Special issue on Semantic Web Meets Health Data Management

Editorial

Kostas Stefanidis [a,*], Haridimos Kondylakis [b] and Praveen Rao [c]

[a] *Tampere University, Finland*
E-mail: konstantinos.stefanidis@tuni.fi
[b] *FORTH-ICS, Heraklion, Crete, Greece*
E-mail: kondylak@ics.forth.gr
[c] *University of Missouri, USA*
E-mail: praveen.rao@missouri.edu

1. Introduction

The COVID-19 pandemic has catapulted healthcare research as a top priority for many nations. Researchers have used data-driven approaches to better understand COVID-19, develop effective vaccines, and mitigate the spread of the virus. Healthcare data management continues to evolve as mankind faces the biggest public health crisis of modern times. As new kinds of data emerge, new models, algorithms, and techniques are needed to better harness the value of healthcare data for advanced decision making.

On the other hand, Semantic Web technologies can provide effective solutions for enabling interoperability and common language among healthcare systems, and can lead to the disambiguation of the information through the adoption of various terminologies and ontologies available. In addition, AI and machine learning can enable data-driven decision making and extracting meaningful insights from complex healthcare datasets. Thus, knowledge representation and reasoning on healthcare data become even more important. Semantic Web technologies have matured over the years and can provide these capabilities by design.

The purpose of this Special Issue is to collect contributions on the cross-cutting the fields of Semantic Web, data science, data management, and health informatics to discuss the challenges in healthcare data management and to propose novel and practical solutions for the next generation of data-driven healthcare systems. The ultimate goal is to enable new innovations in Semantic Web, knowledge management, and data management for healthcare systems to move the needle to achieve the vision of precision medicine.

After two rigorous review rounds, 6 papers have been accepted for publication in this Semantic Web – Interoperability, Usability, Applicability special issue.

The first paper, titled "Evaluating the usability of a semantic environmental health data framework: approach and study" and authored by Albert Navarro-Gallinad, Fabrizio Orlandi, Jennifer Scott, Mark Little and Declan O'Sullivan, studies a semantic framework that links health events with environmental data to explore the environmental risk factors of rare diseases. The usability study results indicate that the proposed framework is useful in allowing researchers themselves to link health and environmental data whilst hiding the complexities of semantic web technologies.

---

*Corresponding author. E-mail: konstantinos.stefanidis@tuni.fi.

The second paper, titled "ciTIzen-centric DatA pLatform (TIDAL): Sharing distributed personal data in a privacy-preserving manner for health research" and authored by Chang Sun, Marc Gallofré Ocaña, Johan van Soest and Michel Dumontier, proposes to give individuals ownership of their own data, and connect them with researchers to donate the use of their personal data for research while being in control of the entire data life cycle, including data access, storage and analysis. For doing so, the platform integrates a set of components for requesting subsets of RDF data stored in personal data vaults based on SOcial LInked Data technology and analyzing them in a privacy-preserving manner.

The third paper, titled "Terminology and ontology development for semantic annotation: A use case on sepsis and adverse events" and authored by Melissa Yan, Lise Tuset Gustad, Lise Husby Høvik and Øystein Nytrø, considers that annotations enrich text corpora and provide necessary labels for natural language processing studies, and studies how a terminology and a corresponding ontology are developed. They use a terminology that represents annotated documents and assists annotators in annotating text and an ontology that is intended for clinician use and captures domain knowledge needed to reason and infer implicit information from data, to make the ontology development understandable and accessible to domain experts without formal ontology training.

The fourth paper, titled "Empowering machine learning models with contextual knowledge for enhancing the detection of eating disorders in social media posts" and authored by Jose Alberto Benitez-Andrades, Maria Teresa García-Ordás, Mayra Russo, Ahmad Sakor, Luis Daniel Fernandes Rotger and Maria-Esther Vidal, proposes an approach where knowledge encoded in community-maintained knowledge graphs is combined with deep learning to categorize social media posts using existing classification models. Knowledge graph embeddings are utilized to compute latent representations of the extracted entities, which result in vector representations of the posts that encode these entities' contextual knowledge extracted from the knowledge graphs. The approach is applied to detect whether a publication is related to an eating disorder and uncover concepts within the discourse that could help healthcare providers diagnose this type of mental disorder.

The fifth paper, titled "Context-aware query derivation for IoT data streams with DIVIDE enabling privacy by design" and authored by Mathias De Brouwer, Bram Steenwinckel, Ziye Fang, Marija Stojchevska, Pieter Bonte, Filip De Turck, Sofie Van Hoecke and Femke Ongenae, presents DIVIDE, a component for a semantic IoT platform that adaptively derives and manages the queries of the platform's stream processing components in a context-aware and scalable manner, and that enables privacy by design. The results of an evaluation on a homecare monitoring use case demonstrate how activity detection queries derived with DIVIDE can be evaluated in on average less than 3.7 seconds and can successfully run on low-end IoT devices.

The last paper, titled "Knowledge Graphs for Enhancing Transparency in Health Data Ecosystems" and authored by Fotis Aisopos, Samaneh Jozashoori, Emetis Niazmand, Disha Purohit, Ariam Rivas, Ahmad Sakor, Enrique Iglesias, Dimitrios Vogiatzis, Ernestina Menasalvas, Alejandro Rodriguez Gonzalez, Guillermo Vigueras, Daniel Gomez-Bravo, Maria Torrente, Roberto Lopez, Mariano Provencio Pulla, Athanasios Dalianis, Ana Triantafillou, Georgios Paliouras and Maria-Esther Vidal, presents a data ecosystem, DE4LungCancer, of health data sources for lung cancer. Knowledge extracted from heterogeneous sources, e.g., clinical records, scientific publications, and pharmacologic data, is integrated into knowledge graphs. Ontologies describe the meaning of the combined data, and mapping rules enable the declarative definition of the transformation and integration processes. DE4LungCancer is also assessed in terms of the methods followed for data quality assessment and curation.