

Enhancing Patient Empowerment and Health Literacy: Integrating Knowledge Graphs with Language Models for Personalized Health Content Delivery

Christine KAKALOU^{a,b,1}, Christina KARAMANIDOU^b, Theodore DALAMAGAS^c
and Manolis KOUBARAKIS^a

^a*Department of Informatics and Telecommunications, National and Kapodistrian
University of Athens, Athens, Greece*

^b*Psychology Lab, Institute of Applied Biosciences, Centre for Research & Technology
Hellas, Thessaloniki, Greece*

^c*Information Management Systems Institute, ATHENA Research Center, Marousi,
Greece*

ORCID ID: Christine Kakalou <https://orcid.org/0000-0001-9272-410X>, Christina
Karamanidou <https://orcid.org/0000-0002-2849-3690>, Theodore Dalamagas
<https://orcid.org/0000-0002-5002-7901>, Manolis Koubarakis <https://orcid.org/0000-0002-1954-8338>

Abstract. Health literacy empowers people to access, understand and apply health information to effectively manage their own health and to be an active participant in healthcare decisions. In this paper we propose a conceptual model for cognitive factors affecting health literacy and related socioeconomic aspects. Then we develop the HEALIE Knowledge Graph to represent the model, drawing from various medical ontologies, resources, and insights from domain experts. Finally, we combine the Knowledge Graph with a Large Language Model to generate personalised medical content and showcase the results through an example.

Keywords. Knowledge Graphs, Natural Language Generation, Retrieval Augmented Generation, Patient Empowerment, Health Literacy

1. Introduction

Patient empowerment is strongly affected by health literacy, i.e., the patient's capacity to navigate healthcare systems, make informed decisions, and engage effectively in promoting personal and community health [1]. Even healthcare and wellness apps and printed medical informational material contain content that almost never meets the minimum readability and comprehensibility requirements. This paper focuses on personalised medical content delivery to support patient empowerment. The aim is to assist patients to better grasp the relevance of the content to their specific health conditions, lifestyle, and background (e.g. problem-solving skills, resilience, emotional regulation, control, gender, cultural background, income level etc.). To this end, we propose and develop **HEALIE** (**HEAL**th **Information Enhancement**), a Knowledge Graph-enabled, personalised medical content generation system that leverages Large

¹ Corresponding Author: Christine Kakalou; E-mail: ch-kakalou@di.uoa.gr.

Language Models (LLMs). The main contributions of this work are a) a **conceptual mapping approach** to integrate cognitive competencies and socio-economic determinants of health literacy, building upon a widely adopted health literacy model, b) the design and development of the **HEALIE Knowledge Graph (KG)**, c) methods for the **retrieval of necessary information from the KG** (factors relevant to health literacy that affect medical content generation) and d) both **template-based** and **Large Language Model (LLM)** enabled **content generation** through an example.

2. The HEALIE model

We build on the integrated conceptual model established by Sørensen and colleagues [2], which consolidates comprehensive evidence-based dimensions of health literacy. The model's four dimensions cover four fundamental aspects for effective health literacy: *Accessing, Understanding, Appraising, and Applying* information. In this work, we focus only on mapping the *Understand* (comprehending and processing health knowledge effectively) and *Appraise* dimension (critical assessment and evaluation of information for relevance, reliability, and credibility), since the other two dimensions involve circumstances out of scope for our work. Our model adds specific cognitive factors to build upon Sørensen's model, which does not provide a break-down of certain skills to elevate health literacy. Cognitive factors are characteristics of an individual that affect performance and learning. To capture more fine-grained health literacy elements, the Psychology Lab team assessed the literature and selected a series of cognitive factors that are connected to each dimension. The importance of each factor has also been ranked and incorporated into the model in the form of weights.

3. The HEALIE Knowledge Graph and Personalised Content Generation

The HEALIE Knowledge Graph, as seen in [Figure 1](#), implements the HEALIE model, providing a multimodal representation of clinical entities, cognitive factors, social determinants of health and elements to determine the generation of personalised health content. For example, for an elderly person with lower health literacy skills, we will replace the term "hematological cancer" to "blood cancer and due to the decline of various memory-related cognitive skills, we will repeat basic information multiple times and offer a bullet-point summary. The KG's development involved: a) *Literature Review and Available Data Resources Mapping.*, b) *Conceptual Model Creation* comprising 102 nodes, 405 properties, 272 relationships) organized in 5 Node Clusters: Clinical Data, Patient Profile, Social Determinants of Health (SDoH), Cognitive Factors, Text Generation. and c) *KG Instatiation*.

Personalised Content Generation. Neo4j was used to store the HEALIE KG and navigate it in order to extract the appropriate information for each use case. Initially, for the automatic content creation, we used a template-based NLG software (RosaeNLG) that uses predefined text templates, bespoke rules and placeholders where the tailored information from the KG is inserted. Aiming for a more sophisticated content output, we switched to Retrieval Augmented Generation (RAG), an Artificial Intelligence framework (Langchain) that combines LLMs (ChatGPT-4) and external knowledge bases to improve the quality and accuracy of the generated content.

4. Pathways for Patient Understanding: Otto’s Use Case

4.1. Graph Processing and Content Generation

A RAG framework is used to process the HEALIE KG, gather the relevant information and finally automatically create personalized content based on templates provided by domain experts, throughout a series of distinct steps. First we create a system that can understand a user’s natural language question, turn it into a query and run it within the KG. Generating accurate KG queries from natural language, particularly for complex graph structures, involves several challenges discussed in Section 6.

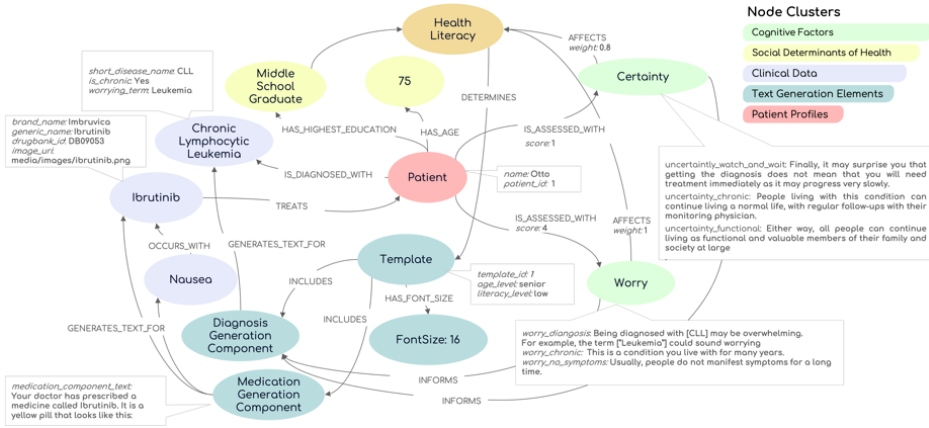


Figure 1. Example of Paths in the HEALIE Knowledge Graph.

Thus we provide the language model with detailed information about the graph’s structure and possible query scenarios (Figure 2a). Using a structured script, we ask natural language questions to extract key health-related information from the HEALIE KG, aiming to get comprehensive insights into the patient’s disease details, cognitive assessments, and socio-economic factors affecting health literacy. After retrieving responses from the KG (Figure 2b), the system combines this information into a detailed text. Then, based on a template created by domain experts, the system crafts a final prompt for the ChatOpenAI model to generate accessible and informative health content (Figure 2c). The generated content explains Otto’s medical condition in simple terms and provides practical tips for managing his disease, as HEALIE is designed to empower patients in their healthcare journey. The expert-designed templates provide a strong foundation, showing the model’s ability not just in generating queries but also in creating structured and informative responses to user questions.

4.2. Otto’s case (CLL older patient, low health literacy)

Patient profile. Otto, a 75-year-old retiree, is a middle school graduate who is diagnosed with Chronic Lymphocytic Leukemia with high levels of worry and uncertainty. Otto is handed out a leaflet for his condition to read in the comfort of his own home.

KG processing to identify personalized information. First, we gather basic details about the CLL disease Otto has and his cognitive scores. Otto has a high *Worry* score and a low *Certainty* score. Experts recommend adding specific sentences to ease Otto’s distress and help him understand better. Lastly, his health literacy level, age, and education determine the content’s structure and style, such as using a larger font size for senior patients.

```

cypher_generation_template > ""
Task: Generate Cypher query for a Neo4j graph database.
Instructions: Use only the provided relationship types and properties in the schema. Do not use any other relationship types or properties that are not provided.
Schema: (schema)
Note:
Do not include any explanations or apologies in your responses. Do not respond to any questions that might ask anything other than for you to construct a Cypher statement. Do not include any text except the generated Cypher statement. Make sure the direction of the relationships is correct in your queries. Make sure you alias both entities and relationships properly. Do not run any queries that would add to or delete from the database. Make sure to alias all statements that follow as with statement (e.g. WITH d as disease, d.disease_description as disease_description). If you need to divide numbers, make sure to filter the denominator to be non zero.
Examples:
# What is the full disease name, the disease description as well as the short disease name
MATCH (p:Patient{name:Mayo})-[:r_IS_DIAGNOSED_WITH]-> (dis:Condition)
RETURN dis.full_name, dis.disease_description, dis.short_name
LIMIT 1
a)

# Generate new Graph(Cypher)Chain chain...
Generated Cypher:
MATCH (p:Patient {name: "Otto"})-[:r_IS_DIAGNOSED_WITH]-> (dis:Condition)
RETURN c.full_name AS FullDiagnosis, c.short_name AS ShortDiagnosis
Full Context:
[{"FullDiagnosis": "Chronic Lymphocytic Leukemia",
"ShortDiagnosis": "CLL"}]
b)

# Prepare the final prompt for the chat model with content included
final_prompt = "You job is to synthesize the information in the context below into a comprehensive and coherent summary suitable for explaining the diagnosis to the patient. Write a cohesive and informative text addressed to the patient, in the form of an informational leaflet. It should be understandable and informative. Use the following context to create the text. Be as detailed as possible, but don't make up any information that's not from the context. If you can't know an answer, say you don't know. In the end of the text, give a couple of bullet points as key takeaways. Include some of the tips for living with the disease. [context]"
c)
    
```

Figure 2. a) Instructions, KG schema illustration and query examples (few-shot prompting), b) Generated query for "What is Otto's diagnosis, disease name, description and character, c) Final prompt combining all the results

Text generation through dynamic content. The queries add dynamic elements to our text generation, creating detailed narratives. Unlike fixed text templates, our new LLM approach turns KG queries into nuanced, context-rich text. **Personalized vs generic medical content.** For comparison, in Figure 3 three versions are juxtaposed: a) HEALIE's LLM generated tailored content, b) HEALIE's template-based tailored content c) a generic text from Mayo Clinic's website.

a) LLM Generated Patient-Friendly Content Tailored to Otto's needs

You have been diagnosed with Chronic Lymphocytic Leukemia (CLL), a type of cancer that affects the lymphocytes in your immune system. These abnormal cells can build up in your blood and bone marrow, leading to symptoms such as fatigue, swollen lymph nodes, and frequent infections. Treatment options for CLL may include chemotherapy, targeted therapy, and stem cell transplant. It is important to work closely with your healthcare team to determine the best course of action for managing your CLL.

It is important to note that CLL is a chronic condition, meaning that it can be lived with for many years. In some cases, immediate treatment may not be necessary due to the slow progression of the disease. This approach, known as "watch and wait," involves regular follow-ups with your monitoring physician to track the progression of the disease. Despite uncertainties, individuals with CLL can still lead normal lives and continue to be valuable members of their family and society.

Key takeaways:

- CLL is a type of cancer that affects the lymphocytes in your immune system.
- Treatment options may include chemotherapy, targeted therapy, and stem cell transplant.
- "Watch and wait" approach may be used for managing CLL, with regular follow-ups with your healthcare team.

Tips for living with CLL:

- Stay informed about your condition and treatment options.
- Maintain a healthy lifestyle with regular exercise and a balanced diet.
- Seek support from family, friends, and support groups to help you cope with the emotional aspects of living with CLL.

Remember, you are not alone in this journey. Your healthcare team is here to support you every step of the way.

b) Template Generated Patient-Friendly Content Tailored to Otto's needs

You have been diagnosed with Chronic Lymphocytic Leukemia, or CLL for short. Being diagnosed with CLL may be overwhelming. For example, the term "Leukemia" could sound scary. However, contrary to acute forms of leukemia, CLL is chronic. Meaning, it is a condition you live with for many years.

CLL is the most common type of blood cancer in adults. It is important to acknowledge that people living with CLL can continue leading a normal life, with regular follow-ups with their monitoring physician. CLL can vary from person to person. It may take various forms, each of which can have a different impact on everyday life. CLL develops very slowly for some people. This means CLL may remain stable for years, without major symptoms or treatment. Other people may develop bothersome symptoms and may need treatment earlier. Either way, all people can continue living as functional and valuable members of their family and society at large. Keeping that in mind some information about CLL is that it usually affects older adults; however, it can also present in younger adults. Specifically, it greatly increases the number of B-lymphocytes in the blood. Even though the cause of CLL are not completely clear, one thing is for certain: it is not transmissible through every-day contact. Finally, it may surprise you that getting the diagnosis does not mean that you will need treatment immediately as it may progress very slowly. In fact, the majority of people with CLL won't receive treatment right after diagnosis.

c) simple-language content, as provided on Mayo Clinic's website.

Chronic lymphocytic leukemia (CLL) is a type of cancer of the blood and bone marrow -- the spongy tissue inside bones where blood cells are made.

The term "chronic" in chronic lymphocytic leukemia comes from the fact that this leukemia typically progresses more slowly than other types of leukemia. The term "lymphocytic" in chronic lymphocytic leukemia comes from the cells affected by the disease -- a group of white blood cells called lymphocytes, which help your body fight infection. Chronic lymphocytic leukemia most commonly affects older adults. There are treatments to help control the disease.

Figure 3. a) LLM Generated Patient-Friendly Content Tailored to Otto's needs b) Template Generated Patient-Friendly Content Tailored to Otto's needs, c) simple-language content, as provided on Mayo Clinic's website.

5. Related Work

A multitude of healthcare Knowledge Graphs have been developed, addressing various applications in healthcare, but also patient education focused on disease management [3], health promotion, lifestyle adjustments, medical Q&A services, alongside personalized health-related Knowledge KGs [4] that recommend information based on patient-stated preferences, fundamental literacy skills, and demographic divisions. Work has also been achieved towards representing personality and cultural characteristics in KGs [5]. Our approach distinguishes itself by considering more profound elements affecting a patient's comprehension and engagement with the provided content, including cognitive competencies, socio-economic factors, and cultural insights. Natural language generation methods have been employed to generate medical and patient explanatory materials [6,7]. Following the introduction of LLMs, a surge of health-related generative-AI solutions have emerged, lacking however in depth, accuracy and reliability [8,9].

6. Conclusion and Further Work

The HEALIE model leverages health psychology to help healthcare professionals provide personalized and empowering medical information to patients, which would be otherwise unrealistic to produce manually at a large scale. Our unique contribution to health informatics is developing a Knowledge Graph that combines cognitive and socioeconomic factors with natural language generation. One current limitation is the clinical information that relies heavily on existing datasets not developed for HEALIE's purposes. Cognitive and literacy level-appropriate information is curated by the domain experts - a slow and difficult process to scale. The cultural competence dimension of the current KG version is limited and lacks nuance. In developing our HEALIE RAG process, we navigated significant challenges in dynamically generating Cypher queries from natural language inputs, specifically when incorporating node and relationship properties directly within queries. Our exploration of the available techniques revealed limitations in template parsing and the complexity of mapping nuanced language to specific graph elements. However, the specificity of the instructions, schema illustration, iterative refinement of the prompt templates and inclusion of query examples and node property categorizations, greatly boosted the results. HEALIE's next iterations will focus on developing a user-friendly interface to process natural language requests. The HEALIE system will also be validated using benchmarking, evaluation metrics, and validation tools, tailored to the patient-facing healthcare domain [10]. Finally, future work could focus on validating the KG's factors and relationships in real-life scenarios with actual patients, followed by refining the weights in HEALIE's KG based on empirical findings.

References

- [1] Hickmann E, Richter P, Schlieter H. All together now – patient engagement, patient empowerment, and associated terms in personal healthcare. *BMC Health Services Research*. 2022 Sep 2;22(1):1116. doi: 10.1186/s12913-022-08501-5
- [2] Sorensen K, Van Den Broucke S, et al. Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health*. 2012;12(1). doi: 10.1186/1471-2458-12-80
- [3] Li L, Wang P, et al. Real-world data medical knowledge graph: construction and applications. *Artif Intell Med*. 2020 Mar;103:101817. doi: 10.1016/j.artmed.2020.101817
- [4] Skjæveland MG, Balog K, Bernard N, Lajewska W, Linjordet T. An Ecosystem for Personal Knowledge Graphs: A Survey and Research Roadmap. Rochester, NY; 2023. doi: 10.48550/arXiv.2304.09572
- [5] Awantee V, Deshpande, Dana Ruiters, Marius Mosbach, D. Klakow. StereoKG: Data-Driven Knowledge Graph Construction for Cultural Knowledge and Stereotypes. *WOAH*. 2022; doi: 10.48550/arxiv.2205.14036
- [6] Bala S, Keniston A, Burden M. Patient Perception of Plain-Language Medical Notes Generated Using Artificial Intelligence Software: Pilot Mixed-Methods Study. *JMIR Form Res*. 2020 Jun 5;4(6):e16670.
- [7] Lambert R, Choo ZY, Gradwohl K, Schroedl L, Luzuriaga ARD. Assessing the Application of Large Language Models in Generating Dermatologic Patient Education Materials According to Reading Level: Qualitative Study. *JMIR Dermatology*. 2024 May 16;7(1):e55898. doi: 10.2196/55898
- [8] Roster K, Kann RB, Farabi B, Gronbeck C, Brownstone N, Lipner SR. Readability and Health Literacy Scores for ChatGPT-Generated Dermatology Public Education Materials: Cross-Sectional Analysis of Sunscreen and Melanoma Questions. *JMIR Dermatol*. 2024 Mar 6;7:e50163. doi: 10.2196/50163
- [9] Nielsen JPS, von Buchwald C, Grønhoj C. Validity of the large language model ChatGPT (GPT4) as a patient information source in otolaryngology by a variety of doctors in a tertiary otorhinolaryngology department. *Acta Otolaryngol*. 2023 Sep;143(9):779–82. doi: 10.1080/00016489.2023.2254809
- [10] Abbasian M, Khatibi E, Azimi I, Oniani D, Shakeri Hossein Abad Z, Thieme A, Sriram R, Yang Z, Wang Y, Lin B, Gevaert O, Li LJ, Jain R, Rahmani AM. Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI. *npj Digit Med*. 2024 Mar 29; doi: 10.1038/s41746-024-01074-z