GEAR-Up: Generative AI and External Knowledge-based Retrieval Upgrading Scholarly Article Searches for Systematic Reviews

Kaushik Roy*[1], Vedant Khandelwal*[1], Valerie Vera[2], Harshul Surana[1], Heather Heckman[2], Amit Sheth[1]

[1]Artificial Intelligence Institute, University of South Carolina
[2]Library Sciences, University of South Carolina

Abstract

This paper addresses the time-intensive nature of systematic reviews (SRs) and proposes a solution leveraging advancements in Generative AI (e.g., ChatGPT) and external knowledge augmentation (e.g., Retrieval-Augmented Generation). The proposed system, GEAR-Up, automates query development and translation in SRs, enhancing efficiency by enriching user queries with context from language models and knowledge graphs. Collaborating with librarians, qualitative evaluations demonstrate improved reproducibility and search strategy quality. Access the demo at https://youtu.be/zMdP56GJ9mU.

Introduction

The proliferation of artificial intelligence (AI) technologies, e.g., Microsoft's recent integration of ChatGPT into its Bing search architecture, showcases AI's immense potential for shaping information search and discovery for educational purposes. For example, researchers at higher education institutions often need to review various subjects and topics systematically. For this, the researchers consult expert librarians trained in finding and evaluating the information in university libraries. AI systems have the potential to assist librarians through the systematic review process. Moreover, if developed responsibly and with input from librarians, such systems could help alleviate significant employee availability concerns (e.g., time and bandwidth limitations) (Borah et al. 2017; Bullers et al. 2018). We demonstrate a pipeline to assist librarians with structured, systematic review search processes. Our method is modular and can be described as follows:

(1) *Query Expansion Module* - processes queries in natural language and attaches additional context by querying external knowledge graphs and pretrained language models (Kawintiranon and Singh 2021).

(2) *Additional Related Query Generation Module* - Our system then uses this expanded query to prompt ChatGPT, generating queries related to the original input query (Lewis et al. 2020).

(3) *Article Search and Retrieval Module* With this list of queries, we first obtain a list of articles using PubMed searches and implement a FAISS-powered retriever that narrows the search down to the most relevant articles (Title, abstracts, and relevant passages) (Komeili, Shuster, and Weston 2021).

*These authors contributed equally.

The Systematic Review Process

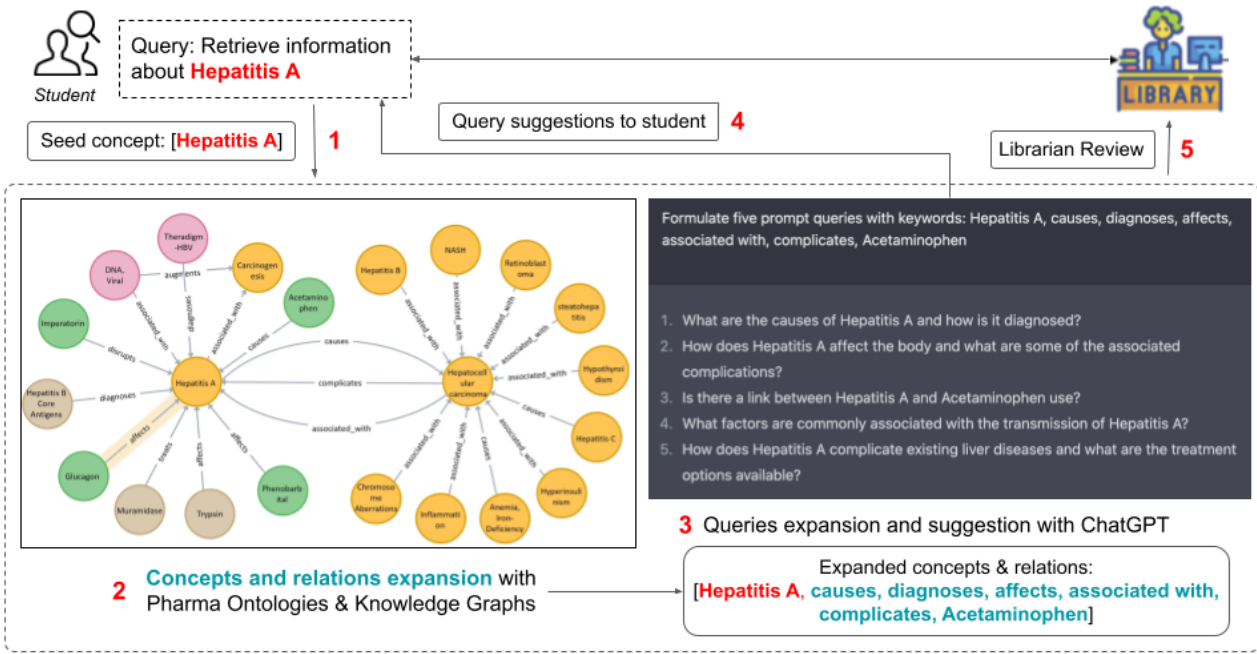
The review process involves the following steps (Koffel 2015):

(i) Defining the research question: Formulating a clear, well-defined research question of appropriate scope. Often, a researcher needs help to define a research problem precisely and interacts with the expert librarian for help with this effort. *(ii) Developing a review protocol/criteria:* This step is often carried out in parallel with the first step and results in defining the terminology and topics that inform the development of the research question. *(iii) Developing inclusion and exclusion criteria:* The student needs to understand and determine whether the review will include a particular study. For this, they provide well-defined inclusion-exclusion criteria.

Steps *(i), (ii), and (iii)* correspond to identifying the issue, determining the question, and writing a plan for the review (protocol): First, the research question is formulated in the "Identify the issue and determine the question" phase. Second, the review protocol is determined, i.e., the related articles to include based on query concepts and their relation to other concepts. This protocol is used to obtain a targeted search query. Lastly, the targeted search query is executed, the relevant data are extracted, the quality is assessed, and a systematic review is compiled and disseminated. Our proposed system seeks to automate the first two steps of issue identification and review protocol-based targeted query formulation. This functionality is carried out by modules (1) and (2) introduced in Section . The remaining steps involve searching a database and using existing machine learning tools to help with the later stages, including article screening, data extraction, and the risk of bias assessment. This is carried out by module (3) in Introduction.

Details of Our Implementation and Evaluation

Due to space concerns, we ask the reader to refer to the caption of Figure 1 for details and illustration of our system and its submodules (introduced in Introduction)



Knowledge-guided query construction with ChatGPT

Figure 1: (1) We use natural language processing tools to obtain seed concepts from the input query. In the figure example, the seed concept is Hepatitis A. (2) We query KGs (e.g., PubMed) using the seed concepts in the input query to obtain additional context. The figure shows subgraphs of concepts connected to the seed concept Hepatitis A and the relationships that connect them. We also query pretrained masked language models to obtain additional context terms. The figure shows the relationships and concepts obtained for the seed concept of Hepatitis A as causes, diagnoses, affects, associated with, complicates, and Acetaminophen. Steps (1) and (2) form the *Query Expansion Module* from Section . (3) The concepts and relationships from the previous steps are fed into ChatGPT with an appropriate prompt to obtain a set of reformulated queries. For example, the prompt in the figure is "Formulate five prompt queries with the keywords: causes, diagnoses, affects, associated with, complicates, Acetaminophen". The figure also shows the reformulated queries that ChatGPT generates, e.g., "What are the causes of Hepatitis A and how is it diagnosed?" Step (3) forms the *Additional Related Query Generation Module* in Section . In Step (4), With this list of queries, we first obtain a list of articles using PubMed searches and implement a FAISS-powered retriever that narrows the search to the most relevant articles, including the article titles, abstracts, and relevant passages. This forms the *Article Search and Retrieval Module* in Section . Finally, the list of retrieved outputs is presented to the librarian for feedback-based refinements to the overall system.

What does the Librarian Review? The librarians' review can be incorporated at several points in the system's execution. The librarian's suggestions can be used to refine the *Additional Related Query Generation Module* to extract more contextual and relevant responses, without which ChatGPT might generate irrelevant and incoherent queries. The librarian may also analyze the *safety* of the generated queries. For example, the external knowledge contains information about how acetaminophen could be misused. The system may then incorporate controls (via KG paths) to avoid showing information due to sensitive content. Similar controls can be placed to extend safety, including the relevant ethics and bias issues. Note that safety is context-sensitive. For example, it may be appropriate for addiction researchers to learn how a drug is abused (e.g., through higher doses, snorting, etc.). These aspects can be incorporated in the *Query Expansion Module*.

Conclusion

We demonstrate a system for assisting librarians in the SR process. It inputs natural language queries and leverages external knowledge to return a set of relevant research articles relevant to the review. We evaluated the retrieved articles with an in-house librarian. Our system shows favorable reviews for reducing the librarian burden by providing high-quality articles like a human librarian.

Acknowledgements

This research was supported in part by NSF Award 2335967 "EAGER: Knowledge-guided neurosymbolic AI" with guardrails for safe virtual health assistants". Opinions are those of the authors and do not reflect the opinions of the sponsor. Special thanks to Ritvik Garimella for their contributions to the user interface.

References

Borah, R.; Brown, A. W.; Capers, P. L.; and Kaiser, K. A. 2017. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ open*, 7(2): e012545.

Bullers, K.; Howard, A. M.; Hanson, A.; Kearns, W. D.; Orriola, J. J.; Polo, R. L.; and Sakmar, K. A. 2018. It takes longer than you think: librarian time spent on systematic review tasks. *Journal of the Medical Library Association: JMLA*, 106(2): 198.

Kawintiranon, K.; and Singh, L. 2021. Knowledge enhanced masked language model for stance detection. In *Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: human language technologies*, 4725–4735.

Koffel, J. B. 2015. Use of recommended search strategies in systematic reviews and the impact of librarian involvement: a cross-sectional survey of recent authors. *PloS one*, 10(5): e0125931.

Komeili, M.; Shuster, K.; and Weston, J. 2021. Internet-augmented dialogue generation. *arXiv preprint arXiv:2107.07566*.

Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.-t.; Rocktäschel, T.; et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33: 9459–9474.