Revista Latinoamericana de Tecnología Educativa, 24(2)
https://doi.org/10.17398/1695-288X.24.2.87

RELATEC
Revista Latinoamericana de Tecnología Educativa

Received: April 28, 2025
Reviewed: July 1, 2025
Accepted: July 17, 2025

Corresponding author's address:

[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

ARTICLE / ARTIGO

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

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

Rafael Antunes dos Santos[1] & Eliseo Berni Reategui[2]

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.

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.

1. Introduction

At different levels of higher education, defining the scope of a research project is an activity in which both undergraduate and graduate students face numerous challenges when searching for and identifying bibliographic information related to their projects. These students must strike a balance between the specificity of their research topics and the openness required to find useful information that aligns with their scientific propositions. Studies indicate that students with stronger skills in refining keywords tend to achieve more successful search results in their tasks (Tu, Shih, & Tsai, 2008). Practices such as location-based hierarchical navigation and the use of social tags are additional examples that can enhance search effectiveness (Liu et al., 2011). These studies are significant as they highlight the importance of keywords in students' learning contexts, even though their findings were based on high school students conducting web searches.

Technological tools for defining the scope of research projects are becoming increasingly critical due to the growing volume and complexity of scientific literature. One such tool, word co-occurrence analysis, identifies terms that frequently appear together, revealing conceptual associations (Callon et al., 1983). This method enables an understanding of semantic relationships and provides a solid foundation for generating keywords and defining research scope (Klarin, 2024). Large Language Models (LLMs), on the other hand, represent a significant advancement, capable of processing, generating, and structuring complex information due to their training on vast textual datasets. In the scientific context, LLMs show potential to broaden analytical perspectives and inspire researchers by suggesting connections with analogous fields (Peres et al., 2023). The combination of these technologies enhances conceptual clarity, deepens understanding, and expands methodological possibilities in contemporary academic research.

In light of this context, this study aims to investigate how the automated evaluation of academic research projects can assist students in understanding the scope of their proposals and in defining keywords that accurately represent their field of study. The increasing volume and complexity of scientific literature demand approaches that support students in active learning experiences. In this scenario, it is essential that they feel empowered to relate concepts, identify patterns, and explore research possibilities in an autonomous and critical manner. Accordingly, this article poses two research questions:

RQ1. Did the keyword analysis, combined with graphical visualization of co-occurrences, contribute to the reformulation of the list and to increased confidence in the selection of keywords?

RQ2. What are students' perceptions of the usefulness of rethinking the set of keywords as a way to refine the scope of research projects?

Together, these questions form an analytical framework to understand how the combination of word co-occurrence analysis and Generative Artificial Intelligence (GenAI) can support students in the iterative refinement of their research projects. These dynamics can be examined in light of the Theory of Information Seeking and Use Behavior, which describes how users interact with information systems, select, and refine search terms to find relevant content (Wilson, 1999).

In this process, the formulation and iterative adjustment of queries play an essential role in knowledge construction, enabling students to expand or narrow the scope of their investigation based on new insights and semantic relationships. The method presented here evaluates whether the integration of GenAI with data extracted from the scientific literature can, in fact, serve as an educational resource for defining the scope of research. The aim is thus to contribute to both the quality of academic work and the autonomy of the research process.

Related works that support research project development include the dissertations of Simone de Oliveira (2017) and Ederson Bastiani (2022), although these do not specifically focus on keyword analysis.

The following sections present the theoretical foundations of scope definition, word co-occurrence analysis, and GenAI, which are essential to understanding the research objectives.

1.1. Defining the Scope of a Research Project

Academic research is essential to higher education, shaping skilled professionals and contributing to innovation, knowledge advancement, and social development (Schwartzman, 2022). A structured definition of research scope, within the context of information behavior, refers to the "systematic delimitation of the conceptual, methodological, and practical boundaries of an investigation, establishing what will be included and excluded from the study" (Wilson, 1999, pp. 256–257). Kuhlthau (2004) adds that the scope evolves during the information search process, particularly during the formulation stage, which is often marked by initial feelings of uncertainty and confusion that gradually give way to greater confidence as information is assimilated.

Proper definition of scope and the configuration of a research project are crucial, especially at the graduate level. Advisors play a key role in helping students shape and refine their projects to ensure they are feasible and relevant; however, exploring methodological approaches can be beneficial in finding an optimal balance (Cooksey & McDonald, 2019). This process offers students opportunities to define the boundaries and extent of their studies (Mbuagbaw et al., 2020; Peters et al., 2020), identify relevant variables to be investigated, or map out concepts within a knowledge domain (Pessini, Yamane, & Siman, 2022), and to provide a framework for decision-making (Tricco et al., 2018).

Whether exploratory or systematic, the process of preparing literature reviews requires the construction of search strategies in databases to retrieve the best information to help formulate the project scope. Methodological strategies described in the literature aim to facilitate the development of search models for reviews (Delaney & Tamás, 2017; Bramer et al., 2018; Grames et al., 2019), but these approaches are generally targeted at professional audiences familiar with systematic review protocols and who typically possess greater fluency in scientific vocabulary. This thematic and semantic structuring is essential for developing search strategies aligned with the research objectives. However, due to the increasing complexity of scientific data, specialized tools have become important to support students in the early stages of research development.

Defining the research scope involves setting goals and boundaries that guide the investigation. According to Wilson (1981), information needs arise from personal, social, and environmental contexts, and may be physiological, cognitive, or affective, all of which drive the information-seeking process. Barriers such as psychological, demographic, social, or environmental factors can interfere with this process. For instance, the perception of self-efficacy influences the motivation to seek information, while limited access to resources may restrict it. Understanding students' context is essential to support their searches. Dervin (1998) emphasizes that users face knowledge gaps and use information to "construct meaning." Ellis's model (1989) describes strategies such as starting, chaining, browsing, differentiating, monitoring, and extracting as ways to locate and filter relevant information.

When defining research scope, it is crucial to consider intervening variables, information needs, knowledge gaps, and use strategies. Word co-occurrence analysis and Generative AI can assist in mapping the focus of study. The following sections explore these approaches and their potential benefits.

1.2. Keyword Co-occurrence

Keywords provided by authors are essential elements in scientific articles, widely used to optimize text retrieval in bibliographic databases. Typically composed of three to five terms, they may appear literally or in varied forms within the text, representing concepts that go beyond the direct content of the work (Pęzik et al., 2023).

Keyword co-occurrence, a key concept in bibliometrics, refers to the frequency with which two or more terms appear together within a textual corpus, such as scientific articles (Zupic & Cater, 2015). This technique enables the mapping of semantic relationships and the identification of conceptual patterns, offering a comprehensive view of the thematic structures within a field (Callon et al., 1983). In bibliometrics, it has been widely used to construct scientific maps and thematic hierarchies, aiding in the understanding of connections between research areas and in tracking emerging trends (He, 1999; Zawacki-Richter & Latchem, 2022).

Beyond thematic mapping, co-occurrence analysis plays a central role in identifying and refining keywords, which are crucial for information search and retrieval (Zupic & Cater, 2015). Its application in research projects contributes to defining scopes with greater clarity and alignment with the current scientific discourse. Recently, the integration of Large Language Models (LLMs), applied in language generation tasks (a form of Generative AI), has enhanced this technique by detecting complex semantic relationships with high precision (Klarin, 2024). This synergy represents a promising approach for students and researchers in improving search strategies and scope definition.

1.3. Generative AI for Information Extraction

Large Language Models (LLMs) have become powerful tools for Natural Language Processing (NLP) tasks such as Information Extraction (IE) and keyword selection (Han et al., 2024). Trained on vast datasets, they understand and generate language in a way that closely resembles human communication and have been widely adopted in academic settings (Chang et al., 2024). Their main advantage lies in handling ambiguous and complex requirements across various domains (Lei et al., 2024). When used to create new texts, summaries, responses, or suggestions, they fall under the category of Generative AI.

Despite their strong performance in tasks without specific training (zero-shot or few-shot), LLMs tend to underperform in traditional IE when compared to fine-tuned models such as those based on BERT (Bidirectional Encoder Representations from Transformers). However, they stand out in Open Information Extraction (OpenIE), as they allow the identification of entities and relationships without predefined categories, making them adaptable to different contexts (Xu et al., 2024).

In the e-commerce domain, Maragheh et al. (2023) proposed LLM-TAKE, a model that improves keyword selection by filtering out uninformative or hallucinated terms, outperforming benchmarks on real-world data. Limitations include dependency on reference data, high computational cost, and the risk of hallucination.

For scientific contexts, Dagdelen et al. (2024) developed a method for Named Entity and Relation Recognition and Extraction (NERRE), capable of generating outputs in both natural language and JSON format. It stands out for its usability and integrated entity normalization but struggles with formatting issues and the generation of non-existent information.

Lee et al. (2023) addressed the lack of keywords in AI conference papers by proposing the use of Meta's Galactica model for automatic keyword generation. Based on similarity metrics, the study showed that 42.7% of the generated keywords matched those defined by the authors, outperforming previous approaches. GPT-based models (Generative Pre-trained Transformer) and RAG (Retrieval-Augmented Generation) have also demonstrated proficiency in ontology generation and semantic relationship identification (Pisu et al., 2024). BERT-based architectures, in turn, offer accuracy in creating topic hierarchies and performing tasks such as indexing specific research domains (Yang et al., 2023).

These studies highlight the potential of LLMs in trend analysis and in supporting the definition of research scopes, although challenges such as keyword relevance and applicability remain. The next section presents the methodological procedures adopted in this study, detailing the strategies, instruments, and analytical approach.

2. Method

The methodological decisions adopted in this study reflect an exploratory-descriptive typology and a qualitative approach. The object of analysis is the definition of the research scope by undergraduate and graduate students. Students were approached and invited to participate through prior contact with professors of higher education courses. At the undergraduate level, the courses involved research methodology and aimed at preparing students for their final thesis projects in Pedagogy and various teacher education programs. At the graduate level, students were also enrolled in a research methodology course. The invitation emphasized the voluntary nature of participation, ensuring confidentiality and the protection of students' personal data.

The study sample consisted of 23 students selected "intentionally based on pre-established criteria" (Fragoso, Recuero, & Amaral, 2012, p. 80) relevant to the research project.

The research design involved a data collection process that included the completion of two questionnaires, interspersed with the delivery of a personalized report via email regarding each participant's research scope. To support the preparation of this report, the first questionnaire (Q1) asked students to declare their research topics and questions, along with a brief initial set of keywords, as shown in Table 1.

Table 1. Questions from Questionnaire F1

Question	Format
F1.1. What is the topic of your research project?	Open-ended response
F1.2. What is your research question?	Open-ended response
F1.3. What keywords are related to your research project?	Open-ended response
F1.4. How confident are you in the keywords you have selected?	Likert scale: completely confident; confident; neutral; unsure; completely unsure
F1.5. Which databases do you use to find bibliography?	Multiple choice: Google; Google Scholar; Scielo; CAPES Portal; Repository; Library; Other

In addition to these initial questions, the questionnaire also explored students' level of confidence in their keyword choices and the databases they used during their search process, aiming to better understand participants' research practices.

The collected information served as a starting point for analyzing the preliminary outlines of each student's research project, with the goal of constructing an individualized report. After completing the questionnaire, ChatGPT-4o was used to generate personalized feedback for each student through four prompts:

Prompt 1. Commentary on the research topic and initial keyword set:

> I need to evaluate a research project. The research topic is: [...]. The research question is: [...]. The initially defined keywords are: [...]. Comment on the relevance of the keywords to the "research question" and the "research topic," but do not mention the keywords directly. If there are any issues, suggest adjustments and recommend that the student discuss them with their advisor.

Prompt 2. Suggestions for alternative keywords:

> Define a set of terms to better delimit the research project. Do not mention the keywords provided by the student. Create an initial outline for the project, presenting other related keywords that may help better frame the research context. Provide a maximum of six keywords in Portuguese, their English translations, and a brief description in Portuguese.

Prompt 3. Creation of a simplified search string:

> Create a simplified search string that combines English terms to be used in Scopus to find literature in the relevant field. Combine project-related terms with those suggested in the previous prompt. Include a brief explanation of the purpose of the proposed search string.

Prompt 4. Glossary creation using graph keywords:

> Extract the keywords from the graph and create a short glossary including the term in English, its translation into Portuguese, and a brief description of its meaning.

The personalized feedback thus constitutes an initial analysis of the research scope, as illustrated in Figure 1. In Section A of the image (left label), the data originally provided by the student can be seen. Based on this input, the report presented a short review of the keyword selection and suggested expanding it with a complementary set, including definitions and brief comments on their relevance (Section B of the image).

At this stage, Generative AI (GenAI) was used to support the construction of the report. First, an evaluation of the research topic and the initially defined keywords was conducted, including comments on their adequacy and suggestions for improvement. Next, new keywords were proposed, accompanied by brief descriptions, to help expand and refine the project scope. A simplified search string was also created by combining relevant terms for use in scientific databases, along with a short explanation of its intended use.

To compose Section C of the image, an advanced search was performed in the Scopus database using the string defined in the previous step. From this, metadata from the search results was retrieved in BibTeX format—that is, the keywords assigned by the authors of the returned papers, extracted from the DE (Descriptor) field. To ensure effective analysis, a maximum of the first 500 retrieved items was set as a limit.

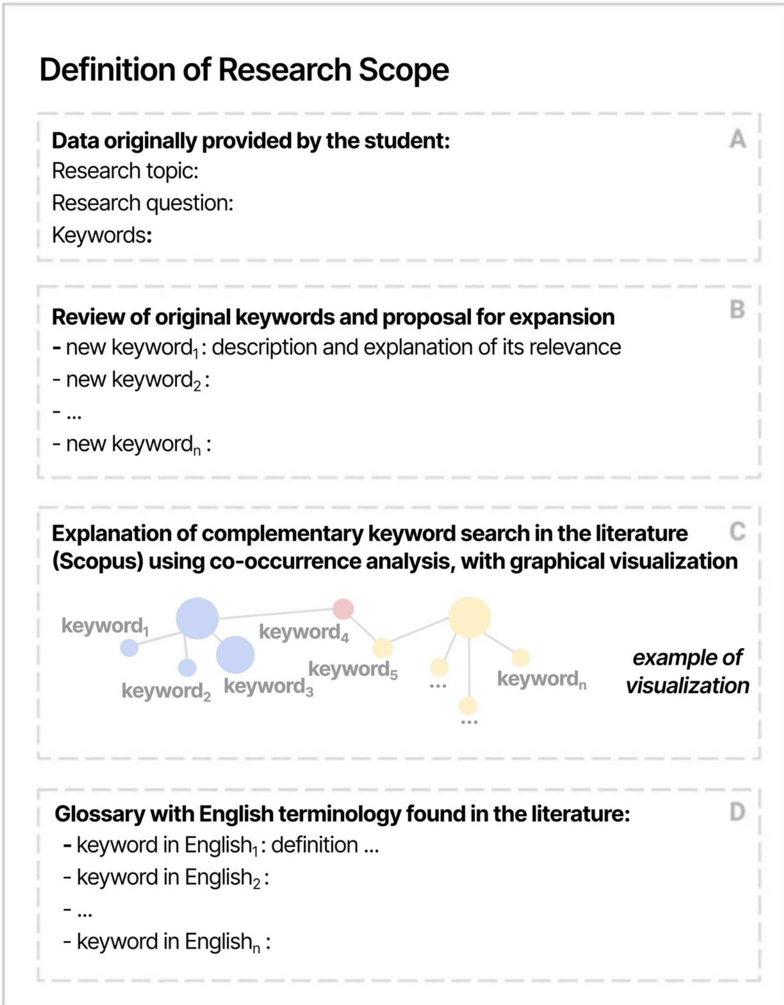

Figure 1. Report Model.

Subsequently, the bibliometric analysis itself was carried out using the Bibliometrix package (Aria & Cuccurullo, 2017) in RStudio. In this phase, specific keyword co-occurrence functions from Bibliometrix were employed, combined with the Fruchterman-Reingold algorithm, which is recognized for its ability to organize nodes (keywords) into clusters, as illustrated in Figure 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, labelsize=1.4)
```

Figure 2. Code used in Bibliometrix.

As a result of this process, keyword co-occurrence graphs were generated to illustrate the most relevant semantic associations within the context of students' research topics. Figure 3 shows a fictional example of a co-occurrence graph generated for a research project in the field of Education related to the topic "games as a learning strategy."

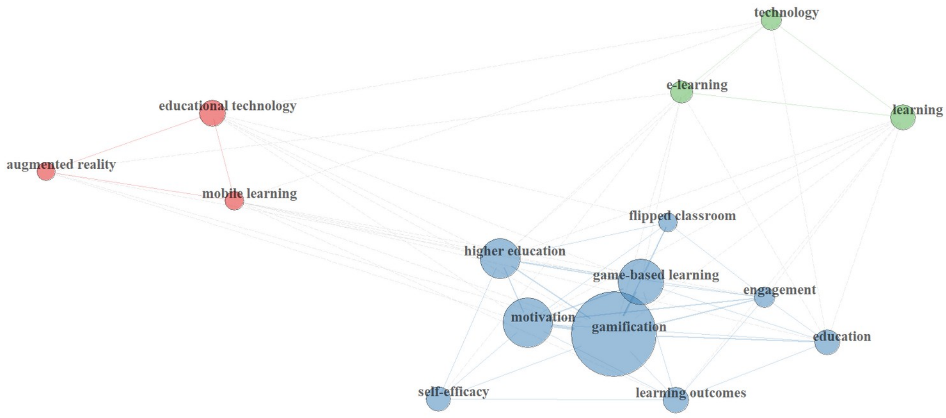

Figure 3. Co-occurrence graph.

Figure 3 displays a co-occurrence network with three thematic clusters: educational technology (red), digital learning (green), and game-based learning (blue). By providing students with a complementary view of the key concepts related to their research topic, the goal was to support scope refinement and improve the bibliographic search strategies adopted, assisting in the identification of connections between relevant keywords. The lines between nodes represent the frequency of keyword co-occurrence. The size of each node indicates how frequently that keyword appears in the corpus. For example, the graph shows that the concept gamification is the most frequent and central, associated with innovative educational strategies in higher education. There is a clear separation between clusters related to methodological innovation, applied technology, and foundations of digital learning, suggesting thematic diversity and meaningful points of convergence for defining research projects. After generating the keyword co-occurrence visualization, the resulting image was saved in PNG format and added to the student's report.

To create Section D of the report (Figure 1), Generative AI was once again used to generate an explanatory glossary of the terms present in the graph, with the aim of supporting the development of search strategies and further refining the scope of the project.

After the report was completed, students received a copy for review and were invited to respond to a second questionnaire (F2), in order to evaluate how much the analysis contributed to the refinement of their initially defined research scope. Table 2 presents the questions and response formats used.

Table 2. Questions from the Second Questionnaire (F2)

Question	Format
F2.1. Did the report on project scope make you reconsider your keyword list?	Simple choice: yes, no, neutral
F2.2. Can you estimate how many of the keywords shown in the graph seemed relevant for use in bibliographic searches?	Ordinal scale by numerical range: all; 9 to 12; 5 to 8; 1 to 4; none
F2.3. After reading the report, how confident are you in the keywords defined for your project?	Likert scale: completely confident; confident; neutral; unsure; completely unsure
F2.4. Did the keyword graph visualization help you understand the scope of your research project?	Likert scale: very helpful; helpful; neutral; not helpful; not helpful at all
F2.5. What is your perception of the usefulness of scope definition for understanding the research topic based on the report received?	Likert scale: extremely useful; useful; neutral; useless; extremely useless
F2.6. Share any positive or negative impressions regarding the keyword definition process for your project.	Open-ended response

The questions in the F2 questionnaire aimed to understand students' perceptions regarding the usefulness of the feedback received and whether they reconsidered their initial keywords. For the initial, quantitatively based questions, descriptive statistical techniques were applied, considering the absolute and relative frequencies of the data collected from both questionnaires. These analyses provided an initial understanding of participants' confidence levels in keyword selection, the perceived usefulness of expanding the research scope, and the estimated number of keywords considered relevant.

For the qualitative analysis, the content analysis method proposed by Bardin (2016) and Saldaña (2015) was adopted to systematize and interpret students' open-ended responses in relation to the two research questions posed in this article.

3. Results

This study investigated higher education students' perceptions of the experience of refining the scope of their research projects through a specific procedure. This procedure combined keyword co-occurrence analysis from bibliographic searches with the use of Generative AI (GenAI) to examine textual elements and generate personalized feedback.

To answer the first research question (RQ1), the study analyzed how the evaluation of the keyword list, along with the graphical visualization of co-occurrences, contributed to revising the list and increasing students' confidence in their keyword choices. For this, responses to questions F2.1, F2.2, and F2.3 from the questionnaire were considered.

The first of these (F2.1) showed that 82.6% of students (19 participants) reported reconsidering their keyword list after reading the report. Two students (8.7%) were unsure, and two others indicated they did not intend to change their list. Regarding the estimated use of terms from the co-occurrence graph (F2.2), nine participants identified 5 to 8 relevant terms, and eight identified 1 to 4; three stated that all terms were useful, and two pointed to 9 to 12 significant keywords. No participants considered all keywords irrelevant. These findings suggest that, although not all students fully adopted the suggested terms, the visualization helped expand their repertoire and supported the refinement of their search strategies.

Data from the second questionnaire indicated that the majority of students (18, or 78.3%) felt confident about their choice of sources for conducting searches; 3 students (13%) felt completely confident, while only 2 (8.7%) adopted a neutral stance, with no reports of insecurity among participants. When analyzing question F2.3, which explored students' confidence in their keyword selection after receiving the report, it was observed that although the overall absolute frequencies remained stable (Figure 4), significant individual shifts occurred.

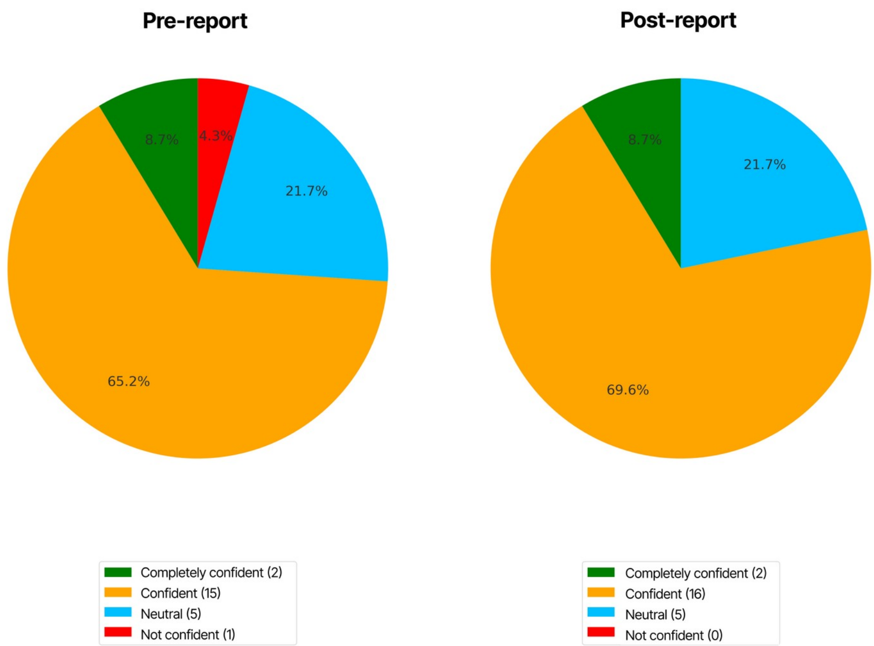

Figure 4. Confidence level in keyword selection.

Absolute counts varied only slightly; for example, the number of students who rated themselves as "confident" showed a slight increase. However, the most noteworthy finding was the internal shifts—i.e., students who, after receiving the report, changed their perception of confidence, either increasing or decreasing their level of certainty (Figure 5).

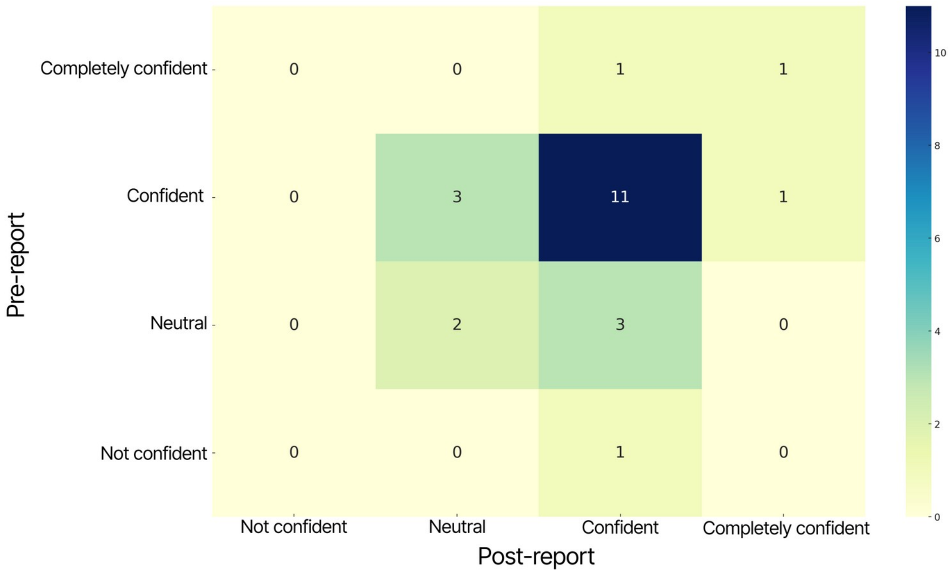

Figure 5. Change in confidence level (pre- and post-report).

The transition matrix in Figure 5 shows that while most students maintained their initial position (11 remained confident), 10 out of the 23 participants changed their self-assessment: three moved from "neutral" to "confident," one from "insecure" to "confident," and two progressed from "confident" to "completely confident," indicating increased self-confidence following the report.

Conversely, three "confident" students repositioned themselves as "neutral," and one student who had been "completely confident" now rated themselves only as "confident." These movements—both in the direction of increased and decreased confidence—reveal a reflective process triggered by the report, suggesting that the intervention prompted critical reassessment of prior decisions. Even transitions indicating reduced confidence may be interpreted as a refinement in students' personal evaluations of their projects.

The analysis of open comments from the questionnaire (F2.5) further reinforced the perception that the intervention contributed to strengthening students' confidence in their keyword selection. For instance, Student 11 noted that “it was possible to rethink the topic and the keywords more clearly,” suggesting increased clarity and assurance after receiving the report. Other students, such as Student 8, emphasized the ease of modifying the search string based on the automated suggestions, which may also indicate enhanced instrumental confidence in the refinement process.

Some students, however, noted limitations. Student 16, for example, mentioned that the suggested keywords did not fully align with their project proposal. In this regard, it's important to acknowledge that the application of GenAI to expand keyword sets and the use of co-occurrence analysis in the literature have limitations and may occasionally suggest terms that do not entirely reflect the student's intended project direction. Refining term selection criteria could enhance the system's precision and reduce these mismatches.

Despite such shortcomings, the overall results indicate that the personalized evaluation and keyword visualization contributed positively to expanding students' search repertoire, refining their keyword lists, and either strengthening or critically reassessing their confidence in the choices made for their research projects.

With respect to the second research question (RQ2), which assessed students' perceptions of the usefulness of rethinking their keyword sets for refining the scope of their research projects, responses to questions F2.3 and F2.4 were considered.

Students' perceptions of the keyword co-occurrence graph visualization in helping them understand their research scope (F2.3) were mostly positive. Although "Neutral" was the most frequent response (9 participants), 13 students (56.5%) stated that the graph "helped" (7) or "helped a lot" (6) in understanding their project's scope, indicating that most students recognized the tool's contribution to project refinement. Only one student said the graph "did not help," and none indicated it was "not helpful at all." These results suggest that even when the visual aid did not have a direct impact on all participants, it served to stimulate reflection and reinforce previously made decisions, showing a positive effect on understanding the thematic contours of their projects.

In response to the question about the perceived usefulness of the report for understanding the research topic (F2.4), the evaluation was overwhelmingly positive: 91.3% of students considered the scope report useful (34.8%) or extremely useful (56.5%), while only two participants (8.7%) remained neutral. No students rated the report as useless.

Analysis of the open-ended responses to question F2.6 further reinforced the positive aspects already identified regarding the contribution of the process to understanding the research topic and structuring the project. Statements such as "the scope definition was very positive" (Student 9), "it helped with the development of the research" (Student 17), and "it helped me identify more refined topics" (Student 23) indicate a favorable reception. These comments support the idea that the activity contributed to a clearer and more objective view of their projects. Student 1's comment exemplifies this when stating, "with the scope definition, I feel I have a better overview."

Although most comments were positive, some limitations were also noted by participants. Student 15, for instance, described the scope-definition process as "a bit complicated," and Student 9 acknowledged the richness of the suggested keywords but noted that this also made it harder to identify the central focus of the project. These comments highlight that scope refinement is a dynamic process requiring a balance between breadth and precision. Despite these occasional challenges, the overall responses indicate that the approach was effective in fostering reflection and increasing students' awareness of their research themes. The high rate of positive perception suggests that the intervention not only helped expand their thematic repertoire but also strengthened students' critical understanding of how to define their objects of study.

4. Discussion

The analysis of the results revealed a predominantly positive perception among participants, indicating the formative potential of these technologies—especially when integrated with reflective procedures such as the selection and revision of keywords. The visualization of the co-occurrence graph helped students reconsider their understanding of their project's scope and reformulate their keywords, reinforcing the link between metacognitive skills and information literacy, as pointed out by Tu, Shih, and Tsai (2008).

The use of Generative AI (GenAI) in the educational context, as outlined in this study, created a semi-guided interaction space that fostered students' critical reflection on the feedback received. More than a tool for automatic responses, GenAI functioned as a mediator, providing inputs that required active interpretation and conscious adjustments by students—contributing to the development of autonomy in defining their project scopes. Regarding confidence in keyword definition, the apparent stability in aggregate data masked significant individual shifts: students who were initially uncertain became more confident, while others, initially confident, adopted a more neutral stance after the intervention. These transitions suggest a genuine process of critical reassessment and refinement in understanding the scope of one's own research, aligning with Kuhlthau's theory (2004), which describes the formulation phase as a moment of instability and progressive reconstruction of understanding.

Consistent with studies such as Haman and Školník (2023), this work demonstrates that GenAI can help reduce cognitive load during research planning by supporting the organization and structuring of conceptual repertoires, making it easier to visualize relevant term connections. In the present study, although most students found 5 to 8 graph terms useful, some considered nearly all of them relevant—indicating variations in the degree of alignment between the automated output and the individual project. This variation may be related to project maturity or the initial clarity of the research topic. Thus, rather than merely expanding the number of keywords, the intervention proved effective in supporting students to structure and refine their thinking about their research themes.

The critical perceptions raised by students are also noteworthy, particularly concerning the disconnect between some of the suggested keywords and the specific focus of their projects. This kind of limitation echoes points raised by Peres et al. (2023), who emphasize the dependency on input quality and the need for contextual interpretation as ongoing challenges in effectively applying GenAI in educational settings. These findings reinforce that the use of generative AI technologies must be accompanied by critical mediation—both by students and instructors—to ensure that suggestions are appropriately interpreted and aligned with each research project's goals.

The results also prompt reflection on the integration of technologies such as keyword co-occurrence analysis and GenAI into scientific training curricula. When systematically embedded, these tools can contribute to the development of skills in information searching, organization, and critical evaluation—aligning with Wilson's theory of information behavior (1999) and the principles of meaningful learning. However, their adoption requires more than the mere introduction of new tools: it calls for pedagogical strategies that promote reflective and critical use of these resources. As Klarin (2024) points out, the iterative and exploratory approach promoted by these technologies is particularly valuable in the early stages of research, when students are still building their understanding of the topic. Nevertheless, their impact will depend on the ability of curricula to integrate these practices contextually, avoiding the risk of purely instrumental adoption.

5. Conclusion

This study demonstrated that the combination of keyword analysis, graphical visualization of co-occurrences, and personalized feedback mediated by Generative AI (GenAI) can be an effective strategy to support students in refining the scope of their research projects. The main contribution of this approach was to enhance students' self-awareness through critical reflection on thematic delimitation, expansion of conceptual repertoire, and increased confidence in keyword selection—essential elements for the development of more structured and well-founded research.

The results also indicated that the effectiveness of the approach described here depends heavily on the quality of the initial information provided by students and the level of maturity of their projects. These factors directly influence the relevance and applicability of the suggestions generated. Therefore, the need for active pedagogical supervision is reinforced, ensuring that these tools are used critically and that the results are contextualized within each student's learning process.

In conclusion, this work shows that the integration of information science and AI opens new methodological possibilities to support students and researchers in navigating the increasing complexity of scientific literature. By structuring strategies such as keyword co-occurrence analysis and AI-mediated personalized feedback generation, it becomes possible not only to organize large volumes of information, but also to foster conceptual refinement and strengthen information literacy.

One limitation of the study concerns the size and composition of the sample, which consisted of 23 undergraduate and graduate students with research projects at different stages of development. This heterogeneity may have influenced perceptions of the usefulness of the tools, particularly in relation to the clarity of research topics and the ability to incorporate the suggestions provided. Furthermore, the intervention was conducted at a single point in the research scoping process, limiting the analysis of long-term effects on project development. Future studies could include ongoing student follow-up to more comprehensively assess the impact of these technologies throughout the various stages of academic research.

6. References

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., Rochwerg, 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. Mikyška 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/eum0000000007145

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