DOI: 10.20472/EFC.2022.016.009

SUELA MAXHELAKU
University of Tirana, Albania

ELDA XHUMARI
University of Tirana, Albania

ENDRIT XHINA
University of Tirana, Albania

EXPLOITING KNOWLEDGE GRAPHS IN HEALTHCARE

Abstract:

The use of knowledge graphs in health care has increased the performance in which information is retrieved and patients' diagnoses are made. Because the healthcare domain is made up of complex concepts and many relationships between them, developing knowledge graphs presents numerous challenges. Another issue is the large number of medical standards, such as HL7, ICD-10, SNOMED CT, DICOM, LOINC, and so on. In this paper are analyzed the technologies used in the construction of knowledge graphs in healthcare. In addition, a model for generating knowledge graphs from the APHRO (Albanian Patient Healthcare Records) Ontology is proposed.

Keywords:

Knowledge graphs, Ontology, Healthcare

JEL Classification: C45, I19

1 Introduction

The exponential growth of clinical data, that is mostly heterogenous and unstructured, requires semantic, statistical, and predictive analysis to convert this multidimensional data into structured and machine-understandable knowledge, known as Knowledge Graph. (Khalid, et al., 2020) The Notion of a Knowledge Graph originates from scientific advances in several research areas, including the Semantic Web, databases, knowledge representation and reasoning, NLP as well as machine learning. Knowledge Graphs can be considered as the coming of age of the integration of data and knowledge at large scale with heterogeneous formats. The integration of ideas and techniques from such disparate disciplines poses a challenge to researchers and practitioners to understand how current advances develop from and are rooted in in early techniques. (Gutierrez & Sequeda, Knowledge Graphs, 2021) A knowledge graph is a graph of data meant to accumulate and communicate knowledge of the real world. Nodes represent entities of interest, and edges represent potentially different relationships between these entities. (Hogan, et al., 2021) Knowledge Graph use data and logic to structure information. (Li, Horiguchi, & Sawaragi, 2022)

The typical trends using knowledge graph consist of: Use knowledge graphs to integrate large scale data from numerous sources as well as leverage it; and the combination of deductive (rules, ontologies, etc.) and inductive techniques (machine learning, analytics, etc.) to represent as well as accumulate knowledge. (Gutierrez & Sequeda, Knowledge Graphs, 2021).

2 Related work

Knowledge graph promotes the development of healthcare management as it can provide knowledge services to many stakeholders including health consumers, health service providers, and online healthcare platforms. For an intelligent question answering system, the knowledge graph can provide semantic analysis and search functions. (Huang, Yu, Chi, & Xu, 2019)

The Knowledge graph can also be used for misinformation detection in healthcare. DETERRENT is a knowledge guided graph that leverages additional information from a medical knowledge graph, to guide the article embedding with a graph attention network. The network can capture both positive and negative relations, and automatically assign more weights to important relations in differentiating misinformation from fact. The node embedding is used for guiding text encoder. Experiments on two real-world datasets demonstrate the strong performance of DETERRENT. (Cui, et al., 2020)

Disease ontologies for knowledge graphs is a knowledge base solution that uses Grakn core with its logical inference and disease ontologies cross-references to allow easy switching between ontology hierarchies for data integration purpose. Grakn core with pre-installed *"Disease ontologies for knowledge graphs"* facilitates the biomedical knowledge graph build and provides an elegant solution for the multiple disease ontologies problem. This software makes it straightforward to run common ontological queries. It is relatively easy to add new ontologies due

to the python loading scripts, and the Grakn reasoning rules are easy to extend. This solution tends to integrate data that uses multiple ontologies. (Kurbatova & Swiers, 2021)

3 Technologies used in construction of knowledge graph in healthcare

There are different technologies used in the construction of the knowledge graph from general domain to specific domain.

Health Knowledge Graph Builder (HKGB) is an end-to-end platform which could be used to construct the disease-specific and extensible health knowledge graphs from multiple sources. The authors adopted ML algorithms to semi -auto-constructed health knowledge graph framework builder. The HKGB has four components as shown in the figure above: (Zhang, et al., 2020)

- Clinician in the loop
- Construction of CKG
- Construction of IKG
- DummyTXdummy (Graph based tools)

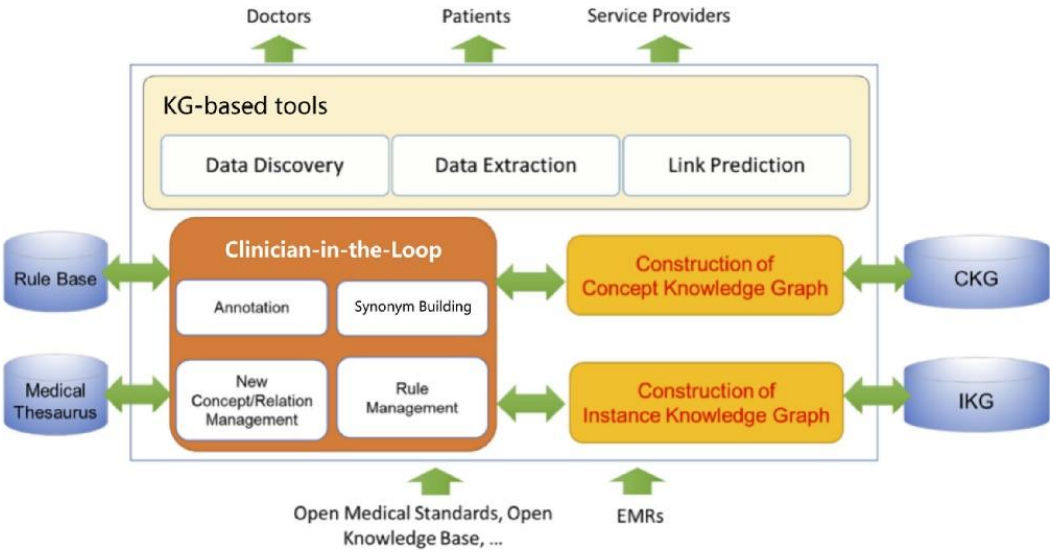

Figure 1 The framework of HKGB

Sources (Zhang, et al., 2020)

cTakes (clinical Text Analysis and Knowledge Extraction System) is a modular system of pipelined components combining rule-based and machine learning techniques aiming at information extraction from the clinical narrative. The cTAKES accepts either plain text or clinical document architecture-compliant XML documents. The open-source release in http://www.ohnlp.org consists of the following components/annotators: Sentence boundary detector, Tokenizer, Normalizer, Part-of-speech (POS) tagger, Shallow parser, Named entity recognition (NER) annotator, including status and negation annotators. The components in cTAKES are specifically trained for the clinical domain and create a rich linguistic and semantic annotations. (Savova, et al., 2010)

Ali, et al. proposed a knowledge authoring environment called Intelligent-Knowledge Authoring Tool (I-KAT), which overcomes the limitations of the state-of-the-art systems. They also introduced a semantic reconciliation model (SRM) to normalize the knowledge acquisition complexity that adopts HL7 MLM in combination with vMR and SNOMED CT. The incorporate SNOMED CT codes into the MLMs to avoid the intrinsic vocabulary binding issue of Arden Syntax. Using SNOMED CT and vMR in MLM creation enhances shareability and interoperability on one hand but increase system complexity with respect to knowledge creativity. (Ali, et al., 2017)

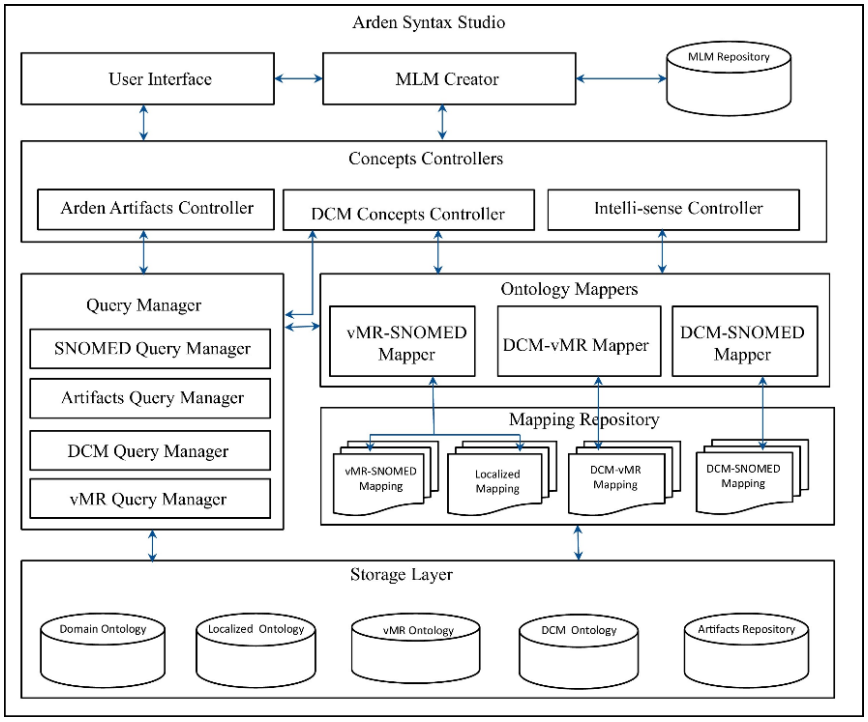

Figure 2 Rule creation and mapping generation and use in rule creation workflow.

Author (Ali, et al., 2017)

Maghawry, Emara, Shaaban, Ghoniemy developed a framework for automatic disease knowledge graph (KG) construction and intelligent ontology integration with standard human disease ontology (DO). They proposed a system in order to reach an integrated knowledge graph that weill be the base for the intelligent healthcare systems. It is composed of three stages, the first stage is designing and implementing a Mayo-clinic scraper that is responsible for scraping the available diseases in the encyclopedia, their causes, risk factors and prevention factors, and building the suitable knowledge graph. The second stage is extracting concepts from the DO, where the third stage is integrating the generated knowledge graph from stage one with the standard ontology concepts. (Maghawry, Emara, Shaaban, & Ghoniemy, 2022)

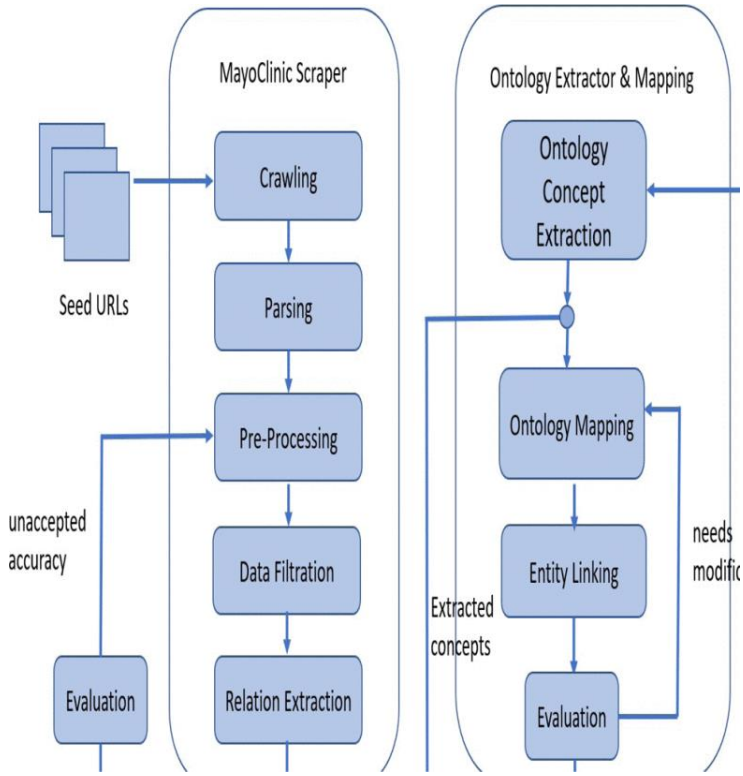

Figure 3 Proposed system

(Maghawry, Emara, Shaaban, & Ghoniemy, 2022)

The Knowledge Collaboratory is a web service to query and publish Nanopublications for the NCATS Biomedical Data Translator project. It can Browse Nanopublications through the web UI, or query them using the Translator Reasoner API (TRAPI) specifications or Publish Nanopublications after login with ORCID, and uploading your authentication keys to the server. Developed and hosted by the Institute of Data Science[1] at Maastricht University. The Knowledge Collaboratory consists in an OpenAPI service and a user-friendly web UI to query the Nanopublications network, store Nanopublication authentication keys, and publish Nanopublications.[2] It consists of:

- **Backend API**: an OpenAPI built with Python and FastAPI, to store the keys on the server, and run the process to publish a Nanopublication.
- **Webapp frontend**: a website built with TypeScript and React, to provide a user-friendly access to the Nanopublication network.

Soest, Choudhury, Gaikwad, Sloep, Dumontier and Dekker provided a framework and tooling where users can easily make their data (more) FAIR (findable, accessible, interoperable, reusable), while reducing the amount of information needed to get started. Specifically, to annotate existing datasets using RDFS and OWL statements. The tool to extract the ontology,

[1] https://www.maastrichtuniversity.nl/research/institute-data-science

[2] https://collaboratory.semanticscience.org/

and to materialize the triples is publicly available at https://gitlab.com/UM-CDS/fair/tools/triplifier. This tool is available as a stand-alone java application or can be executed as a service in a docker container. (Soest, et al., 2019)

4 From APHRO Ontology to knowledge graphs

In (Maxhelaku, Shehu, & Xhina, APHRO Ontology for managing patient data, 2021) (Maxhelaku, Integrating medical data in designing e-health architecture, 2021) we have developed the APHRO (Albanian Patient Healthcare Records Ontology) Ontology. This ontology is designed to manage patient health records. APHRO Ontology is an ontology that is developed to different records of the patients like demographics data, risk factors, vital signs, patient visits, ect.

Figure 4 Vital Signs of the Patient

Source: Author (Maxhelaku, Shehu, & Xhina, APHRO Ontology for managing patient data, 2021)

The APHRO Ontology can be used to add information about hospitalized patients, such as specific symptoms, sensory findings, diagnoses, head findings, skin findings, and so on. The knowledge graph can be created using real patient records based on diagnosis and vital sign results. So, for each of our ontological relationships, we can create unique instances. The individual instance of the relationship Patient ⟶ *has_skin_finding* ⟶ Skin Finding in the APHRO Ontology, should be as follows:

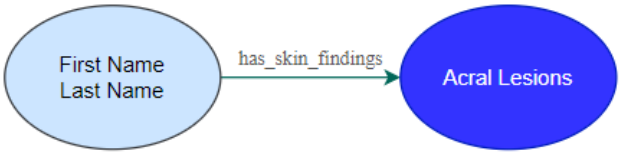

Figure 5 First individual instance of knowledge graph

Source: Author

We can view our data as a web of relationships with new connections between data points using the information from this knowledge graph. We can also get more information about the various data that make up the knowledge graph by using SPARQL.

5 Conclusions and future work

The development of healthcare management is facilitated by knowledge graphs, which can provide knowledge services to a variety of stakeholders, including health consumers, health service providers, and online healthcare platforms.This paper analysises the technologies used in the construction of knowledge graphs in healthcare.

A model for generating knowledge graphs from the APHRO (Albanian Patient Healthcare Records Ontology) Ontology is also proposed. In the future, we intend to add patient information to the APHRO ontology in order to complete the knowledge graph presentation.

Bibliography

Ali, T., Hussain, M., Wajahat, A. K., Muhammad, Afzal, Jamil, H., . . . Lee, S. (2017). Multi-model-based interactive authoring environment for creating shareable medical knowledge. *Computer Methods and Programs in Biomedicine, 150*, 41-72.

Cui, L., Seo, H., Tabar, M., Ma, F., Wang, S., & Lee, D. (2020). Knowledge Guided Graph Attention Network for Detecting Healthcare Misinformation. *Proceedings of the 26th ACM SIGKDD International Conference on knowledge discovery & data mining. ACM*, 492-502.

Gutierrez, C., & Sequeda, J. F. (2021). Knowledge Graphs. *Communications of the ACM, 64*(3).

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., . . . Zimmermann, A. (2021). Knowledge graphs. *Communications of the ACM, 64*, 96-104.

Huang, L., Yu, C., Chi, Y. Q., & Xu, H. (2019). Towards Smart Healthcare Management Based on Knowledge Graph Technology. *In Proceedings of the 2019 8th International Conference on Software and*

Computer Applications (ICSCA '19). Association for Computing Machinery, New York, NY, USA,, 330-337.

Khalid, M. M., Madan, K., Mazen, A., Maqbool, H., Fakhare, A., & Ghaus, M. (2020). *Automated domain-specific healthcare knowledge graph curation framework: Subarachnoid hemorrhage as phenotype,* (Vol. 145). Expert Systems with Applications.

Kurbatova, N., & Swiers, R. (2021). Disease ontologies for knowledge graphs. *BMC Bioinformatics, 22*(377).

Li, J., Horiguchi, Y., & Sawaragi, T. (2022). Counterfactual inference to predict causal knowledge graph for relational transfer learning by assimilating expert knowledge --Relational feature transfer learning algorithm. *Advanced Engineering Informatics, 51*.

Maghawry, N., Emara, K., Shaaban, E., & Ghoniemy, S. (2022). *Automated intelligent online healthcare ontology Integration.* 2022 5th International Conference on Computing and Informatics (ICCI).

Maxhelaku, S. (2021). *Integrating medical data in designing e-health architecture.* Albania: PhD Thesis, University of Tirana.

Maxhelaku, S., Shehu, J., & Xhina, E. (2021). APHRO Ontology for managing patient data. *RTA-CSIT 2021.*

Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & Chute, C. G. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association, 17*(5), 507-513.

Soest, J. V., Choudhury, A., Gaikwad, N., Sloep, M., Dumontier, M., & Dekker, A. (2019). *Annotation of existing databases using Semantic Web.* SWAT4HCLS.

Zhang, Y., Sheng, M., Zhou, R., Wang, Y., Han, G., Zhang, H., . . . Dong, J. (2020). HKGB: An Inclusive, Extensible, Intelligent, Semi-auto-constructed Knowledge Graph Framework for Healthcare with Clinicians' Expertise Incorporated. *Information Processing & Management, 57*(6).