CHAPTER 1

INTRODUCTION

This chapter includes the following subtopics, namely: (1) Rationale; (2) Theoretical Framework; (3) Conceptual Framework/Paradigm; (4) Statement of the problem; (5) Hypothesis (Optional); (6) Assumption (Optional); (7) Scope and Delimitation; and (8) Importance of the study.

1.1 Rationale

Electronic Health Records (EHR) are large repositories of clinical and operational data that have a variety of use cases from population health, clinical decision support, risk factor stratification and clinical research[19]. However, health record systems store large portions of clinical information in unstructured format or proprietary structured formats, resulting in data that is hard to summary, extract and analyse. There is a need for a platform to accurately extract information from freeform health text in a scalable manner that is generalized so that it is interoperable among various systems.[22]

On the other hand, When taking two or more drugs at the same time or in succession, the activity of one drug may be alerted significantly due to the presence of other drugs, which is described as DDI. DDI databases offer crucial information about the potential interactions that may be happened between drugs that are given to treat various diseases, which is vital for the patient safety and in the making of effective treatment plans.[22] However, these two important databases are frequently not integrated from one another, leading to gaps in information that could enhance patient care outcomes. A domain-specific knowledge graph that integrates EHR and drug interaction data offers a novel approach to bridge these gaps, enabling healthcare professionals with a comprehensive perspective on patient information and enhancing the decision-making processes to give the patient the best patient care that available.

A graph database is a type of database that utilizes graph structures for semantic queries, with nodes, edges, and properties to represent and store data. The key concept behind a graph database is the graph, which directly associates entities (nodes) and the relationships (edges) that connect them. Each node represents an entity (such as a person, place, or thing), and each edge represents a connection or relationship between two nodes. This structure allows for a highly flexible and intuitive representation of complex networks of relationships, like social networks, organizational structures, or communication networks. Graph databases offer a significant performance boost in these scenarios, providing a compelling argument for their adoption over traditional models or even within the context of knowledge graphs.

The emerging scope of knowledge graphs can contribute to the healthcare industry and academia by assisting in clinical decision-making, accelerating drug discovery, detecting adverse drug effects, recommending medicine, and predicting differential diagnosis.[10]

Pharmaceutical care is “the responsible provision of drug therapy for the purpose of achieving definite outcomes that improve a patient’s quality of life” and it “involves the process through which a pharmacist co-operates with a patient and other professional (Healthcare systems) in designing, implementing and monitoring a therapeutic plan that will produce specific therapeutic outcomes for the patient”. Nowadays, Healthcare systems heavily rely on the knowledge and the experiences of the physicians for drug prescription based on diagnosed symptoms of patients.[5]

Conventional Process Challenges:

- Several different types of drugs may be appropriate to treat the same disease.
- Many healthcare professionals who are not physicians are not supposed to prescribe drugs in normal situations.[14]
- Physicians may want to search for all potential existing drugs which may have positive effects given the observed symptoms of the disease. [20]

While a set of very valuable sources is publicly available, no comprehensive database exist that can accommodate the listed challenges. Unfortunately, as no comprehensive knowledge graph (with both medical and non-medical) knowledge exists today, much manual effort is required to consolidate knowledge across sources, that are heterogeneous in content and formats.[14]

A real-world knowledge source such as HER data is crucial to provide better diagnostic support. As EHR contains heterogeneous information on the health status of patients, such knowledge graphs can better furnish the evidence required in a clinical practice. Learning medical knowledge graphs from EHR data incorporates real-world data with medical knowledge and makes the process effortless compared to scientific article-based medical knowledge graphs.[10]

1.2 Theoretical Framework

Graph theory is a mathematical framework that studies the properties and applications of graphs, which are structures made up of nodes (or vertices) connected by edges. Graph theory is instrumental in modeling the relationships between various entities in the knowledge graph that is included in the Electrical Health Records, such as patients, diseases, medications, drug interactions doctor’s name, apothecary’s name, etc. By representing these entities and their relationships as a graph, graph theory enables efficient querying, visualization, and analysis of complex data structures.

Knowledge fusion helps to combine all data of a specific entity from multiple sections and unify all the collected information into the base entity. Knowledge fusion is highly required when there are different data sources for knowledge graphs. The knowledge support provided by multiple source knowledge fusion is highly needed in intelligent data processing and extracting hidden knowledge patterns from knowledge graphs. Mapping of differently expressed entities to a single original entity is a significant task known as entity resolution. Knowledge integration from web sources and medical literature helps enrich the graph with recent advances and helps to enhance the information in EHR. Integrating clinical concepts from multiple institutes is useful for establishing cooperation and shareability of clinical knowledge.

Health informatics is an interdisciplinary field that focuses on the acquisition, storage, retrieval, and use of healthcare information to improve patient outcomes and enhance healthcare delivery. This research draws on health informatics principles to understand the needs and requirements of healthcare providers when integrating EHR and drug interaction data. Health informatics emphasizes the importance of data quality, patient privacy, and security, all of which are critical considerations when constructing a domain-specific knowledge graph. By aligning with health informatics standards and best practices, the research aims to ensure that the knowledge graph is robust, reliable, and compliant with healthcare regulations.

1.3 Conceptual Framework/Paradigm

The conceptual framework for this study outlines the relationships between the key variables involved in constructing and evaluating a domain-specific knowledge graph that integrates Electronic Health Records (EHR) and drug interaction data. The framework illustrates how these variables interact to address the research problem and achieve the study's objectives.

The Input Variables is Electronic Health Records (EHR) Data. This includes structured data from EHR systems, such as patient demographics, medical history, diagnoses, medications, and laboratory results. EHR data provides the foundational information necessary for building the knowledge graph. Drug Interaction Data, his involves data from drug interaction databases that describe the potential adverse interactions between different medications. This data is crucial for understanding how medications interact with each other and with patient conditions.

After collecting the data, we process the variables, these includes Data Standardization and Preprocessing The process of cleaning, transforming, and standardizing the EHR and drug interaction data to ensure compatibility and integration within the knowledge graph. Then, Knowledge Graph Construction, the development of a graph-based model that represents the relationships between entities in the EHR and drug interaction data. This involves selecting appropriate graph-based algorithms and defining the nodes and edges that represent different entities and their relationships.

The Output Variables will be Integrated Knowledge Graph, A domain-specific knowledge graph that integrates EHR and drug interaction data, providing a comprehensive and interconnected representation of clinical information. Clinical Decision Support Enhancement: The ability of the integrated knowledge graph to enhance clinical decision-making processes, particularly in

1.4 Statement of the Problem

The integration of Electronic Health Records (EHR) and drug-drug interaction data is crucial for enhancing patient safety and improving clinical decision-making. Despite advancements in healthcare data management, there remains a significant gap in how these datasets are combined and utilized to support clinical practices effectively. This study aims to address this gap by constructing and evaluating a domain-specific knowledge graph that integrates EHR and drug interaction data. In this research we will try to answer how can a domain-specific knowledge graph be constructed and evaluated to effectively integrate Electronic Health Records (EHR) and drug interaction data.

Beside that, we also try to answer several questions regarding the integration of Electronic Health Records (EHR) and drug interaction data such as: What are the existing data sources for EHR and drug interactions, and how can these datasets be standardized and prepared for integration into a domain-specific knowledge graph?, What methods and technologies can be employed to construct a knowledge graph that accurately represents the relationships between entities within EHR and drug interaction data? And also, what criteria and metrics should be used to evaluate the effectiveness of the constructed knowledge graph in terms of its accuracy, completeness, and usability in clinical settings?

Each problem is designed to address a specific aspect of the research, from the initial stages of data collection and preparation to the construction, integration, evaluation, and practical application of the knowledge graph. By systematically addressing these the research aims to provide a comprehensive solution to the general problem of integrating EHR and drug interaction data for improved clinical outcomes.

1.5 Objective and Hypotheses

The objective of this research is to construct a Knowledge Graph (KG) based on EHR (Electronic Health Records) and fuse it with Drug Interaction knowledge graph that will be built on the knowledge of disease, prescription, and drug-drug interactions and then validate the fused knowledge graph. This fused Knowledge Graph will be evaluated using coverage, correctness, and clinical correctness.

Hypothesis :

1. The EHR (Electronic Health Records) knowledge graph will be build based on Unstructured Data and Structured Data.

2. The Validation of the fused knowledge graph between EHR (Electronic Health Records) and fuse it with Drug Interaction knowledge graph will give a high percentage.

3. The fused KG will demonstrate high percentage clinical correctness when evaluated by a panel of healthcare professionals, based on its ability to support clinical decision-making.

The study proposes building a Knowledge Graph by integrating EHR that based on Unstructured and structured data and Drug Interaction data using Neo4j. Then aim to validate it based on comprehensive metrics. The hypotheses anticipate successful construction of the EHR KG, high validation rates for the fused KG, and strong coverage, completeness, and clinical relevance, affirming its utility in healthcare decision support.

1.6 Assumption

It is assumed that sufficient and high-quality Electronic Health Record (EHR) and drug interaction data are available for integration into the knowledge graph. This includes assuming that the data sources are reliable, up-to-date, and comprehensive enough to build a meaningful knowledge graph. It is assumed that the integrated knowledge graph will provide clinically relevant information that is useful for enhancing decision-making and improving patient safety. This includes the assumption that healthcare professionals will find the knowledge graph valuable for identifying potential drug interactions and making informed decisions. It is assumed that healthcare professionals are willing to adopt new technologies that improve clinical workflows and patient outcomes, provided that these technologies are user-friendly and demonstrably effective.

1.7 Scope and Delimitation

Scope

1. Principal Variables

   Knowledge Graph Integration: The primary focus of this study is the integration of Electronic Health Records (EHR) and drug interaction data into a domain-specific knowledge graph. The key variables include the structure of the knowledge graph, the data integration techniques used, and the resulting relationships between entities (e.g., patients, medications, and drug interactions).

2. Location

The study will be conducted within a select number of healthcare institutions that use EHR systems and have access to drug interaction databases. These institutions will be chosen based on their willingness to participate in the study and their ability to provide the necessary data for integration.

3. Timeframe

   The research will be carried out over a period of 12 months. This timeframe includes stages such as data collection and preparation, knowledge graph construction, integration of data, and evaluation of the knowledge graph

4. Justification

   The scope of this study is justified by the need to enhance clinical decision-making through better data integration. The focus on EHR and drug interaction data is particularly relevant given the critical role of medication management in patient safety. Additionally, limiting the study to a specific set of healthcare institutions allows for a controlled environment in which the effectiveness of the knowledge graph can be thoroughly evaluated.

Delimitation

1. Data Sources

   The study will be limited to data obtained from specific EHR systems that range between 15th August 2022 until 18th August 2023. It does not account for other potential sources of clinical data, such as imaging data, genomic data, or data from wearable devices.

   The study will focus primarily on Antidiabetic Drug Interactions, because the data is limited to this specific disease. It does not account for other drug interactions database but can be added in the future when it's available.

2. Technology Scope

   The research is restricted to the use of specific technologies for knowledge graph construction and integration, such as ontologies, and graph-based algorithms. It does not explore alternative data integration methods or emerging technologies beyond those specified.

3. Evaluation Metrics

   The evaluation of the knowledge graph will be based on predefined metrics, including accuracy, coverage, and usability. Other potential benefits, such as cost savings, workflow efficiency, or long-term patient outcomes, are outside the scope of this study.

These delimitations ensure a focused and manageable research scope, allowing for a detailed exploration of the integration of EHR and drug interaction data within a specific context while acknowledging the limitations and boundaries of the study.

1.8 Significance of the Study

This study contributes to the field of healthcare informatics by exploring novel methods for integrating heterogeneous datasets, specifically EHR and drug interaction data, into a cohesive and comprehensive knowledge graph. By developing a knowledge graph tailored to the healthcare sector, this study provides a deeper understanding of how knowledge graphs can be utilized to represent complex relationships among clinical entities, such as patients, medications, and potential drug interactions. The findings of this study have the potential to make clinical decision-making more conclusive by providing healthcare professionals with a more comprehensive view of patient data and potential drug interactions. This can lead to more informed decisions, reduced medication errors, and ultimately, enhanced patient safety.

In summary, this research will answer few important questions regarding the construction and evaluation of a domain-specific Knowledge Graph

- KG construction. How to construct a domain-specific and high-quality KG? How to refine the KG to ensure the data quality?
- KG evaluation. How to evaluate the quality of a KG quantitatively and rigorously? What are the most important quality dimensions and their relations? How can those dimensions be measured? We reuse two quality dimensions from existing literature then conduct a comprehensive evaluation of the Drug KG we created

The main contributions of this research are threefold:

- The construction of Drug Knowledge Graph with structuring the unstructured data that we have from EHR records such as medical notes. The KG offers an urgent need of data sources for building high-quality knowledge discovery systems Based on EHR
- Integrating between two data sources Electonical health records and Anti-diabetic Drug-drug interactions.
- Developing a platform for analyzing the EHR. The proposal of a feasible quality evaluation method of KG with an empirical validation of its effectiveness

Hopefully with this integration help computer science professionals in performing operations, and handling tasks related to medical aspects in Patient Medical Record, such as determining the most suitable drugs, identifying potentially harmful substances for patients, and other matters related to medical data.