The CEKG: A Tool for Constructing Event Graphs in the Care Pathways of Multi-Morbid Patients⋆

Milad Naeimaei Aali[1,*], Felix Mannhardt[2] and Pieter Jelle Toussaint[1]

[1]*Norwegian University of Science and Technology, Trondheim, Norway*
[2]*Eindhoven University of Technology, Eindhoven, Netherlands*

Abstract

One of the challenges in healthcare processes, especially those related to multi-morbid patients who suffer from multiple disorders simultaneously, is not connecting the disorders in patients to process events and not linking events' activities to globally accepted terminology. Addressing this challenge introduces a new entity to the clinical process. On the other hand, it facilitates that the process is interpretable and analyzable across different healthcare systems. This paper aims to introduce a tool named CEKG that uses event logs, diagnosis data, ICD-10, SNOMED-CT, and mapping functions to satisfy these challenges by constructing event graphs for multi-morbid patients' care pathways automatically.

Keywords

Healthcare, Process mining, Event knowledge graph, Multi-morbid patients

Metadata description	Value
Tool name	CEKG
Current version	1.0
Legal code license	Apache 2.0
Languages, tools and services used	Python, JavaScript, CSS, HTML, Django, Channel, GraphViz
Supported operating environment	Microsoft Windows, GNU/Linux, MacOS
Download/Demo URL	https://cekg-db1cc0d27386.herokuapp.com/
Documentation URL	https://github.com/mnaeimaei/ClinicalEventKnowledgeGraphs_Web/blob/main/README.md
Source code repository	https://github.com/mnaeimaei/ClinicalEventKnowledgeGraphs_Web/tree/main/Backend_Scripts
Screencast video	https://drive.google.com/file/d/1vik3K8XJ1LV5xqdb-wwtv3IUQr1nM3rG/view?usp=sharing
Test Dataset	https://github.com/mnaeimaei/ClinicalEventKnowledgeGraphs_Web/tree/main/Dataset

1. Introduction

For both ethical reasons and economic progress, it is essential to foster a society with healthy people. Achieving this vision necessitates the presence of robust and effective healthcare services. These services are crucial to ensure timely medical treatment and preventative care for

all. Patients with multi-morbidity in particular need such care. This patient group, who have multiple chronic conditions at the same time [1], is expanding due to socio-economic deprivation and an aging population [1], and they require coordinated care from various specialties. They also demand significantly more resources due to the complexity of their conditions. Therefore, enhancing healthcare services for such patients can be a cornerstone of achieving truly effective healthcare for all. One of the approaches for enhancing healthcare service for these multi-morbid patients is enhancing the clinical process they are subject to.

A clinical process or care pathway outlines the events involved in diagnosing, treating, managing, and following up with patients [2]. It can be considered a type of business process and, consequently, techniques like process mining may be used to improve a multi-morbid patient's clinical process. Still, there are a lot of challenges when applying process mining methods to the care pathways of multi-morbid patients, which often spread several caregivers in multiple organizations and involve the simultaneous treatment of multiple conditions. Among these challenges are connecting emerging entities to events and linking relevant terminology to them, Addressing both of these challenges may significantly enhance the delivery of care paths [3].

- **Connecting clinical entities.** Emerging clinical entities are clinical attributes that are not connected directly to events but can potentially be used as new entities. For example, multi-morbid patient disorders are connected to the patient entity but not attached to events. Multi-morbid patients have many different (sometimes emerging) disorders, which we see as entities connected to the patient. By connecting events to the relevant disorders, thereby getting a multi-entity event data, we can better query the event data of a patient to find relevant insights[3].
- **Linking terminology.** Standardized clinical coding and nomenclature systems provide useful terminology that are often not linked to the clinical process's activities and entities. For example, sources like Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [7], International Classification of Diseases Clinical Modification (ICD CM) [8], and diagnosis-related groups (DRG) that store event activities and entities terminology in a standardized way are not linked directly to event activities and entities. By aligning terminology with activities and events, we can standardize clinical processes, enabling global interoperability for patient diagnoses and event activities. This may also allow for various levels of abstraction and standardized categorization, ensuring a more organized and segmented process[3].

In this paper we introduce a tool for the Clinical Event Knowledge Graph (CEKG) framework presented in our previous work [3], which addresses the two mentioned challenges. We developed this tool to support constructing CEKGs for enhancing the process analysis of multi-morbid patient care pathways. The tool utilizes inputs such as low-dimensional clinical event logs, diagnostic data (indicating each patient's disorders), ICD codes, and SNOMED CT terminology. It allows to map the different inputs through constrained node mappings, which are functions derived from various sources, including empirical data, domain expertise, professional insights, and documentation.

In Section 2, we describe the innovations of this tool and in Sect. 3 we show its application in a case study.

2. The Overview of the Tool

Since the tool needs to support terminologies such as SNOMED-CT and ICD-10 as parts of its inputs a graph database was chosen for storage since it supports a linked-data structure for these terminologies. Furthermore, the need for path-based traversal of data makes the graph database an ideal choice. Additionally, the tool requires the storage of entity attributes and other semantic patient data, further reinforcing the suitability of a graph database for these functions. The CEKG was proposed using the Labelled Property Graph Model. However, a challenge in creating a CEKG using Neo4J is integrating data from different sources (SNOMED, hospital information systems, etc.), which requires using several complex Cypher Query Language (CQL) queries manually. [4] introduced an open-source Python library for exploring graph-based, object-centric process discovery, but this approach requires the deployment of clinical data and terminology.

In addressing usability challenges, process mining tools frequently exhibit operational complexities, particularly when applied by healthcare professionals managing patients with multi-morbid conditions. To address these difficulties, the CEKG tool incorporates a user interface that is designed, enabling users to rapidly assimilate its functionalities from the initial steps. Furthermore, the tool generates outputs in the LPG format and employs the Graphviz library for visualization purposes.

For the implementation of the tool, as illustrated in Fig. 1, Python with Django and Django Channels was used as the backend framework along with several libraries such as Pandas, Neo4j, and Graphviz. For frontend development, vanilla JavaScript and HTML and CSS were used.

The CEKG tool offers several features for discovering various types of care pathways that integrate both connecting entities and linking terminology:

C1 Independent graphs for each patient without consolidating patient activities.
C2 Combined graphs for patients without consolidating patient activities.
C3 Consolidated patient activities to identify repeated activities for patients with the same multi-morbidity or for specific disorders.
C4 Consolidated patient activities to determine how frequently activities related to the treatment of each disorder are repeated for a group of patients with the same multi-morbidity.
C5 Consolidated Patient Care Pathways to identify the most frequently repeated activities in the treatment of a group of patients with the same multi-morbidity.
C6 Care Pathways that indicate which disorders are treated, untreated, or newly discovered in each admission.

Additionally, we can determine whether to include properties of activities in the graph. For example, should the graph only indicate that a specific clinical test, such as the ABG test, was conducted, or should it also include the test results (e.g., the values of Oxygen, Hemoglobin, ...). Furthermore, we can segment the graph by relating it to the domain or scope of activities.

The tool is designed to handle large datasets, with the primary dataset extracted from the entire MIMIC-IV database. However, it can also be used with alternative datasets. To facilitate testing and provide a template for users to create their own datasets, a test dataset was prepared

by extracting a portion of the MIMIC-IV data. The activity titles and patient IDs were modified for de-identification. This test dataset includes essential ICD codes and SNOMED-CT IDs, although the tool is capable of processing the full range of ICD codes and SNOMED-CT ID databases. For convenience, all data in the test dataset was consolidated into a single spreadsheet. However, for practical use, separate CSV files can be imported into the tool.

For building the clinical event knowledge graph, different steps were defined:

1. Creating a Labeled Property Graph for the Event Log
2. Creating a Labeled Property Graph for Entities Attributes
3. Creating Relationships Between Entities and Entities Attributes
4. Creating a Labeled Property Graph for Activity Attributes
5. Creating a Labeled Property Graph for Activities Domains
6. Creating a Labeled Property Graph for ICD Codes
7. Creating Nodes for SNOMED-CT Concepts
8. Creating Relationships Between Nodes of SNOMED-CT Concepts
9. Connecting Diagnosed Disorders to ICD Codes
10. Connecting ICD Codes to SNOMED-CT ID
11. Connecting Activities to SNOMED-CT Codes
12. Connecting Activities to Domains and Connecting Domains to SNOMED-CT Codes
13. Connecting Events to Disorders and Defining Diagnosed Disorders as New Entities
14. Creating Directly Follows (DF) Relationships and Finalizing the Clinical Event Knowledge Graph
15. Discovering Care Pathways from the Clinical Event Knowledge Graph

At each step, Neo4j queries are automatically generated based on the input dataset to create graphs or establish relationships between two graphs. These queries will adapt if the dataset is changed. Some queries are designed to clear the database, remove or create constraints, create or modify nodes, or establish relationships between nodes. Since the tool sends the queries to the Neo4j Aura Database, it is possible to view the final clinical event knowledge graph within Neo4j Aura. However, the tool also facilitates the creation of the clinical event knowledge graph offline in a local Neo4J instance by providing all the necessary queries for the user.

3. Use cases overview

In this section, we validated the CEKG tool with a case based on the MIMIC-IV dataset [5]. Two patients with multi-morbidity were considered, including only two entities: *PATIENT* and *ADMISSION*. We used the tools to discover care pathways, denoted as C2 and C3, as examples of the care pathways that we can identify. By using the tools, we not only discovered another entity, *Disorder*, but also connected all its activities and entities to ICD-10 and SNOMED-CT to facilitate standardized analysis.

The C2 care pathway discovered from the tool, as shown in Fig. 2, is the dependent care pathways of two multi-morbid patients, consisting of three entities: *PATIENT* with red circles, *ADMISSION* with blue circles, and *Disorder* with green circles. All activities in the process
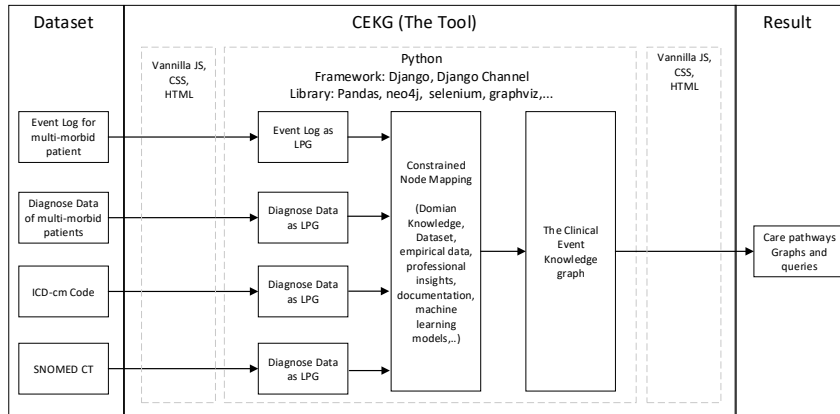
Figure 1: The tool was developed using Python with Django and Django Channels as the backend framework, supplemented by libraries including Pandas, Neo4j, Selenium, and Graphviz. The frontend was created using vanilla JavaScript, HTML, and CSS.
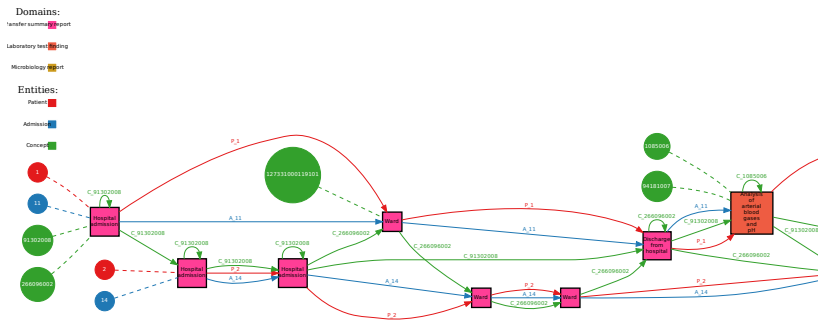

Figure 2: The resulting event graph includes disorder as a new entity. Activities and processes are categorized in a standardized manner, making them interpretable across different healthcare systems.

are mapped to concepts from SNOMED-CT. Additionally, the domains of activities are shown with different colors in the graph. With this type of care pathway, we can determine which activities that happened for these patients are related to which disorders. For example, the graph "Analysis of Arterial blood gases and pH" relates to two disorders with SNOMED-CT 1085006 and 94181007. Furthermore, we can categorize the activities using SNOMED-CT concepts.

The C3, as shown in Fig. 3, is the identification of the most frequent activities in the treatment of two patients. For example, we can find out how many times the "Microbiology Procedure" happened after the "Analysis of Arterial Blood Gases and pH" for these two patients. Using SNOMED-CT concepts as a label of activities facilitates the interpretation of the resulting care pathways universally across all health organizations.

To sum up, the tool streamlines the creation of standardized care pathways by integrating any event log with ICD codes and SNOMED CTs using a graph database. It also automates the
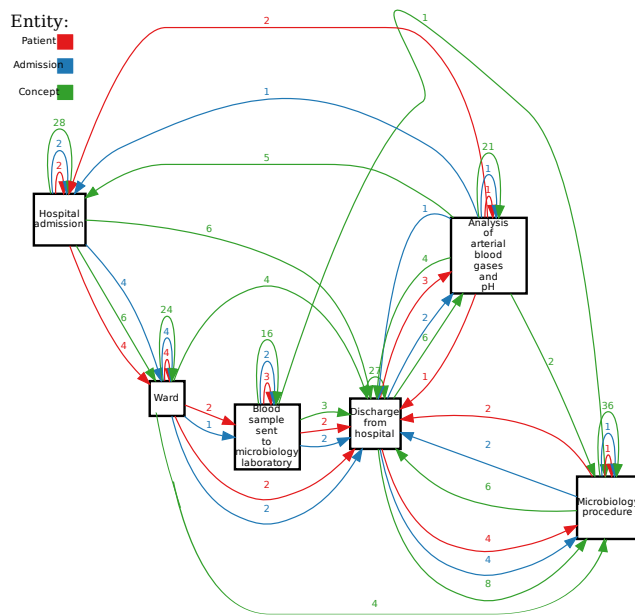
Figure 3: The resulting event graph identifies the most frequently repeated activities in treating two patients, all mapped to SNOMED-CT for standardized analysis.

generation and execution of the necessary queries for building the graph database, ensuring a seamless process. One area of future research could focus on identifying additional care pathways from the clinical event knowledge graph.

References

[1] A. Marengoni, S. Angleman, R. Melis, F. Mangialasche, A. Karp, A. Garmen, B. Meinow, L. Fratiglioni, Aging with multimorbidity: a systematic review of the literature, Ageing research reviews 10 (2011) 430–439.

[2] J. Munoz-Gama, N. Martin, C. Fernandez-Llatas, O. A. Johnson, M. Sepúlveda, E. Helm, V. Galvez-Yanjari, E. Rojas, A. Martinez-Millana, D. Aloini, et al., Process mining for healthcare: Characteristics and challenges, Journal of Biomedical Informatics 127 (2022) 103994.

[3] M. Naeimaei Aali, F. Mannhardt, P. Jelle Toussaint, Clinical event knowledge graphs: Enriching healthcare event data with entities and clinical concepts-research paper, in: International Conference on Process Mining, Springer, 2023, pp. 296–308.

[4] A. Swevels, E. L. Klijn, D. Fahland, Object-centric process mining (and more) using a graph-based approach with promg., in: ICPM Doctoral Consortium/Demo, 2023.

[5] A. E. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow, et al., Mimic-iv, a freely accessible electronic health record dataset, Scientific data 10 (2023) 1.