis relevantes a serem estudadas ou mapear conceitos de uma área de conhecimento (Pessini; Yamane; Siman, 2022), e fornecer uma estrutura para a tomada de decisões (Tricco et al., 2018).

De forma exploratória ou sistemática, o processo de preparação de revisões requer a construção de estratégias de busca em bases de dados para encontrar as melhores informações que ajudarão a formular o escopo dos projetos de pesquisa. Há estratégias metodológicas descritas na literatura para facilitar o desenvolvimento de modelos de buscas para revisões (Delaney; Tamás, 2017; Bramer et al, 2018; Grames et al., 2019), mas até onde se sabe, ações como as citadas acima são voltadas para um público profissional acostumado com os protocolos das revisões sistemáticas e

geralmente estas pessoas agregam maior conhecimento do vocabulário científico. Esse delineamento temático e semântico é essencial para formular buscas alinhadas aos interesses da pesquisa. Porém, devido ao aumento da complexidade dos dados científicos, ferramentas especializadas tornaram-se importantes para apoiar estudantes no desenvolvimento inicial da pesquisa.

O delineamento do escopo de pesquisa envolve definir objetivos e limites que orientam a investigação. Segundo Wilson (1981), as necessidades de informação emergem de contextos pessoais, sociais e ambientais, sendo fisiológicas, cognitivas ou afetivas, e impulsionam a busca informacional. Barreiras como fatores psicológicos, demográficos, sociais ou ambientais podem interferir nesse processo. A percepção de autoeficácia, por exemplo, influencia a motivação para buscar informações, enquanto o acesso limitado a recursos pode restringi-la. Compreender o contexto dos estudantes é essencial para apoiar suas buscas. Dervin (1998) destaca que os usuários enfrentam lacunas de conhecimento e usam a informação para “construir sentido”. O modelo de Ellis (1989) descreve estratégias como iniciar, encadear, navegar, diferenciar, monitorar e extrair como formas de localizar e filtrar informações relevantes.

Ao delinear o escopo de pesquisa, é crucial considerar variáveis intervenientes, necessidades informacionais, lacunas e estratégias de uso. A análise de coocorrência e a IAGen podem contribuir para o mapeamento do foco de estudo. As próximas seções exploram essas abordagens e seus potenciais benefícios.

1.2. Coocorrência de palavras-chave

As palavras-chave fornecidas pelos autores são elementos essenciais em artigos científicos, amplamente utilizadas para otimizar a recuperação de textos em bases bibliográficas. Geralmente compostas por três a cinco termos, podem aparecer literal ou variavelmente no texto, representando conceitos que extrapolam o conteúdo direto do trabalho (Pezik et al., 2023).

A coocorrência de palavras, conceito-chave na bibliometria, refere-se à frequência com que dois ou mais termos ocorrem conjuntamente em um corpus textual, como artigos científicos (Zupic; Cater, 2015). Essa técnica permite mapear relações semânticas e identificar padrões conceituais, oferecendo uma visão abrangente das estruturas temáticas de um campo (Callon et al., 1983). Na bibliometria, tem sido amplamente usada para construir mapas científicos e hierarquias temáticas, auxiliando na compreensão das conexões entre áreas de pesquisa e no rastreamento de tendências emergentes (He, 1999; Zawacki-Richter; Latchem, 2022).

Além do mapeamento temático, a análise de coocorrência desempenha papel central na identificação e no refinamento de palavras-chave, essenciais à busca e recuperação de informação (Zupic; Cater, 2015). Sua aplicação em projetos de pesquisa contribui para delimitar escopos com maior clareza e alinhamento ao debate científico atual. Recentemente, a integração com Grandes Modelos de Linguagem (LLMs), aplicados em tarefas de geração de linguagem (um tipo de IAGen), potencializa essa técnica ao detectar relações semânticas complexas com alta precisão (Klarin, 2024). Essa sinergia representa uma abordagem promissora para estudantes e pesquisadores no aperfeiçoamento das estratégias de busca e definição de escopo.

1.3. IA generativa para extração de informações

Grandes Modelos de Linguagem (LLMs) tornaram-se ferramentas poderosas para tarefas de Processamento de Linguagem Natural (PLN), como Extração de Informações (IE) e seleção de palavras-chave (Han et al., 2024). Treinados com grandes volumes de dados, compreendem e geram linguagem de forma semelhante à humana, sendo amplamente adotados no meio acadêmico (Chang et al., 2024). Sua principal vantagem é lidar com requisitos ambíguos e complexos em diferentes domínios (Lei et al., 2024). Quando utilizados para criar novos textos, resumos, respostas ou sugestões, eles se inserem no campo da IAGen.

Apesar do bom desempenho em tarefas sem treinamento específico (zero-shot ou few-shot), os LLMs tendem a ter resultados inferiores em IE tradicional, quando comparados a modelos ajustados como os baseados em BERT (Bidirectional Encoder Representations from Transformers). No entanto, destacam-se na Extração Aberta de Informações (OpenIE), por permitir a identificação de entidades e relações sem categorias fixas, sendo adaptáveis a diferentes contextos (Xu et al., 2024).

No e-commerce, Maragheh et al. (2023) propuseram o LLM-TAKE, que melhora a seleção de palavras-chave ao filtrar termos não informativos ou alucinados, superando benchmarks em dados reais. Suas limitações incluem dependência de dados de referência, alto custo computacional e riscos de alucinação.

Para contextos científicos, Dagdelen et al. (2024) desenvolveram um método para Extração de Entidades e Relações (NERRE), capaz de gerar saídas em linguagem natural e em JSON. Destaca-se por sua usabilidade e normalização integrada de entidades, mas enfrenta dificuldades com formatação e geração de informações inexistentes.

Lee et al. (2023) enfrentaram a ausência de palavras-chave em artigos de conferências em IA, propondo o uso do modelo Galactica, da Meta, para geração automática. Com base em métricas de similaridade, o estudo mostrou que 42,7% das palavras-chave geradas coincidiram com as definidas pelos autores, superando abordagens anteriores. Modelos baseados em GPT (Generative Pre-trained Transformer) e RAG (Retrieval-Augmented Generation) também demonstram competência na geração de ontologias e identificação de relações semânticas (Pisu et al., 2024). Já estruturas com BERT oferecem precisão na criação de hierarquias de tópicos e tarefas como indexação de pesquisas específicas (Yang et al., 2023).

Esses estudos evidenciam o potencial dos LLMs na análise de tendências e no suporte ao delineamento de escopos, embora desafios como relevância e aplicabilidade de palavras-chave persistam. A seção seguinte apresenta os procedimentos metodológicos adotados neste estudo, detalhando estratégias, instrumentos e abordagem de análise.

2. Método

As decisões metodológicas adotadas na pesquisa implicaram em um estudo de tipologia exploratório-descritiva e abordagem qualitativa. Define-se como objeto o delineamento da pesquisa acadêmica de estudantes de graduação e de pós-

graduação. A aproximação e o convite à participação dos estudantes ocorreu através do contato prévio com professores de disciplinas cursadas no ensino superior. Na graduação, as disciplinas envolviam metodologia de pesquisa e visavam a preparação do projeto de pesquisa para o trabalho de conclusão dos cursos de Pedagogia e de licenciaturas diversas. Na pós-graduação, os alunos também cursavam uma disciplina de metodologia de pesquisa. O convite foi realizado enfatizando seu caráter voluntário, com ênfase na garantia de sigilo e na preservação da privacidade dos dados dos alunos. A amostra do estudo foi composta por 23 estudantes selecionados de forma “[...] intencional por critérios pré- determinados” (Fragoso; Recuero; Amaral, 2012, p. 80) ao projeto de pesquisa.

O desenho da pesquisa envolveu um fluxo de coleta de dados que compreendeu a solicitação ao preenchimento de dois formulários intercalados ao recebimento via e-mail de um relatório sobre o escopo da pesquisa de cada participante. Para embasar sua elaboração, o primeiro formulário (F1) solicitou a declaração dos temas e as questões de pesquisa junto a um pequeno repertório de palavras-chave iniciais, como descrito na Tabela 1.

Tabela 1. Perguntas do formulário F1

Pergunta	Formato
F1.1. Qual o tema do seu projeto de pesquisa?	Resposta aberta
F1.2. Qual sua questão de pesquisa?	Resposta aberta
F1.3. Quais são as palavras-chave relacionadas ao seu projeto de pesquisa.	Resposta aberta
F1.4. Qual o nível de segurança em relação às palavras-chave que você definiu?	Escala Likert: totalmente seguro; seguro; neutro; inseguro; totalmente inseguro
F1.5. Quais bases de dados você utiliza para encontrar bibliografia?	Múltipla escolha: Google; Google Scholar; Scielo; Portal CAPES; Repositório; Biblioteca; Outra

Além dessas perguntas iniciais, o formulário também investigou o nível de segurança dos estudantes na escolha das palavras-chave e as bases de dados utilizadas em seus processos de busca, com o objetivo de compreender melhor as práticas de pesquisa dos participantes.

As informações coletadas servem como ponto de partida para a análise das propostas preliminares do projeto, visando à construção de um relatório individual. Após o preenchimento do formulário, o ChatGPT4o foi usado para elaborar um feedback para cada aluno, com a execução de quatro prompts:

Prompt 1. Comentário ao tema e ao repertório inicial de palavras-chave:

Preciso fazer a avaliação de um projeto de pesquisa. O tema da pesquisa é: [...]. A questão de pesquisa é: [...] As palavras-chave inicialmente definidas são: [...]. Comente a adequação das palavras-chave à “questão de pesquisa” e ao “tema de pesquisa”, mas sem citá-las. Caso haja problemas, sugerir algum ajuste e recomendar ao aluno que converse com seu orientador a respeito.

Prompt 2. Sugestão de palavras-chave alternativas para a pesquisa:

Defina um conjunto de termos para melhor delimitar o projeto de pesquisa. Não mencione as palavras-chave indicadas pelo projeto de pesquisa. Faça um delineamento inicial para o projeto, mostrando outras palavras-chave relacionadas que podem ajudar a melhor compreender o contexto da pesquisa. Considere informar um limite de apenas 6 palavras-chave em português, a tradução do termo para o inglês e uma descrição breve do significado em português.

Prompt 3. Elaboração de uma chave de busca simplificada:

Faça uma chave de busca simplificada que combine termos em inglês para ser utilizada na Scopus e buscar trabalhos na área. Combine termos do projeto e termos sugeridos pelo prompt. Apresente uma descrição do objetivo da chave de busca oferecida.

Prompt 4. Elaboração do glossário com as palavras-chave do grafo:

Extraia as palavras desse grafo e elabore um pequeno glossário mantendo a entrada do termo em inglês, a tradução para português e uma descrição breve do seu significado.

O feedback personalizado configura, então, uma análise inicial do delineamento de sua pesquisa, como ilustrado na Figura 1. Na porção A da imagem (rótulo à esquerda), observam-se os dados originalmente informados pelo aluno. Com base nestes dados, o relatório apresentava um breve parecer sobre a escolha das palavras-chave e sugeria ampliá-las com um conjunto complementar, trazendo uma definição e breve comentário sobre sua relevância (porção B da imagem). Nesta etapa, a IAGen foi utilizada para apoiar a elaboração do relatório. Primeiro, foi feita uma avaliação do tema de pesquisa e das palavras-chave inicialmente definidas, com comentários sobre sua adequação e sugestões de ajustes. Em seguida, foram sugeridas novas palavras-chave, acompanhadas de breves descrições, para ampliar e qualificar o delineamento do projeto. Também foi construída uma chave de busca simplificada, combinando termos relevantes para pesquisas em bases de dados, com uma breve explicação de sua finalidade.

Para compor a porção C da imagem, foi executada busca avançada na base Scopus inserindo-se a string definida na etapa anterior. A partir disso, recuperaram-se os metadados dos resultados obtidos em formato BibTeX, ou seja, as palavras-chave atribuídas pelos autores no conjunto de trabalhos retornados, a partir do campo DE (Descriptor). Para garantir uma análise eficiente, definiu-se como limite máximo os 500 primeiros itens resultantes da busca.

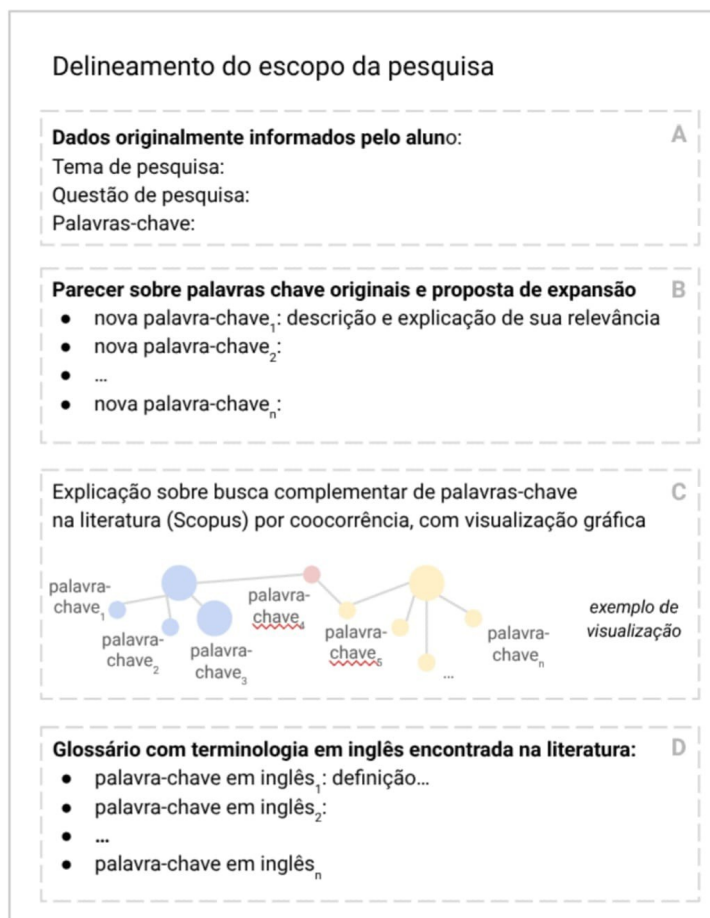


Figura 1. Modelo de relatório.

Na sequência, procedeu-se à análise bibliométrica propriamente dita, empregando-se o pacote Bibliometrix (Aria e Cuccurullo, 2017) do software RStudio. Nesta etapa, utilizaram-se funções de coocorrência de palavras-chave específicas do Bibliometrix, associadas ao algoritmo Fruchterman-Reingold, reconhecido pela sua capacidade de organizar os nós (palavras-chave) em agrupamentos (clusters), conforme a Figura 2.

```
library(bibliometrix)
df.scopus <- convert2df("local/scopus.bib",
                        dbsource = "scopus",
                        format = "bibtex")
NetMatrix <- biblioNetwork(df.scopus, analysis = "co-occurrences",
                           network = "author_keywords", sep = ";")
net=networkPlot(NetMatrix, normalize="association", weighted=T,
                n = 15, Title = "Co-Word Graph",
                type = "fruchterman", size=T,
                edgesize = 5, labels=1.4)
```

Figura 2. Código usado no Bibliometrix.

Como resultado desse processo, obtiveram-se grafos de coocorrência de palavras-chave que ilustravam as associações semânticas mais relevantes no contexto dos temas pesquisados pelos estudantes. A Figura 3 traz um exemplo fictício de grafo de coocorrência de palavras gerado para um projeto de pesquisa em Educação relacionado ao tema “jogos como estratégia de aprendizagem”.

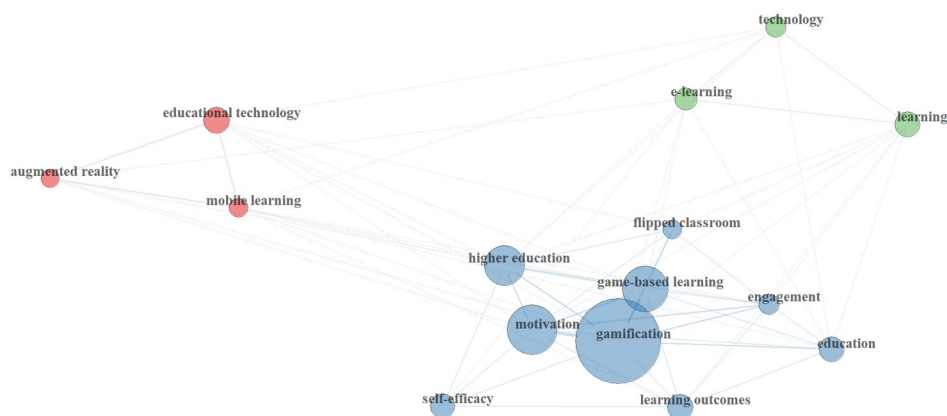


Figura 3. Grafo de coocorrência.

A Figura 3 mostra uma rede de coocorrência com três agrupamentos temáticos: tecnologia educacional (vermelho), aprendizagem digital (verde) e aprendizagem baseada em jogos (azul). Ao fornecer aos estudantes uma visão complementar dos conceitos-chave relacionados ao tema de sua pesquisa, buscava-se contribuir no refinamento do escopo e no aprimoramento das estratégias de busca bibliográfica adotadas, auxiliando-os na identificação de conexões entre palavras-chave relevantes. As linhas entre os nós indicam frequência de coocorrência entre palavras-chave. O tamanho de cada nó indica a frequência com que a palavra-chave aparece no corpus. Por exemplo, o grafo revela que o conceito gamification é o mais recorrente e central, associado a estratégias educacionais inovadoras no ensino superior. Há uma separação clara entre clusters ligados à inovação metodológica, tecnologia aplicada e fundamentos da aprendizagem digital, sugerindo diversidade temática e pontos de convergência relevantes para o delineamento de projetos de pesquisa. Após a geração da visualização das coocorrências de palavras, a imagem resultante era salva em formato PNG e adicionada ao relatório do aluno.

Para compor a porção D do relatório (Figura 1), utilizou-se novamente a IAGen para elaborar um glossário explicativo dos termos presentes no grafo, com o objetivo de apoiar a construção de estratégias de busca e o refinamento do escopo do projeto.

Após a finalização do relatório, os estudantes receberam uma cópia para apreciação e foram convidados a responder a um segundo formulário (F2), com o objetivo de avaliar em que medida a análise apresentada contribuiu para o refinamento do escopo da pesquisa inicialmente delineada. A Tabela 2 apresenta as questões e os formatos de resposta utilizados.

Tabela 2. Perguntas do segundo formulário (F2)

Pergunta	Formato
F2.1. O relatório sobre o delineamento do projeto lhe fez repensar a lista de palavras-chave?	Ponderação simples: sim, não e neutro
F2.2. Existe uma estimativa numérica de palavras-chave dispostas no grafo que lhe pareceram interessar para uso em buscas bibliográficas?	Escala ordinal por faixas numéricas: todas; de 9 a 12; de 5 a 8; de 1-4; nenhuma
F2.3. Após a leitura do relatório, qual seu nível de segurança em relação às palavras-chave definidas no projeto?	Escala Likert: totalmente seguro; seguro; neutro; inseguro; totalmente inseguro
F2.4. A visualização do grafo de palavras-chave auxiliou a compreender o escopo do projeto de pesquisa?	Escala Likert: auxiliou bastante; auxiliou; neutro; não auxiliou; não auxiliou em nada
F2.5. Qual sua percepção quanto à utilidade do delineamento do escopo para a compreensão do tema de pesquisa a partir do relatório recebido?	Escala Likert: extremamente útil; útil; neutro; inútil; extremamente inútil
F2.6. Comente impressões positivas ou negativas em relação ao delineamento das palavras-chave do projeto.	Resposta aberta

As perguntas do formulário F2 buscaram conhecer a percepção dos estudantes sobre a utilidade do feedback recebido e a reconsideração das palavras-chave iniciais. Para as questões iniciais de base quantitativa, foram utilizadas técnicas de estatística descritiva, considerando as frequências absolutas e relativas dos dados coletados nos dois formulários. Essas análises permitiram uma compreensão inicial sobre o nível de segurança dos participantes na escolha de palavras-chave, a utilidade percebida na ampliação do escopo de pesquisa e a quantidade estimada de palavras-chave consideradas relevantes.

Para a análise qualitativa, adotou-se a técnica de análise de conteúdo proposta por Bardin (2016) e por Saldaña (2015), visando sistematizar e interpretar as respostas abertas dos estudantes em torno das duas questões de pesquisa propostas neste artigo.

3. Resultados

A pesquisa investigou as percepções de estudantes do ensino superior sobre a experiência de refinar o escopo de seus projetos de pesquisa a partir de um procedimento específico. Esse procedimento combinou a análise de coocorrência de palavras em buscas bibliográficas com o uso de IAGen para examinar elementos do texto e gerar feedback personalizado.

Para responder à primeira questão de pesquisa (QP1), foi analisado como a avaliação da lista de palavras-chave, associada à visualização gráfica de coocorrências, contribuiu para a reformulação da lista e para o aumento da segurança na escolha das palavras-chave. Para isso, foram consideradas as respostas às perguntas F2.1, F2.2 e F2.3 do formulário.

A primeira delas (F2.1) mostrou que 82,6% dos estudantes (19) afirmaram considerar a modificação de sua lista de palavras-chave após a leitura do relatório. Dois

(8,7%) mostraram-se incertos, e outros dois indicaram não pretender alterá-la. Em relação ao uso estimado das palavras do grafo de coocorrência (F2.2), nove participantes indicaram de 5 a 8 termos relevantes e oito de 1 a 4; três disseram que todas foram úteis, e dois apontaram de 9 a 12 palavras-chave significativas. Ninguém considerou todas irrelevantes. Esses dados indicam que, ainda que nem todos tenham adotado integralmente os termos sugeridos, a visualização contribuiu para ampliar o repertório e favorecer o refinamento das estratégias de busca.

Os dados obtidos do segundo formulário indicaram que a maior parte dos alunos (18, 78,3%) declarou-se segura quanto à escolha das fontes para realizar suas buscas; 3 (13%) consideraram-se totalmente seguros, enquanto apenas 2 (8,7%) adotaram uma postura neutra, sem registros de insegurança entre os participantes. Ao analisar a questão F2.3 do segundo formulário, que investigava o nível de segurança dos estudantes em relação à seleção de palavras-chave após o recebimento do relatório, observou-se que, embora as frequências absolutas gerais tenham permanecido estáveis (Figura 4), individualmente ocorreram movimentações significativas.

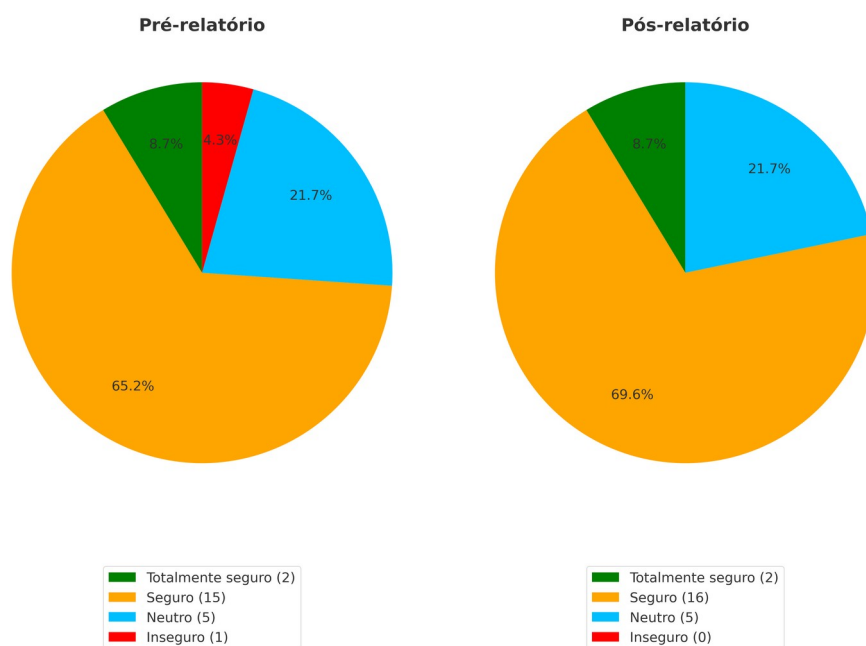


Figura 4. Nível de segurança na seleção de palavras-chaves.

As contagens absolutas variaram pouco; por exemplo, o número de estudantes que se classificaram como "seguros" apresentou um leve aumento. No entanto, o que mais chamou atenção foram as mudanças internas. Ou seja, os estudantes que, após o recebimento do relatório, alteraram sua percepção de segurança, seja ampliando, seja reduzindo seu nível de confiança (Figura 5).

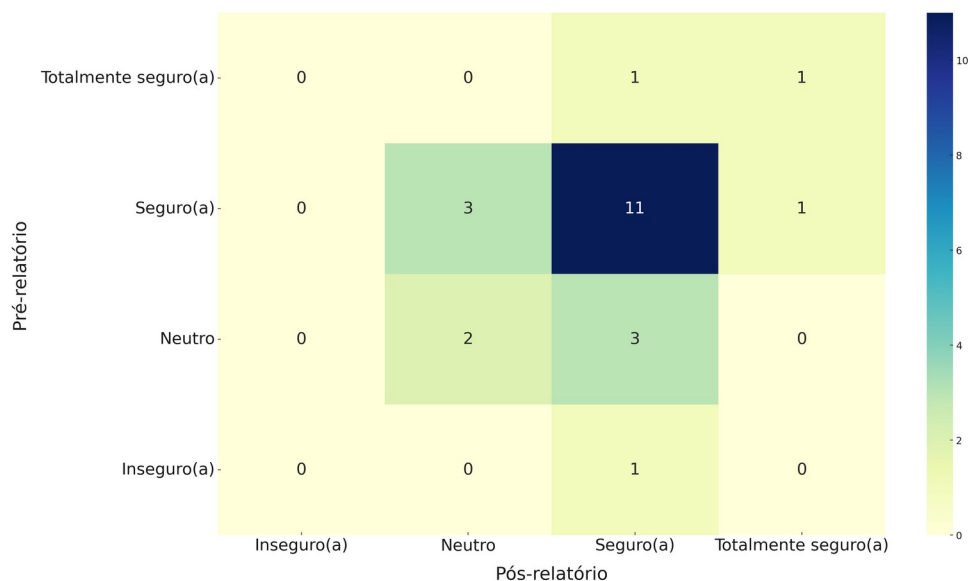


Figura 5. Mudança no nível de segurança (pré e pós-relatório).

A matriz de transição da Figura 5 mostra que, embora a maioria dos alunos tenha mantido seu posicionamento (11 permaneceram seguros), 10 dos 23 participantes alteraram sua autopercepção: três migraram de “neutro” para “seguro”, um de “inseguro” para “seguro” e dois que avançaram de “seguros” para “totalmente seguros”, indicando fortalecimento da autoconfiança após o relatório. Em sentido oposto, três estudantes “seguros” se reposicionaram como “neutros” e um “totalmente seguro” passou a se considerar apenas “seguro”. Essas movimentações, tanto de fortalecimento quanto de recuo na autopercepção, revelam a ação reflexiva estimulada pelo relatório, sinalizando que a intervenção promoveu uma revisão crítica das escolhas anteriores. Mesmo nas transições que sugerem redução no nível de segurança, o movimento pode ser interpretado como um refinamento na avaliação pessoal do projeto.

A análise dos comentários abertos no formulário (F2.5) reforçou a percepção de que a intervenção contribuiu para aumentar a segurança dos estudantes na escolha das palavras-chave. O relato do aluno 11, por exemplo, aponta que “foi possível repensar com mais clareza sobre o tema e as palavras-chave”, sugerindo um fortalecimento da confiança após o relatório. Outros estudantes, como o aluno 8, destacaram a facilidade de modificar a string de busca com base nas sugestões automatizadas, o que também pode ser interpretado como um indicativo de maior segurança instrumental no processo de refinamento.

Alguns estudantes, no entanto, registraram observações que apontam limitações. O aluno 16, por exemplo, mencionou que as palavras-chave sugeridas não estavam totalmente adequadas à sua proposta. Neste sentido, é importante reconhecer que a aplicação de IAGen para ampliar o conjunto de palavras-chave e a utilização de coocorrência na literatura possuem limitações, podendo eventualmente sugerir termos que não correspondam plenamente à percepção do aluno sobre a direção desejada para seu projeto. Refinamentos nos critérios de localização e sugestão

de termos podem contribuir para aumentar a precisão do sistema e reduzir essas discrepâncias.

Mesmo com essas fragilidades, em conjunto, os resultados indicam que a avaliação personalizada e a visualização gráfica de palavras-chave contribuíram positivamente para a ampliação do repertório de busca, a reformulação da lista de palavras-chave e o fortalecimento, ou o refinamento crítico, da segurança dos estudantes em relação às escolhas realizadas em seus projetos de pesquisa.

Com relação à segunda questão de pesquisa (QP2), que avaliava a percepção dos estudantes sobre a utilidade de repensar o conjunto de palavras-chave para o refinamento do escopo dos projetos de pesquisa, foram consideradas as respostas às perguntas F2.3 e F2.4 do formulário.

A percepção dos participantes sobre a visualização do grafo de coocorrência de palavras-chave para a compreensão do escopo da pesquisa (F2.3) foi, em sua maioria, positiva. Embora "Neutro" tenha sido a resposta mais frequente (9 participantes), 13 estudantes (56,5%) afirmaram que a visualização do grafo "auxiliou" (7) ou "auxiliou bastante" (6) na compreensão do escopo, indicando que a maioria reconheceu efetivamente a contribuição da ferramenta para o refinamento do projeto. Apenas um aluno afirmou que o grafo "não auxiliou" e nenhum respondeu que "não auxiliou em nada". Esses resultados indicam que, mesmo sem impacto direto para todos os estudantes, a proposta visual contribuiu para estimular reflexões e reforçar decisões já tomadas, evidenciando um efeito positivo sobre a compreensão dos contornos temáticos dos projetos.

Em resposta à pergunta sobre a percepção dos estudantes quanto à utilidade do relatório recebido para a compreensão do tema de pesquisa (F2.4), a avaliação foi amplamente positiva: 91,3% dos estudantes consideraram o delineamento útil (34,8%) ou extremamente útil (56,5%), enquanto apenas dois participantes (8,7%) adotaram uma postura neutra. Não houve registros de percepção de inutilidade.

A análise das respostas à questão aberta do formulário (F2.6) permitiu reforçar aspectos positivos já identificados em relação à contribuição do procedimento para a compreensão do tema de pesquisa e o delineamento do projeto. Afirmações como "foi bastante positivo o delineamento" (aluno 9), "ajudou no desenvolvimento da pesquisa" (aluno 17) e "me ajudaram a perceber tópicos mais refinados" (aluno 23) indicam uma recepção favorável. Esses relatos reforçam a percepção de que a atividade contribuiu para uma visão mais clara e objetiva dos projetos. O comentário do aluno 1 exemplifica esse movimento ao afirmar: "com o delineamento sinto que tenho uma visão melhor do panorama geral".

Embora a maioria dos comentários tenha sido positiva nesse sentido, algumas limitações também foram apontadas pelos participantes. O aluno 15, por exemplo, descreveu o processo de delineamento como "um pouco complicado", e o aluno 9 reconheceu a abrangência positiva das palavras-chave, mas ressaltou que isso também o afastou um pouco da identificação do foco central do projeto. Esses relatos evidenciam que o refinamento do escopo é um processo dinâmico, que exige equilíbrio entre ampliação e precisão. Apesar dessas dificuldades pontuais, o conjunto das respostas indica que a proposta foi eficaz ao promover a reflexão e ampliar a consciência dos estudantes sobre os contornos de seus temas. A elevada taxa de

percepção positiva sugere que a intervenção não apenas auxiliou na ampliação do repertório temático, mas também contribuiu para fortalecer a compreensão crítica dos estudantes sobre a delimitação de seus objetos de estudo.

4. Discussão

A análise dos resultados evidenciou uma percepção majoritariamente positiva dos participantes, indicando o potencial formativo dessas tecnologias, especialmente quando integradas a procedimentos reflexivos, como a seleção e revisão de palavras-chave. A visualização do grafo de coocorrências contribuiu para que os estudantes reconsiderassem a compreensão do escopo de seus projetos e reformulassem suas palavras-chave, reforçando a relação entre habilidades metacognitivas e competência informacional, como apontam Tu, Shih e Tsai (2008).

A aplicação de IAGen no contexto educacional, conforme delineado neste estudo, configurou um espaço de interação semi-orientada que favoreceu a reflexão crítica dos estudantes sobre o feedback recebido. Mais do que uma ferramenta de resposta automática, a IAGen atuou como mediadora, oferecendo insumos que exigiram interpretação ativa e ajustes conscientes por parte dos alunos, o que contribuiu para o desenvolvimento da autonomia no processo de delineamento dos projetos. No que se refere à segurança na definição de palavras-chave, a aparente estabilidade nos dados agregados mascarou movimentações individuais significativas: estudantes inicialmente inseguros tornaram-se mais confiantes, enquanto outros, mais seguros, adotaram posições neutras após a intervenção. Essas transições sugerem um processo genuíno de revisão crítica e refinamento da compreensão do próprio escopo de pesquisa, em consonância com a teoria de Kuhlthau (2004), que descreve a formulação de uma pesquisa como uma etapa marcada por instabilidade e reconstrução progressiva do entendimento.

Em consonância com estudos como o de Haman e Školník (2023), este trabalho evidencia que a IAGen pode contribuir para reduzir a carga cognitiva no planejamento de pesquisa ao apoiar a organização e a estruturação do repertório conceitual, facilitando a visualização de conexões relevantes entre termos. No presente estudo, embora a maioria dos estudantes tenha considerado entre 5 e 8 palavras do grafo como úteis, alguns reconheceram quase todas, indicando variações no grau de alinhamento entre a representação automatizada e o projeto individual — o que pode estar relacionado à maturidade do projeto ou à clareza inicial do tema de pesquisa. Assim, mais do que simplesmente expandir o número de palavras-chave, a intervenção mostrou-se eficaz ao oferecer suporte para estruturar e refinar o pensamento dos estudantes sobre seus temas.

As percepções críticas levantadas pelos estudantes também são relevantes, especialmente no que diz respeito ao distanciamento observado entre algumas sugestões de palavras-chave e o foco específico de seus projetos. Esse tipo de limitação confirma apontamentos já feitos por Peres et al. (2023), que destacam a dependência da qualidade dos dados de entrada e a necessidade de interpretação contextual como desafios ainda presentes para a aplicação efetiva da IAGen em ambientes educacionais. Esses resultados reforçam que o uso de tecnologias baseadas em IA generativa não pode prescindir de mediação crítica, tanto por parte dos estudantes quanto dos

docentes, a fim de assegurar que as sugestões sejam adequadamente interpretadas e ajustadas aos objetivos de cada pesquisa.

Os resultados também suscitam reflexões sobre a inserção de tecnologias como a análise de coocorrência de palavras e a IAGen nos currículos de formação científica. Integradas de forma sistemática, essas ferramentas podem contribuir para o desenvolvimento de habilidades de busca, organização e avaliação crítica da informação, alinhando-se à teoria do comportamento informacional de Wilson (1999) e aos princípios da aprendizagem significativa. No entanto, sua adoção exige mais do que a simples introdução de novas ferramentas: requer estratégias pedagógicas que incentivem o uso reflexivo e crítico desses recursos. Como destaca Klarin (2024), a abordagem iterativa e exploratória promovida por essas tecnologias é particularmente valiosa nos estágios iniciais de investigação, quando os estudantes ainda estão construindo a compreensão de seus temas. Ainda assim, seu impacto dependerá da capacidade dos currículos de integrar essas práticas de modo contextualizado, evitando o risco de uma adoção meramente instrumental.

5. Conclusão

Esta pesquisa evidenciou que a combinação entre análise de palavras-chave, visualização gráfica de coocorrências e feedback personalizado mediado por IAGen pode ser uma estratégia eficaz para apoiar estudantes no refinamento do escopo de projetos de pesquisa. A principal contribuição da proposta foi favorecer a auto-percepção dos estudantes na reflexão crítica sobre a delimitação temática, na ampliação do repertório conceitual e no fortalecimento da segurança dos participantes na definição de palavras-chave, elementos essenciais para o desenvolvimento de pesquisas mais estruturadas e fundamentadas.

Os resultados deste estudo também indicaram que a efetividade da proposta aqui descrita depende fortemente da qualidade das informações iniciais fornecidas pelos estudantes e do grau de maturidade de seus projetos. Esses fatores influenciam diretamente a relevância e a aplicabilidade das sugestões recebidas. Assim, reforça-se a necessidade de uma supervisão pedagógica ativa, capaz de orientar o uso crítico dessas ferramentas e de contextualizar os resultados no percurso formativo de cada aluno.

Em conclusão, este trabalho demonstra que a integração entre ciência da informação e IA abre novas possibilidades metodológicas para apoiar estudantes e pesquisadores no enfrentamento da crescente complexidade da literatura científica. Ao estruturar estratégias como a análise de coocorrência de palavras e a geração de feedback personalizado mediado por IA, torna-se possível não apenas organizar grandes volumes de informação, mas também estimular processos de refinamento conceitual e fortalecimento da competência informacional.

Uma limitação do estudo refere-se ao tamanho e à composição da amostra, formada por 23 estudantes de graduação e pós-graduação, com projetos em diferentes níveis de maturidade. Essa heterogeneidade pode ter influenciado as percepções sobre a utilidade das ferramentas aplicadas, especialmente em relação à clareza dos temas e à capacidade de aproveitamento das sugestões oferecidas. Além disso, a intervenção foi realizada em um único momento do processo de delineamento dos projetos, o que

restringe a análise dos efeitos de longo prazo no desenvolvimento das pesquisas. Estudos futuros poderão considerar o acompanhamento contínuo dos estudantes para avaliar de forma mais abrangente o impacto dessas tecnologias ao longo das diferentes etapas da pesquisa acadêmica.

6. Referências

- Bardin, L. (2016). *Análise de conteúdo*. São Paulo: Edições 70.
- Bastiani, E. (2022). *Desenvolvimento e validação de uma ferramenta computacional apoiada por mineração de texto voltada à qualificação da escrita de projetos de pesquisa* (Tese de doutorado, Programa de Pós-Graduação em Informática na Educação, Universidade Federal do Rio Grande do Sul).
- Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L., & Bekhuis, T. (2018). Evaluation of a new method for librarian-mediated literature searches for systematic reviews. *Research Synthesis Methods*, 9(4), 510–520. <https://doi.org/10.1002/jrsm.1279>
- Brasil. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. (2025). *Portal de Periódicos da CAPES*. <https://www.periodicos.capes.gov.br/>
- Callon, M., Courtial, J.-P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22(2), 191–235. <https://doi.org/10.1177/053901883022002003>
- Chang, Y., Xu, C., Wang, W., & Yan, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3), 1. <https://doi.org/10.1145/3641289>
- Cooksey, R., & McDonald, G. (2019). How do I scope, shape and configure my research project? In *Surviving and thriving in postgraduate research* (pp. 417–501). Springer. https://doi.org/10.1007/978-981-13-7747-1_12
- Dagdelen, J., Kononova, O., Horton, M., & Persson, K. A. (2024). Structured information extraction from scientific text with large language models. *Nature Communications*, 15, 1418. <https://doi.org/10.1038/s41467-024-45563-x>
- Delaney, A., & Tamás, P. (2017). Searching for evidence or approval? A commentary on database search in systematic reviews and alternative information retrieval methodologies. *Research Synthesis Methods*, 9(1), 124–131. <https://doi.org/10.1002/jrsm.1282>
- Dervin, B. (1998). Sense-making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, 2(2), 36–46. <https://doi.org/10.1108/13673279810249369>
- Ellis, D. (1989). A behavioural approach to information retrieval system design. *Journal of Documentation*, 45(3), 171–212. <https://doi.org/10.1108/eb026843>
- Fragoso, S., Recuero, R., Amaral, A. (2012). *Métodos de pesquisa na internet*. Porto Alegre: Sulina.
- Grames, E. M., Stillman, A. N., Tingley, M. W., & Elphick, C. S. (2019). An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. *Methods in Ecology and Evolution*, 10(10), 1645–1654. <https://doi.org/10.1111/2041-210x.13268>
- Han, S., Zhang, C., Wang, J., & Wang, X. (2024). A review of large language models: Fundamental architectures, key technological evolutions, interdisciplinary technologies integration, optimization and compression techniques, applications, and challenges. *Electronics*, 13(24), 5040. <https://doi.org/10.3390/electronics13245040>

- Klarin, A. (2024). How to conduct a bibliometric content analysis: Guidelines and contributions of content co-occurrence or co-word literature reviews. *International Journal of Consumer Studies*, 48(2), e13031. <https://doi.org/10.1111/ijcs.13031>
- Kuhlthau, C. C. (2004). *Seeking meaning: A process approach to library and information services* (2nd ed.). Libraries Unlimited.
- Lee, W., Park, K., Kim, S., & Kim, Y. (2023). Toward keyword generation through large language models. In *Proceedings of the International Conference on Intelligent User Interfaces* (pp. 37–40). ACM. <https://doi.org/10.1145/3581754.3584126>
- Lei, G., Docherty, R., & Cooper, S. J. (2024). Materials science in the era of large language models: A perspective. *Digital Discovery*, 3(7), 1257. <https://doi.org/10.1039/d4dd00074a>
- Liu, M. C., Chang, H. C., Wu, J. S., & Huang, Y. C. (2013). Fostering learner's metacognitive skills of keyword reformulation in image seeking by location-based hierarchical navigation. *Educational Technology Research and Development*, 61, 233–254. <https://doi.org/10.1007/s11423-012-9280-3>
- Maragheh, R. Y., Amini, M., & Aghajani, M. (2023). LLM-TAKE: Theme-aware keyword extraction using large language models. In *Proceedings of the International Conference on Big Data* (pp. 4318–4324). IEEE. <https://doi.org/10.1109/bigdata59044.2023.10386476>
- Mbuagbaw, L., Rochweg, B., Jaeschke, R., & Guyatt, G. (2020). A tutorial on methodological studies: The what, when, how and why. *BMC Medical Research Methodology*, 20(226). <https://doi.org/10.1186/s12874-020-01107-7>
- Oliveira, S. (2017). *O apoio de uma ferramenta com base na mineração de texto para a escrita acadêmica* (Tese de doutorado, Programa de Pós-Graduação em Informática na Educação, Universidade Federal do Rio Grande do Sul).
- Peres, R., Matzler, K., Schlegelmilch, B. B., & Vom Brocke, J. (2023). On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice. *International Journal of Research in Marketing*, 40(2), 269–275. <https://doi.org/10.1016/j.ijresmar.2023.03.001>
- Pessin, V. Z., Yamane, L. H., & Siman, R. R. (2022). Smart bibliometrics: An integrated method of science mapping and bibliometric analysis. *Scientometrics*, 127, 3695–3718. <https://doi.org/10.1007/s11192-022-04406-6>
- Peters, M. D. J., Godfrey, C. M., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis*, 18(10), 2119–2126. <https://doi.org/10.11124/jbies-20-00167>
- Pęzik, P., Mikołajczyk, A., Wawrzyński, A., Żarnecki, F., & Ogrodniczuk, M. (2023). Transferable keyword extraction and generation with text-to-text language models. In J. Mikiška et al. (Eds.), *Computational Science – ICCS 2023* (pp. 398–405). Springer. https://doi.org/10.1007/978-3-031-36021-3_42
- Pisu, A., Alessandri, L., Geraci, F., & Motta, E. (2024). Leveraging language models for generating ontologies of research topics. *CEUR Workshop Proceedings*. https://ceur-ws.org/Vol-3747/text2kg_paper6.pdf
- Saldanha, J. (2015). *The coding manual for qualitative researchers* (2nd ed.). Sage.
- Scopus: Abstract and citation database. (2025). Elsevier. <https://www.scopus.com/>
- Schwartzman, S. (2022). Pesquisa e pós-graduação no Brasil: Duas faces da mesma moeda? *Estudos Avançados*, 36(104), 227–254. <https://doi.org/10.1590/s0103-4014.2022.36104.011>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., & Moher, D. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/m18-0850>

- Tu, Y. W., Shih, M., & Tsai, C. C. (2008). Eighth graders' web searching strategies and outcomes: The role of task types, web experiences and epistemological beliefs. *Computers & Education*, 51(3), 1142–1153. <https://doi.org/10.1016/j.compedu.2007.11.003>
- Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, 37(1), 3–15. <https://doi.org/10.1108/eb026702>
- Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, 55(3), 249–270. <https://doi.org/10.1108/eum000000007145>
- Xu, D., Chen, W., Peng, W., & Wang, W. (2024). Large language models for generative information extraction: A survey. *Frontiers of Computer Science*, 18, 186357. <https://doi.org/10.1007/s11704-024-40555-y>
- Yang, H., Lin, J., & Wei, J. (2023). Research on the automatic subject-indexing method of academic papers based on climate change domain ontology. *Sustainability*, 15(5), 3919. <https://doi.org/10.3390/su15053919>