

A Healthcare Knowledge Graph-based Approach to Enable Focused Clinical Search

Maulik R. Kamdar, Will Dowling, Michael Carroll, Cailey Fitzgerald, Sujit Pal, Steve Ross, Katie Scranton, Dru Henke, and Mevan Samarasinghe

Elsevier Health and Commercial Markets, Philadelphia, PA
{m.kamdar, w.dowling, m.carroll, c.fitzgerald, sujit.pal, s.ross.1, k.scranton, dru.henke, m.samarasinghe}@elsevier.com

1 Introduction

For the diagnosis, prognosis, and treatment of their patients, clinicians need access to accurate, succinct, updated, and trustworthy information, provided by renowned medical organizations and disseminated through patient guidelines, medical textbooks, journals, and synoptic overviews. Medical specialists must search and synthesize information on focused, yet esoteric, questions from a broad set of literature sources (textbooks, guidelines, journal articles) in the course of a busy practice using search engines (e.g., ClinicalKey – <https://www.clinicalkey.com/>). There are several barriers for clinicians to do focused clinical search queries at the point of care (e.g., “*Drug for condition X?*”, “*Cause of symptom Y?*”): growth and evolution of medical knowledge, insufficient time, unanswered questions, search on patient comorbidities, lack of awareness on which resource to search, and lack of trust around quality of search results [2].

Advanced technological solutions that automate the search and retrieval of the right excerpts from a corpus of trusted medical literature sources in the right context for such questions are critical for the practice of medicine and patient care. Knowledge graphs can be effective tools to address diverse search and knowledge inference problems in several domains, including healthcare and biomedicine [5, 3]. We present our research and development on a Focused Clinical Search Service (HGFCS), powered by the Elsevier Healthcare Knowledge Graph (termed HG henceforth), that interprets the intent behind focused clinical search queries and retrieves relevant, updated, and trusted medical content from a diverse corpus of medical literature sources.

2 Methods

HG is a knowledge platform built to power diverse content discovery and clinical decision support applications [4, 1]. HG includes knowledge and data from heterogeneous healthcare sources about diseases, drugs, findings, guidelines, cohorts, journals, and books. As of August 2021, HG consists of more than 400,000

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

medical concepts, 1.5 million medical term labels for these concepts, and more than 8 million hierarchical and associative relations of specific relation types. Subject matter experts (SMEs) regularly update medical knowledge in HG using novel exploration interfaces. Excerpts from medical literature are tagged with HG concepts and relations by SMEs or by automated NLP models [4], and are ingested into HG daily through automated pipelines.

Given a clinical search query (e.g., *ketamine for pain dosage, pancreatic pseudocyst management*), the HGFCS parses and identifies the set of medical concepts and their semantic types from HG, as well as relevant medical phrases (e.g., dosage) or cohorts (e.g., pregnancy), in that query. The parser uses HG concept embedding vectors (i.e., representation of concepts in a geometric space) and the HG hierarchy and labels to identify similar concepts and to correct misspelled words for query expansion. Medical phrases are mapped to structural elements in literature sources and HG relation types. The natural language query is rewritten to a structured query using the identified concepts and relation types, which is then executed against multiple HG indexes through a federated querying infrastructure to retrieve the right excerpt in a medical document. Search performance is measured over multiple sets of real-world clinical queries.

3 Conclusion and Lessons Learned

Structured representation of medical content in HG improves performance in focused clinical search over conventional text-based search methods. Moreover, the use of concept embedding vectors, in conjunction with HG semantics, further improved the search performance significantly, since we were able to parse clinical queries more intelligently. Using automated pipelines to regularly update concepts, labels, and relations within HG from several sources and to tag medical excerpts, we can ensure that clinicians are able to retrieve accurate, recent, succinct, and trustworthy medical content. This research signifies the importance of a scalable healthcare knowledge graph ecosystem and the use of machine learning and knowledge representation methods for focused clinical search.

References

1. DeJong, A., et al.: Elsevier’s healthcare knowledge graph and the case for enterprise level linked data standards. In: Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks 2018. (2018)
2. Del Fiol, G., et al.: Clinical questions raised by clinicians at the point of care: a systematic review. *JAMA internal medicine* **174**(5), 710–718 (2014)
3. Kamdar, M.R., et al.: Enabling web-scale data integration in biomedicine through linked open data. *NPJ digital medicine* **2**(1), 1–14 (2019)
4. Kamdar, M.R., et al.: Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings. *AMIA Summits on Translational Science Proceedings* p. 288 (2020)
5. Noy, N., et al.: Industry-scale knowledge graphs: lessons and challenges. *Communications of the ACM* **62**(8), 36–43 (2019)