

Knowledge Graph Solutions in Healthcare for Improved Clinical Outcomes

Jans Aasman, Ph.D.¹ and Parsa Mirhaji, MD, Ph.D.²

¹ Franz Inc, 2201 Broadway, Suite 715, Oakland, California 94612

jans.aasman@franz.com

² Montefiore Medical Center, Institute for Clinical and Translational Research,
6 Executive Plaza, Suite 112, Yonkers, New York, 10701

pmirhaji@montefiore.org

Abstract. Deploying patient Knowledge Graphs based on Semantic Technologies offers improved patient care and revolutionizes care models and medical research. Knowledge Graphs provide the ability to search the information that's available in a much more efficient way in order to find patterns in the data and to use those patterns for clinical purposes to improve clinical outcomes.

Keywords: Knowledge Graph, Graph Database, Machine Learning, RDF, Semantic Web

1. Healthcare Challenges

Located in the Bronx, Montefiore Health System serves one of the most ethnically and socioeconomically diverse populations in the US. Like all healthcare organizations, Montefiore faces complex challenges—from government pressures to reduce costs and stringent regulatory guidelines to diverse patient populations and disruptive technologies. Understanding patients requires information on a complex array of factors, some of which may not even be known during a clinical interaction, such as the home and work environment, nutrition, and genetics.

The industry has long collected data on patients—it is not uncommon for hospitals to gather thousands of data points per patient per day. Data ranges from unstructured free text information to images and waveforms to data from sensors and monitoring devices. Access to accurate data is vital for assessing risk from intubation to drug interactions. But often this data cannot be analyzed quickly, nor can hospital data be easily combined with external data sources such as those from pharmaceutical companies and researchers.

2. Knowledge Graph Solutions

To optimize healthcare based on advanced data analytics and make sure clinicians have the right information available in time to impact patient outcomes, Montefiore in partnership with Franz Inc and Intel Corp. have deployed and deployed PALM - Patient-centered Analytic Learning Machine, a solution that brings together varied and vast amounts of raw data for deeper analysis to flag patients who are at risk or help clinicians identify optimal treatment plans. The PALM Knowledge Graph platform integrates both structured and unstructured data ranging from basic science, clinician records, and population demographics to community, environmental, behavioral, and wellness research data. By assessing a holistic and realistic profile of patients—along with relevant science, clinical population histories, drug information, and medical imaging—PALM has the capability to improve care, identify at-risk patients, and personalize medicine, while reducing error and inefficiency.

3. PALM Technical details

The PALM Knowledge Graph is a 'semantic solution' from the ground up. At the core of the Knowledge Graph are more than 180 different life science and health care taxonomies and ontologies that are interlinked (see Fig.1). Every conceivable type of patient information coming from more static databases or from dynamic HL7 streams are

mapped onto an events based ontology that makes it several orders of magnitude easier to do ad-hoc queries, analytics and feature extraction for machine learning. A frame description language (FDL) makes it straight forward for data scientists to declaratively specify the features needed for analytics and machine learning. The FDL is also used to facilitate the learning aspect of the Knowledge Graph, that is: it specifies how to systematically store the output of analytics back into the Knowledge Graph. We expect the size of the Knowledge Graph to grow to more than 2 trillion triples in the near future so scalability is a very important consideration. We have implemented a uniquely scalable approach to storage of the entire knowledge graph in a distributed graph database. The architecture of this distributed graph database is based on a combination of partitioning and federation. Patient data is partitioned and the partitions are federated with local unpartitionable knowledge bases and terminology systems. A new parallel distributed SPARQL was developed to leverage all hardware resources in executing queries and performing analytics. For queries that need to combine data from various partitions we developed a unique SPARQL pipeline mechanism. Part of the current efforts is deep integration with SPARK to facilitate this pipeline mechanism but also to facilitate machine learning and analytics.

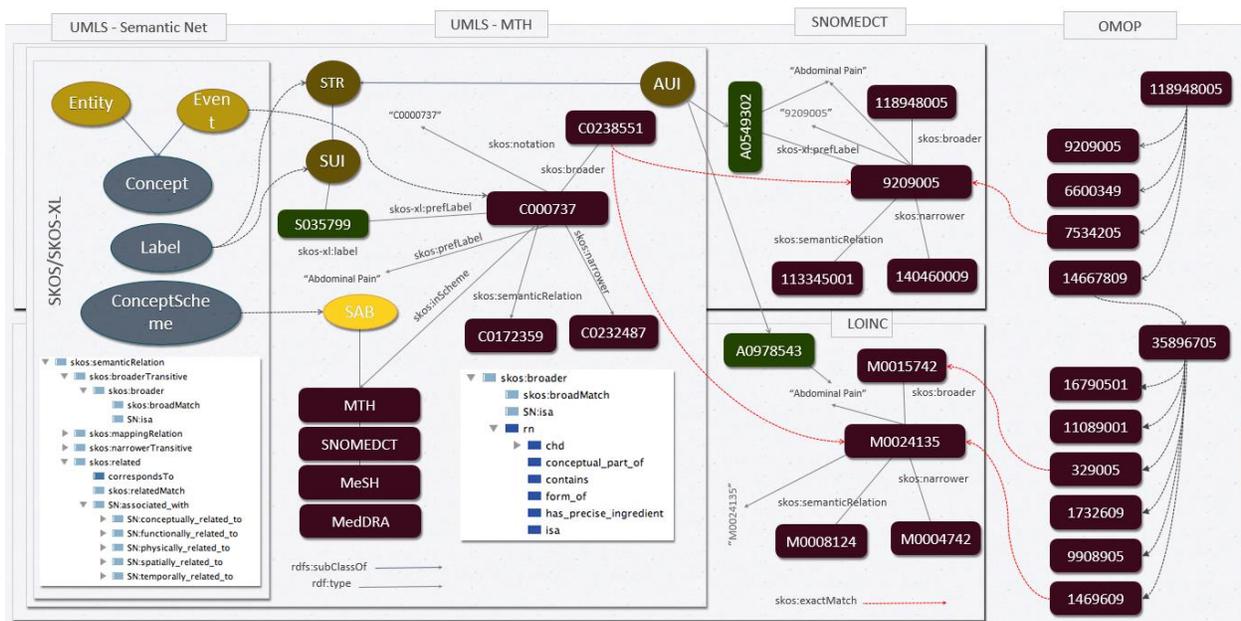


Fig. 1 - The PALM Knowledgebase

4. Summary

The use of Semantic Web technologies to generate automated predictive and preventive approaches has proven effective in identifying patients at highest risk, and providing consistent clinical decision support to relevant practitioners. In one use case, accurate prediction of prolonged ventilation detects patients with more than 70% likelihood of an event, 48 hours in advance of a fatal episode or respiratory failure in the hospital in order to avoid the crisis.

PALM provides the ability to search the information that's available to clinicians in a much more efficient way in order to find patterns in the data and to use those patterns for to improve patient outcomes.