Criteria2Query 3.0 Powered by Generative Large Language Models

Jimyung Park[1], Yilu Fang[1], Chunhua Weng[1]
[1] Department of Biomedical Informatics, Columbia University, New York, New York, USA

Background

The latest advances in generative large language models (LLM) such as ChatGPT have exhibited stunning performance on generative language tasks.[1] More than language generation, generative LLMs have proven their competency over encoder-based language models (i.e., BERT) for information extraction tasks.[2] Furthermore, their application in clinical natural language processing (NLP) has shown the potential of LLMs to revolutionize the biomedical research. Criteria2Query (C2Q) is an NLP pipeline for automating the translation of clinical trial eligibility criteria into executable cohort queries formatted using the Observation Medical Outcome-common data model (OMOP-CDM).[3] C2Q 2.0 enhances user experience by combining human-computer intelligence and demonstrated its usability by clinical research staffs.[4, 5] After seeing the language understanding and generation abilities of LLMs, we hypothesized that LLM could address the existing limitations in C2Q 2.0. Therefore, we present C2Q 3.0, which integrates LLM within its architecture to further enhance eligibility criteria parsing and user experience.

Methods

Overall workflow of C2Q 3.0 is presented in Figure 1. Likewise previous version, current version uses OHDSI's Usagi for concept mapping process.[6] Users can interactively modify parsed clinical concepts and their corresponding mappings within C2Q 3.0. We implemented 'Criteria Reasoning' view to exhibit GPT's concept reasoning and explanation on the eligibility criteria (Figure 2).

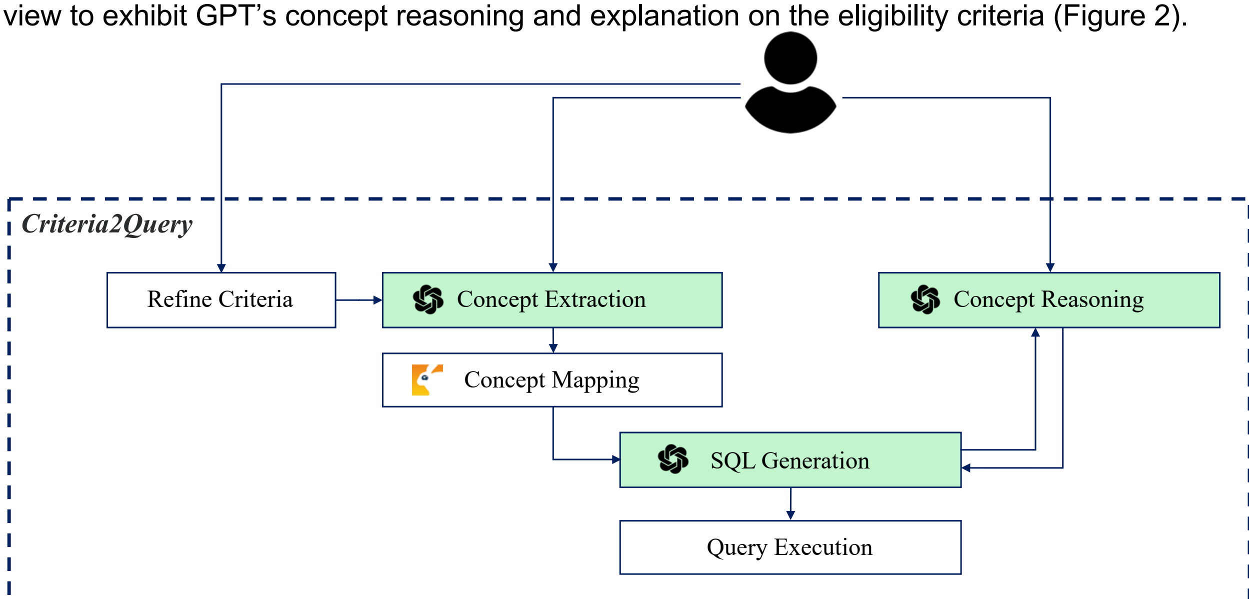

Figure 1. Workflow of Criteria2Query 3.0. GPT-prompts are implemented in the system: 1) Concept Extraction, 2) SQL Generation, and 3) Concept Reasoning. Concept Extraction prompt parses clinical concepts from eligible criteria and SQL generation creates PostgreSQL query based on the parsing results. Concept Reasoning prompt interprets SQL query to provide reasoning information to user. Users can interactively modify the results of Concept Extraction and Concept Reasoning.

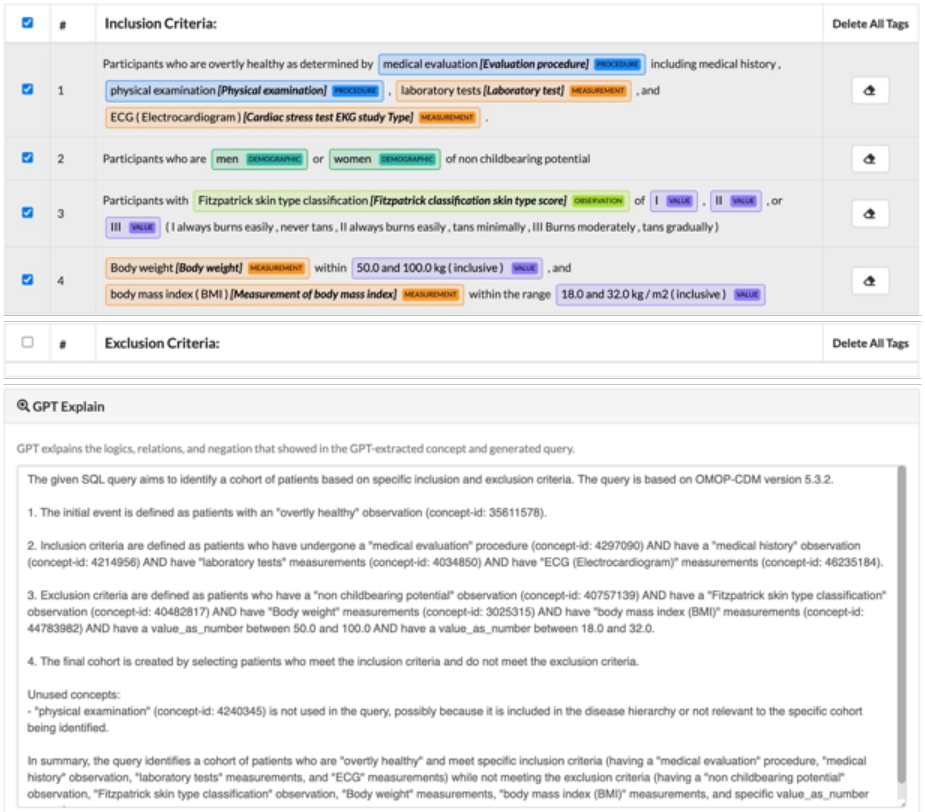

Figure 2. Sample view of Concept Extraction and Concept Reasoning view in the system

C2Q 2.0 integrated variety of NLP functionalities, which included concept extraction, relation extraction, logic analysis, negation detection, temporal/value normalization, and query formulation. In C2Q 3.0, these functionalities were substituted by LLM's three prompts: 1) Concept Extraction, 2) SQL generation, and 3) Concept Reasoning. OpenAI's GPT-4 API was utilized as LLM in the system, and to ensure reproducibility and consistency, temperature was set as 0.0. All prompts were designed as few-shot prompts which employ a single example to optimize model performance (Figure 3).

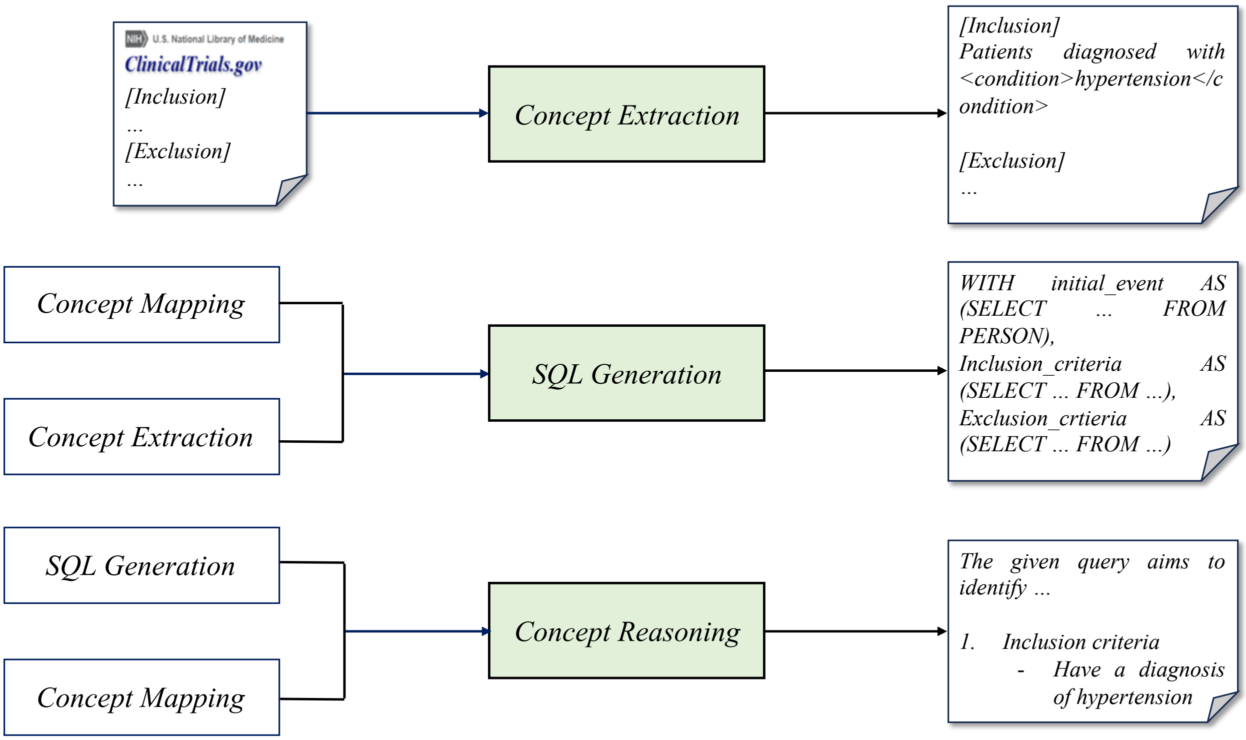

Figure 3. Example of input/output data of prompts in Criteria2Query 3.0. Concept Extraction prompt annotates clinical concepts with OMOP's domain category. SQL Generation prompt generates using parsed and mapped concepts PostgreSQL query. Concept Reasoning prompt generates reasoning on GPT-generated query by inclusion and exclusion criteria.

In Concept Extraction, prompt was designed to extract and annotate clinical concepts as OMOP's domain categories (e.g., demographic, condition, drug, observation, measurement, procedure, and device) and its attribute (value, temporal, visit, and negation). The other information extraction tasks such as relation extraction, logic analysis, negation detection, and temporal/value normalization are incorporated within the task definition in the Concept Extraction prompt. The definitions for each OMOP domain are also incorporated within the prompt. However, if there are concepts that are not conform to the OMOP's domain, we permit LLM to define a new category.
In SQL generation, we designed the prompt to generate PostgreSQL query using Common Table Expression to reduce possible syntax errors. OMOP-CDM version 5.3.2 was designated as target version. The source eligibility criteria and the list of extracted concepts with standardized concept mapping were provided as standardized concept-IDs in the query. To prevent errors occurring during query execution, we implemented the Chain-of-Thoughts approach to automatically update errors and re-execute the query. Concept Reasoning prompt is designed to elucidate the logic, temporal/value attributes, and relations between clinical concepts in the GPT-generated SQL query. Since the query only has concept-IDs, the mapping information coupled with the IDs is provided for the translation into concept-names. The output of Concept Reasoning is formulated as a narrative description, designed to present the reasoning and logic effectively.

Conclusion

This study demonstrated the design and architecture of C2Q 3.0 which leverages generative LLM for concept extraction, SQL generation, and concept reasoning. The three prompts replaced both traditional NLP models and encoder-based language models in the previous version of the system. Though evaluation on each prompt is undergoing, we observed LLM can be effective in processing complex eligible criteria. Future research will be explicitly evaluated LLM performance on the system.

References

1. Introducing ChatGPT. https://openai.com/blog/chatgpt (accessed June 13, 2023)
2. Wang S, Sun X, Li X, Ouyang R, Wu F, Zhang T, Li J, Wang G. Gpt-ner: Named entity recognition via large language models. arXiv preprint arXiv:2304.10428. 2023 Apr 20.
3. Yuan C, Ryan PB, Ta C, Guo Y, Li Z, Hardin J, Makadia R, Jin P, Shang N, Kang T, Weng C. Criteria2Query: a natural language interface to clinical databases for cohort definition. Journal of the American Medical Informatics Association. 2019 Apr;26(4):294-305.
4. Fang Y, Idnay B, Sun Y, Liu H, Chen Z, Marder K, Xu H, Schnall R, Weng C. Combining human and machine intelligence for clinical trial eligibility querying. Journal of the American Medical Informatics Association. 2022 Jul;29(7):1161-71.
5. Idnay B, Fang Y, Dreisbach C, Marder K, Weng C, Schnall R. Clinical Research Staff Perceptions on a Natural Language Processing-driven Tool for Eligibility Prescreening: An Iterative Usability Assessment. International Journal of Medical Informatics. 2023 Jan 6:104985.
6. OHDSI Usagi. https://github.com/OHDSI/Usagi (accessed June 13, 2023)