

Exploring Generative AI and Natural Language Processing to Develop Search Strategies for Systematic Reviews

Sarah G. Park, University of Illinois at Urbana - Champaign

Sarah Park is the Mathematics and Computational Sciences Librarian and Head of the Mathematics Library at the University of Illinois at Urbana-Champaign. She supports active engagement with faculty, students, and staff to promote library services and collections and to support emerging information literacy.

Monica Carroll, University of Illinois at Urbana - Champaign

Monica Carroll is an Engineering Librarian at the University of Illinois at Urbana-Champaign. Originally trained as a geologist, she is interested in the sociology and social history of academic research in STEM disciplines.

Lucy Marie Alice Esteve, Duke University

Lucy Esteve is an Endocrinology Research Fellow in the Department of Medicine at Duke University. Her academic interests focus on the use of technology (in particular wearable smartwatches) for early detection of diabetes and prevention of diabetes-related complications.

Karnika Singh, Duke University

Karnika Singh is a PhD candidate in Biomedical Engineering at Duke University. Her research is focused on the use of digital health technologies for health monitoring.

Exploring Generative AI and Natural Language Processing to Develop Search Strategies for Systematic Reviews

Abstract

Systematic reviews, as well as other comprehensive literature reviews, require a rigorous and structured approach for an exhaustive literature search. Identifying the relevant keywords that encapsulate the research topic and then combining them using the appropriate search technique demands expertise in both search techniques and in-depth domain knowledge. As a result, librarians are frequently asked to join research teams to assist researchers unfamiliar with this intricate process. These subject specialists have expertise in using search tools and techniques to formulate search statements, yet they often need to conduct extensive literature surveys to identify relevant search terms, especially in rapidly evolving fields. Although controlled vocabularies and keywords supplied by authors and database companies are useful, they are often insufficient in capturing all relevant terms used in the literature. This paper explores how Generative Artificial Intelligence (AI) and Natural Language Processing (NLP) could be harnessed by librarians to refine the formulation of search terms for these reviews. This pilot study suggests that the use of Generative AI and NLP helps users identify relevant search terms for developing search strategies, although users must be cautious about the reproducibility of Generative AI's responses.

Introduction

Systematic reviews are a type of literature review that takes a systematic approach to search, review, and synthesis of information on a defined topic. A well-designed review will achieve “exhaustive and comprehensive searching”[1] for relevant evidence while minimizing selection bias. As the volume of scientific publications has surged exponentially over the past several decades [2], with a recent study showing that the number of publications in the Physical and Technical Sciences doubles every 11.9 years [3], the need for systematic reviews to collate and synthesize all this research has become critical. The general field of engineering has seen the rise in published research review papers that many other fields have seen recently [4-6]. Conversely, performing these systematic reviews presents a significant challenge to researchers in achieving both effectiveness and efficiency.

The growing popularity of publishing systematic reviews contrasts with a growing body of clinical medical literature stressing the challenges and common failures evident in reviews. “[Evidence syntheses] are complicated and time-consuming undertakings prone to bias and

errors. Production of a good evidence synthesis requires careful preparation and high levels of organization in order to limit potential pitfalls” [7]. For example, various methods and approaches have emerged to navigate and identify relevant works within scientific literature. Some of these methods include citation chasing utilizing both library databases and Google Scholar [8]. Similarly, researchers use "related articles" or suggested article features available in article databases [9]. While these strategies can uncover numerous sources, they also carry the risk of introducing bias into the literature review process. Adopting more objective methods for conducting literature reviews is just as vital as obtaining comprehensive sources.

In recent decades, the health and medical research communities have established a transparent approach to minimize bias in searching and reviewing existing literature [10]. This objective and SR methodology have gained traction in fields closely related to medicine, such as biomedical engineering and other health-related science disciplines [5]. Engineering has also begun to embrace SR methods, with recent studies highlighting the trends of systematic reviews in engineering education and emphasizing the widespread need for training engineering students and scholars in these methodologies [11].

Research consultations for SRs represent a new and expanding role for engineering librarians. The involvement of librarians is crucial to the systematic review process to train students and researchers, and to help formulate effective search strategies [12]. Some librarians are invited to join research teams to assist in developing search strategies and co-author papers. This growth in demand for library expertise on engineering reviews can represent a significant demand on librarian time. Numerous articles have delineated the level of engagement and responsibilities that librarians conduct in the systematic review process and offered guidance for less experienced librarians to engage with researchers effectively [13, 14]. The initial literature interview, search and review phase can demand over 20 hours [15]. Engineering librarians will be called upon to employ various strategies to bridge any gaps in domain expertise.

In this paper, we explore a variety of innovative methods, supported by technological advances such as natural language processing (NLP) and generative AI, to investigate whether these tools can enhance the efficiency and effectiveness of librarians' work in the systematic review process. Specifically, we attempt to answer the following research questions:

1. Can natural language processing methods be utilized to extract relevant search terms from documents?
2. Is it feasible to use generative AI to suggest relevant search terms?
3. Can either of these technologies assist users in formulating an improved search statement?

Literature Review

The ideal search strategy in a systematic review aims to retrieve all relevant studies – maximizing recall while excluding irrelevant ones to ensure precision. In practice, enhancing one measure usually inversely affects the other. However, this has changed over time. The methodologies for SRs have shifted since 1985, when Cooper outlined 15 methods, such as browsing library shelves and sending general requests to government agencies [16]. Today a search is typically detailed in a SR protocol by specifying the search statement/database pairs, along with any limits or filters and date the search was conducted [17]. The database selection is often a discipline-based consensus, so it is the search statement that controls how effective searching is.

The design of a search query is the most time-consuming aspect of a librarian's task in a systematic review [15]. Creating a sequence of search terms that optimizes both recall and precision requires deep familiarity with the literature, including terms used outside one's regular expertise (White, 2009). This process of digesting a body of text and identifying patterns seems ideally suited to automation. In 2009, Ananiadou et al., [18] discussed using text mining to extract terms and expand queries. Text mining applies statistical analysis to a specific body of text to identify patterns, including associated terms. Natural language processing (NLP) is a closely related concept where computer programs (i.e., machine learning) extract and utilize these patterns on unstructured text to aid understanding. NLP is a field that interacts with artificial intelligence (AI) and employs machine learning (ML) techniques. ChatGPT is an AI application that leverages NLP principles to understand user inputs and generate responses.

The systematic review process can be divided into smaller tasks, and some of them can reasonably be targeted for automation, like screening papers for inclusion or extracting data from them. Software such as Covidence is already aiding researchers in these areas. However, a 2020 review by van Dinter et al. [19] examined 41 systematic reviews using some form of automation and found that only 7 included automating the search query creation process. A 2021 study found that text-mining tools could significantly reduce the time spent creating queries, though it also found a decrease in query sensitivity [20]. Furthermore, a 2022 review of artificial intelligence used in systematic reviews, by Blaizot et al. [21] found 12 health science studies, yet none applied AI to search query creation. This indicates an absence of tools used to create search statements although Toth et al. [22] propose that this could be fully automated.

Given the increase in systematic reviews within engineering there is a large need for engineering librarian involvement. Automation tools might be a way for librarians to efficiently manage serving researchers across a broad range of topics.

Research Methods

The team formation

A team, consisting of a medical doctor and a Ph.D. student in biomedical engineering, reached out to a librarian for research consultation in conducting a systematic review on non-invasive techniques for monitoring human blood glucose levels. Although the team had expertise in both the medical and engineering fields, they required guidance in executing their first systematic review. The team specifically needed guidance in crafting a thorough and unbiased search strategy to review literature and gather relevant studies from databases.

Reference interview and writing the initial search statement.

After the initial reference interview, the team and the librarian pinpointed three major concepts: glucose, non-invasive, and sensor. They agreed to conduct searches in three primary databases: Web of Science, PubMed, and Scopus. The librarian expanded the search terms for these major concepts based on information from the initial reference interview, MeSH review, and examination of seed articles discussed with the team. These key terms were then combined using Boolean operators to represent the concepts, initially for PubMed and later manually translated for use in Web of Science and Scopus. The search strategies were almost identical, except the one for PubMed included MeSH terms.

The preliminary search statement entered for Web of Science Core Collection and MEDLINE

("diabetes mellitus" OR diabet* OR glucose OR hyperglycemi* OR hyperglycaemi* OR hypoglycemi* OR hypoglycaemi* OR glycemi*) AND (noninvasive OR "non invasive" OR "minimally invasive" OR "less invasive") AND ("Biosensing Techniq*" OR sensor* OR sensing OR biosens* OR biomarker*)

Initial search and initial result review

The initial search statements yielded 3,696 results from PubMed, 6,397 from the Web of Science Core Collection plus MEDLINE (limited to topics), and 5,593 from Scopus (limited to titles, abstracts, and keywords) as of September 30, 2023. Despite significant overlap among the three databases, the sheer volume of results was overwhelming and required further refinement. To refine the search, the team decided to review additional keywords. Rather than downloading all records and deduplicating results across the three databases, the librarian opted to download and analyze the largest dataset, which came from the Web of Science, using natural language processing (NLP).

Keyword Extraction

Each full record from the Web of Science Core Collection and MEDLINE is comprised of multiple fields, mainly TI (Title of the Article), AB (Abstract), DE (Author Keywords), and ID (Clarivate's Keywords Plus). Therefore, results were extracted into three files: 1) free text from titles and abstracts, 2) author keywords, and 3) Keywords Plus. Given that both the Author Keywords and Keywords Plus fields consist of words and phrases separated by commas, they were subjected to simple term frequency analysis.

Term Frequency analysis (TF) quantifies how many times a word or phrase appears within a document, which is useful for evaluating keyword lists like Author Keywords and Keywords Plus. However, assigning equal weight to all keywords fails to capture their relative significance across multiple documents. In contrast, TF-IDF (Term Frequency-Inverse Document Frequency) assesses a word's relevance in a document by merging its document frequency (TF) with the rarity of the term across the entire corpus (IDF) [23]. This method emphasizes words that are common in a specific document but otherwise uncommon, more effectively revealing word relevance. As a result, TF-IDF is more appropriate for extracting keywords from free texts such as abstracts and titles.

To conduct the TF-IDF analysis on the titles and abstracts, the authors employed Python's Natural Language Toolkit (NLTK) and programmed it to process n-gram extraction. They then used scikit-learn's TfidfVectorizer to calculate the TF-IDF scores for each n-gram [24]. This approach facilitated a more nuanced analysis of the textual data, enabling the authors to identify and review a list of keywords extracted from the keyword phrases and sentences from the title and abstract fields.

Generative AI

In parallel with the keyword extraction using NLP, the team investigated whether Generative AI could assist in discovering more relevant terms related to wearable sensors for monitoring and sensing blood glucose levels in humans for a systematic review. The team used ChatGPT versions 3.5 and 4 multiple times, prompting the AI with questions for keywords related to their original research questions. This initial exploration was conducted in September 2023, and the same prompts were tested again twice in January 2024. All prompts and responses were documented for future discussions.

Evaluation

After reviewing the four sets of keywords extracted and generated from the methods above, a new search statement was formulated for Web of Science, which was then adapted for Scopus and PubMed. These revised search statements were then presented to the research team for

feedback. This collaborative process led to the finalization of the search strategy, integrating keywords identified through natural language processing and those suggested by ChatGPT. In this paper, the authors compare and analyze the search terms ultimately selected by the research team with the initial sets, to assess the effectiveness of the various methods outlined earlier.

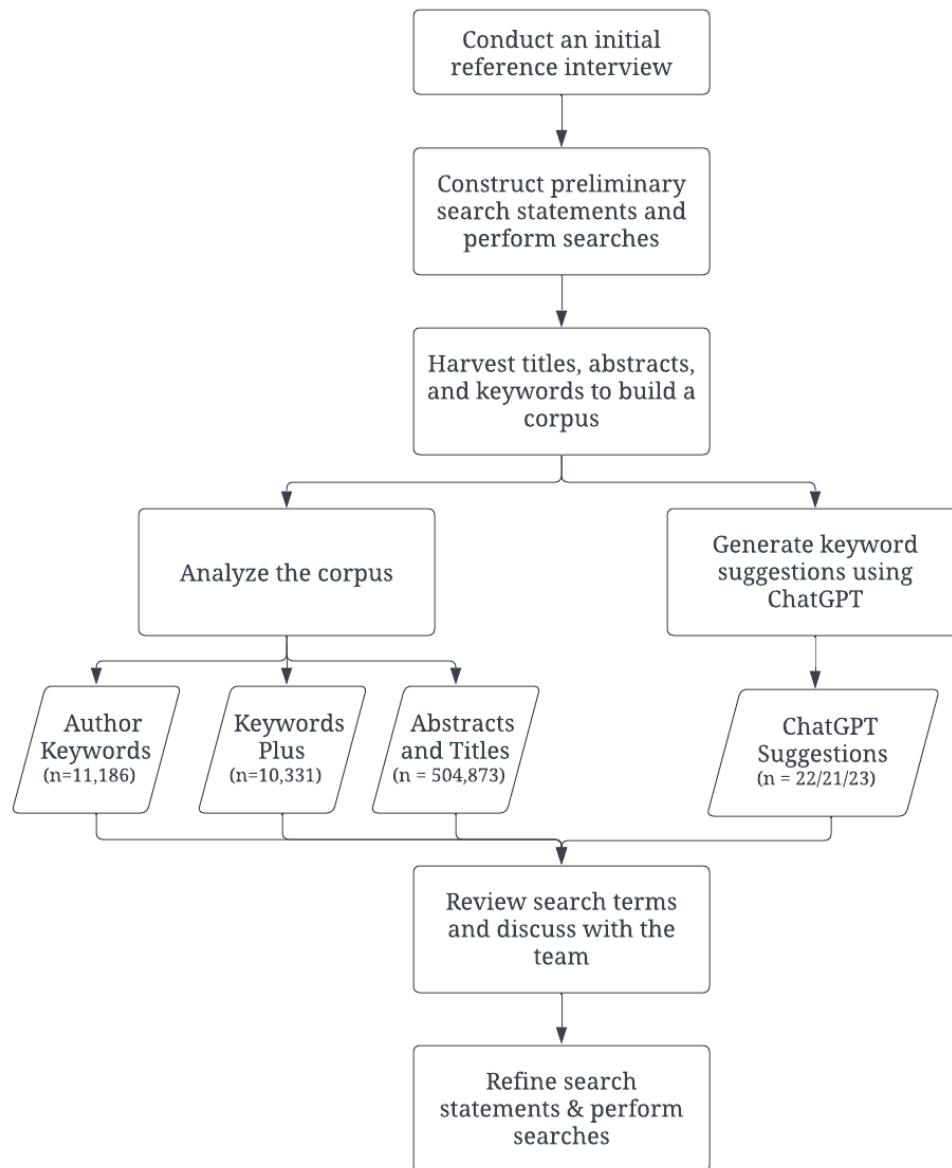


Figure 1. Flow chart of the research process, including the number of keywords in NLP analysis

Results and analysis

The preliminary search statement yielded 6,397 results from the Web of Science Core Collection and MEDLINE. Term Frequency (TF) analysis within the Author Keywords section revealed 11,186 unique keywords, with “diabetes” appearing most frequently, 434 times. Keywords Plus, a unique feature of Web of Science, generates keywords from the titles of cited articles not included in the article's title or Author Keywords [25]. The TF analysis within Keywords Plus revealed 10,331 unique terms, with “GLUCOSE” as the top keyword, occurring 354 times. The top 20 keywords identified from the TF analysis are listed in Table 1.

Table 1: Top 20 terms identified in the Author Keywords and Keywords Plus fields

Ranks	Author Keywords	TF	Keywords Plus	TF
1	diabetes	434	GLUCOSE	354
2	glucose	277	DIAGNOSIS	291
3	diabetes mellitus	190	DISEASE	274
4	biomarkers	189	SENSOR	269
5	biomarker	181	RISK	219
6	non-invasive	180	IN-VIVO	211
7	diabetic retinopathy	142	INSULIN-RESISTANCE	210
8	biosensor	112	EXPRESSION	203
9	noninvasive	96	SENSORS	189
10	type 2 diabetes	95	PREVALENCE	183
11	obesity	79	BIOMARKERS	170
12	glucose sensor	76	BLOOD-GLUCOSE	166
13	diagnosis	72	NANOPARTICLES	161
14	blood glucose	71	ASSOCIATION	159
15	nonalcoholic fatty liver disease	70	SYSTEM	154
16	glucose sensing	69	BIOSENSOR	148
17	saliva	69	MELLITUS	145
18	inflammation	66	DIABETES-MELLITUS	133
19	biosensors	66	MORTALITY	129
20	optical coherence tomography	63	CELLS	123

The Title and Abstract fields were merged and analyzed using TF-IDF, programmed to identify 2-grams, or two-word consecutive word phrases, instead of single words. This analysis resulted in identifying 504,873 unique keywords. Table 2 below lists the top 20 keywords from the TF-IDF analysis in the Title and Abstract fields, sorted by their TF-IDF scores.

Table 2: Top 20 terms sorted by TF-IDF scores, identified through TF-IDF analysis in the Title and Abstract (TiAb) fields

Ranks	TiAb	TF	TF-IDF
1	glucose	4258	280.385
2	patients	2971	156.757
3	non	3962	147.816
4	blood	2614	141.343
5	invasive	4315	140.797
6	diabetes	3297	138.088
7	monitoring	2505	133.607
8	diabetic	2061	122.276
9	non invasive	3484	121.855
10	disease	2393	117.214
11	based	2668	116.698
12	sensor	1628	114.523
13	using	2946	109.131
14	noninvasive	2656	104.309
15	detection	1872	98.875
16	liver	862	97.047
17	blood glucose	1333	94.092
18	biomarkers	1602	84.318
19	retinal	738	82.970
20	sensing	1241	81.943

The librarian initially reviewed the compiled lists, checked for additional keywords, and revised the search statement. These compiled lists and the revised search statements were then shared with researchers to identify any further keywords for the final search statement. To assess whether the analysis helped identify additional keywords, the preliminary search statement was compared with the final one.

The table below compares the search statements before and after revision, including the expanded keywords for biomarkers. For this pilot study analysis, we are presenting only the parts of the revised search statement that pertain to the original concepts, even though additional concepts and keywords were later added by the researchers. Furthermore, the researchers have chosen not to disclose the full list of keywords of biomarkers, sharing only those relevant to the subsequent discussion.

The keywords originated from MeSh Terms, such as “diabetes mellitus” or “Biosensing Techniq*,” are kept in the revised search statement despite being redundant to diabet* or biosens* to facilitate its translation into a PubMed search statement later.

Table 3: Comparison of initial vs. revised search statement.

	Preliminary search statement	Part of the revised search statement
Concept #1: Diabetes	("diabetes mellitus" OR diabet* OR glucose OR hyperglycemi* OR hyperglycaemi* OR hypoglycemi* OR hypoglycaemi* OR glycemi*)	("diabetes mellitus" OR diabet* OR glucose OR hyperglycemi* OR hyperglycaemi* OR hypoglycemi* OR hypoglycaemi* OR glycemi* OR glycaemi* OR "insulin resistan*" OR "impaired glucose" OR metabol* OR "glucose prediction" OR "blood sugar" OR OGTT OR HbA1c OR hba1c OR "hemoglobin A1c" OR "haemoglobin A1c" OR "glycated hemoglobin" OR prediabet* OR pre-diabet*)
Concept #2: Noninvasive	(noninvasive OR "non invasive" OR "minimally invasive" OR "less invasive")	(noninvasive OR "non invasive" OR "minimally invasive" OR "less invasive")
Concept #3: Sensors	("Biosensing Techniques" OR sensor* OR sensing OR biosens* OR biomarker*)	("Biosensing Techniq*" OR sensor* OR sensing OR biosens* OR biomarker* OR smartwatch OR "wearable sens*" OR "smart phone" OR "Apple watch" OR Fitbit OR "smart ring" OR whoop OR Huawei OR aktiv OR "fitness tracker*" OR "activity tracker*" OR "smart watch*")
Concept #4: Biomarkers	-	("heart variab*" OR HRV OR "heart rate variab*" OR "heart rate" OR EKG OR elektrokardio OR elektrodermal OR EDA OR "galvanic skin" OR GSR OR sleep* OR [...] OR Actigraph* OR GPS OR "ambient light" OR "oxygen satura*" OR capacitive OR "blood volume puls*")

Moreover, a Python script was written to determine which fields and methods identified the most relevant search terms within three keyword lists – Author Keywords, Keywords Plus, and Titles and Abstracts.

Of the total 78 search terms in the revised search statement, 68 search terms appeared in the Titles and Abstracts keyword list, with 10 not found. Seventeen search terms were not matched with keywords from Author Keywords, and the Keywords Plus set did not include the corresponding search terms for 41. The high number of search terms not found in the Keywords Plus set is consistent with the findings of Zhang et al.’s [26], which reported that keywords in Keywords Plus alone are not effective at representing the content of the article.

The researchers included product or brand names in the search statements, highlighted in blue in the table below. These names were found in some Titles and Abstracts but were less prevalent in Author Keywords and Keywords Plus. It's important to note that the TF-IDF analysis for Abstracts and Titles was conducted using a 2-gram approach, meaning keywords composed of three or more consecutive words were not matched and are highlighted in red font. Although the final search terms include a small number of phrases longer than two words, the author believes the effect of limiting the analysis to 2-grams is considered negligible.

Table 4: Search terms absent from search fields.

Not found in Titles and Abstracts Keywords	Not found in Author Keywords (DE)	Not found in Keywords Plus (ID)
aktiv	actigraph*	actigraph*
blood volume puls*	activity track*	activity track*
ekg	Aktiv	Aktiv
elektrokardio	ambient light	ambient light
heart rate variab*	apple watch	apple watch
heart variab*	Ekg	biosensing techniq*
Huawei	Elektrokardio	blood sugar
pre-diabet*	Fitbit	blood volume puls*
smart ring	fitness track*	capacitive
whoop	Gps	diabetes mellitus
	haemoglobin a1c	ekg
	heart variab*	electrodermal
	Huawei	elektrokardio
	smart phone*	fitbit
	smart ring	fitness track*
	smartwatch*	galvanic skin
	whoop	glucose prediction
		glycaemia*
		gps
		gsr
		haemoglobin a1c
		heart rate variab*
		heart rate*
		heart variab*
		hemoglobin a1c
		hrv
		huawei
		hyperglycaemi*
		hypoglycaemi*
		less invasive
		non invasive

		ogtt
		oxygen satura*
		pre-diabet*
		prediabetes
		smart phone*
		smart ring
		smart watch
		smartwatch*
		spO2
		whoop

In parallel to NLP, the authors also utilized ChatGPT 4.0, a Generative AI application, to suggest a list of keywords related to wearable devices. In September 2023, ChatGPT provided a list of 22 keywords. When a similar question was asked on January 16, 2024, as part of content verification for this paper, it generated a list of 21 keywords, omitting "Camera" and "IR Sensors" (highlighted in yellow) but adding "Compass" (highlighted in gray). Remarkably, when the question was repeated just 10 days later, ChatGPT expanded the list to include two additional keywords: "hydration sensor" and "stress and fatigue sensors," which are highlighted in green. The variation in responses to the same or similar questions is noteworthy. The prompts for each instance and the corresponding responses are documented in the table below, with changes between different instances highlighted in yellow, gray, and green for clarity.

Table 5: Variations of ChatGPT 4.0 Responses over time

"Can you list a comprehensive list of sensors available through commercial wearable devices including smartwatch?" (2023, September)	"Could you please provide me with a comprehensive list of sensors available on commercial wearable devices, including smartwatches?" (2004, January 16)	"Could you please provide me with a comprehensive list of sensors available on commercial wearable devices, including smartwatches?" (2024, January 26)
Accelerometer	Accelerometer	Accelerometer
Ambient Light Sensor	Ambient Light Sensor	Ambient Light Sensor
Barometer (or Atmospheric Pressure Sensor)	Barometer	Barometer
Bioimpedance Sensor	Bioimpedance Sensor	Bioimpedance Sensor
Blood Oxygen (SpO2) Sensor	Blood Oxygen Sensor (SpO2)	Blood Oxygen (SpO2) Sensor
Camera	Capacitive Sensor	Capacitive Sensor
Capacitive Sensor	Compass	Compass
Electrocardiogram (ECG or EKG) Sensor	ECG (Electrocardiogram) Sensor	ECG (Electrocardiogram) Sensor

Galvanic Skin Response (GSR) Sensor	Galvanic Skin Response Sensor	Galvanic Skin Response Sensor
Glucose Monitor	Glucose Monitoring Sensor	Glucose Monitoring Sensor
GPS (Global Positioning System)	GPS (Global Positioning System)	GPS (Global Positioning System)
Gyroscope	Gyroscope	Gyroscope
Heart Rate Monitor	Heart Rate Monitor (HRM) (Photoplethysmography or PPG).	Heart Rate Monitor (HRM)
IR Sensors	Magnetometer	Hydration Sensor
Magnetometer (or Compass)	Microphone	Magnetometer
Microphone	NFC (Near Field Communication)	Microphone
NFC (Near Field Communication)	Pedometer	NFC (Near Field Communication)
Pedometer	Proximity Sensor	Pedometer
Proximity Sensor	Skin Conductance Sensor	Proximity Sensor
Skin Conductance Sensor	Temperature Sensor	Skin Temperature Sensor
Temperature Sensor	UV Sensor	Stress and Fatigue Sensors
UV Sensor		Temperature Sensor
		UV Sensor

Discussion

The primary aim of this study was to investigate the application of various NLP tools in extracting search terms and formulating search strategies for systematic reviews and other comprehensive literature reviews. The search terms extracted from titles and abstracts closely matched those chosen by researchers and experts, with Author Keywords being the second most useful and Clarivate's Keywords Plus showing the least utility, corroborating Zhang et al.'s earlier findings [26]. Although this study focused on a single SR consultation, it investigated various methods for rapidly reviewing keywords and topics to aid researchers in refining their research focus.

The authors also highlight the challenge of downloading certain database fields, leading to the exclusion of controlled vocabulary from the analysis. The datasets for the NLP method investigation were sourced from the Web of Science Core Collection and MEDLINE. Although Web of Science offers searchable MeSH terms from MEDLINE®, these terms could not be downloaded through a web interface. These controlled vocabulary terms could be valuable in

identifying relevant articles but as they were functionally unavailable, they were excluded from the NLP analysis.

Furthermore, the study examined the potential and limitations of using Generative AI to identify key search terms. Unlike the labor-intensive processes of coding for TF and TF-IDF analyses, Generative AI can rapidly generate responses with search terms with the proper prompts, bypassing the need for extensive data preparation and analysis. However, the inconsistency in Generative AI's responses over time (Table 5) and the lack of reproducibility raise questions about its credibility as a tool for research or scientific reasoning. Despite the promise Generative AI holds in easing systematic review processes, its variability underscores the need for caution. Librarians should advise researchers to perform secondary verifications rather than rely solely on Generative AI's responses.

Conclusion

The increasing adoption of systematic reviews in engineering underscores the critical need for methods that are both effective and methodical in crafting search strategies. This investigation aimed to assess the potential of NLP and Generative AI technologies in aiding librarians and researchers involved in systematic reviews and extensive literature searches, with the goal of making the process of identifying relevant search terms more streamlined. The study explored the use of various NLP techniques to uncover new search terms from bibliographic records. It also examined the role of Generative AI within this framework. While Generative AI can rapidly provide answers to prompts, thereby reducing the time and effort involved in the NLP process, its responses can vary over time, necessitating a secondary verification process. Despite this, both NLP and Generative AI serve as complementary tools in the systematic review process, each contributing to the efficiency of search strategy development as well as the comprehensiveness of the search strategy itself.

References

- [1] M. J. Grant and A. Booth, "A typology of reviews: an analysis of 14 review types and associated methodologies," *Health information & libraries journal*, vol. 26, no. 2, pp. 91-108, 2009.
- [2] P. Larsen and M. Von Ins, "The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index," *Scientometrics*, vol. 84, no. 3, pp. 575-603, Sep 2010, doi: 10.1007/s11192-010-0202-z.
- [3] L. Bornmann, R. Haunschild, and R. Mutz, "Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and

- new literature databases," *Humanities and Social Sciences Communications*, vol. 8, no. 1, p. 224, 2021/10/07 2021, doi: 10.1057/s41599-021-00903-w.
- [4] M. Borrego, M. J. Foster, and J. E. Froyd, "Systematic literature reviews in engineering education and other developing interdisciplinary fields," *Journal of Engineering Education*, vol. 103, no. 1, pp. 45-76, 2014.
 - [5] A. Henry and L. Stieglitz, "An Examination of Systematic Reviews in the Engineering Literature," in *2020 ASEE Virtual Annual Conference Content Access*, 2020.
 - [6] M. Phillips, J. B. Reed, D. Zwicky, and A. S. Van Epps, "A scoping review of engineering education systematic reviews," *Journal of Engineering Education*, 2023.
 - [7] K. Kolaski, L. R. Logan, and J. P. A. Ioannidis, "Guidance to best tools and practices for systematic reviews," *Systematic Reviews*, vol. 12, no. 1, p. 96, 2023/06/08 2023, doi: 10.1186/s13643-023-02255-9.
 - [8] N. R. Haddaway, M. J. Grainger, and C. T. Gray, "Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching," *Research Synthesis Methods*, vol. 13, no. 4, pp. 533-545, Jul 2022, doi: 10.1002/jrsm.1563.
 - [9] A. Booth, "Unpacking your literature search toolbox: on search styles and tactics," *Health information and libraries journal*, vol. 25, no. 4, p. 313, Dec 2008, doi: 10.1111/j.1471-1842.2008.00825.x.
 - [10] E. Alvaro and D. A. Zwicky, "Improving the Quality of Literature Reviews in Chemistry: Tools and Techniques," *ACS Chemical Health & Safety*, vol. 28, no. 5, pp. 303-310, 2021/09/27 2021, doi: 10.1021/acs.chas.1c00046.
 - [11] J. Reed, M. Phillips, A. S. Van Epps, and D. Zwicky, "An early look at a scoping review of systematic review methodologies in engineering," in *2020 IEEE Frontiers in Education Conference (FIE)*, 2020: IEEE, pp. 1-4.
 - [12] T. Muka *et al.*, "A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research," *Eur J Epidemiol*, vol. 35, no. 1, pp. 49-60, Jan 2020, doi: 10.1007/s10654-019-00576-5.
 - [13] A. Ross-White, "An environmental scan of librarian involvement in systematic reviews at Queen's University: 2020 update," *Journal of the Canadian Health Libraries Association/Journal de l'Association des bibliothèques de la santé du Canada*, vol. 42, no. 2, 2021.
 - [14] S. McKeown and A. Ross-White, "Building capacity for librarian support and addressing collaboration challenges by formalizing library systematic review services," (in eng), *J Med Libr Assoc*, vol. 107, no. 3, pp. 411-419, Jul 2019, doi: 10.5195/jmla.2019.443.
 - [15] K. Bullers *et al.*, "It takes longer than you think: librarian time spent on systematic review tasks," *Journal of the Medical Library Association: JMLA*, vol. 106, no. 2, p. 198, 2018.
 - [16] H. Cooper, "Literature searching strategies of integrative research reviewers," *American Psychologist*, vol. 40, no. 11, p. 1267, 1985.
 - [17] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *International journal of surgery*, vol. 88, p. 105906, 2021.
 - [18] S. Ananiadou, B. Rea, N. Okazaki, R. Procter, and J. Thomas, "Supporting systematic reviews using text mining," *Social science computer review*, vol. 27, no. 4, pp. 509-523, 2009.

- [19] R. van Dinter, B. Tekinerdogan, and C. Catal, "Automation of systematic literature reviews: A systematic literature review," *Information and Software Technology*, vol. 136, p. 106589, 2021.
- [20] R. A. Paynter, R. Featherstone, E. Stoeger, C. Fiordalisi, C. Voisin, and G. P. Adam, "A prospective comparison of evidence synthesis search strategies developed with and without text-mining tools," *Journal of Clinical Epidemiology*, vol. 139, pp. 350-360, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0895435621000858?via%3Dihub>.
- [21] A. Blaizot *et al.*, "Using artificial intelligence methods for systematic review in health sciences: A systematic review," *Research Synthesis Methods*, vol. 13, no. 3, pp. 353-362, 2022.
- [22] Á. Tóth, A. Suta, J. Pimentel, and A. Argoti, "A comprehensive, semi-automated systematic literature review (SLR) design: Application to P-graph research with a focus on sustainability," *Journal of Cleaner Production*, p. 137741, 2023.
- [23] A. Jalilifard, V. F. Caridá, A. F. Mansano, R. S. Cristo, and F. P. C. da Fonseca, "Semantic Sensitive TF-IDF to Determine Word Relevance in Documents," in *Advances in Computing and Network Communications*, Singapore, S. M. Thampi, E. Gelenbe, M. Atiquzzaman, V. Chaudhary, and K.-C. Li, Eds., 2021// 2021: Springer Singapore, pp. 327-337.
- [24] scikit-learn. "sklearn.feature_extraction.text.TfidfVectorizer — scikit-learn 1.4.0 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html (accessed January 16, 2024).
- [25] "KeyWords Plus generation, creation, and changes." Clarivate. https://support.clarivate.com/ScientificandAcademicResearch/s/article/KeyWords-Plus-generation-creation-and-changes?language=en_US (accessed 2024, January 15).
- [26] J. Zhang, Q. Yu, F. Zheng, C. Long, Z. Lu, and Z. Duan, "Comparing keywords plus of WOS and author keywords: A case study of patient adherence research," *Journal of the association for information science and technology*, vol. 67, no. 4, pp. 967-972, 2016.