Designing and Validating a Knowledge Graph for Person-Owned and Fully Controlled Health Data

Summary

The overall goal of the project is to examine the opportunities and pitfalls involved in designing and using experimental approaches for person-owned and controlled health data as knowledge graphs in primary (e.g. clinical diagnoses and treatment) and secondary (e.g. medical research, public health, policymaking, AI and ML needs) use. The project addresses the following research objectives: design and evaluate a knowledge graph for personal health data, which remains under individual ownership and control, aligned with both primary and secondary use requirements; develop and evaluate a formal specification of this graph using descriptive logic; mathematically substantiate the coherence and correctness of an RDF-based knowledge representation; and integrate the solution with existing standards (e.g., HL7 FHIR, ContSys) , as well as with the semantic-interoperability and other tools developed in eMedLab, enhancing RDB2RDF and RDF2RDB transformations where necessary.

Research field:	Information and communication technology
Supervisor:	Gunnar Piho
Availability:	This position is available.
Offered by:	School of Information Technologies Department of Software Science
Application deadline:	Applications are accepted between June 01, 2025 00:00 and June 30, 2025 23:59 (Europe/Zurich)

Description

The research

The prerequisite for the continuity of healthcare and medical services is the semantic interoperability of information systems. This requires machine-readable and unambiguously defined data structures, so that systems can exchange data seamlessly, preserving a consistent understanding of semantics and avoiding loss of data or meaning. Effective data exchange is still hindered by the fact that information systems interpret data structures differently. Although international standards (e.g. HL7 FHIR, ContSys, ISO 13606, SNOMED CT) provide the basis for data sharing, no one has been able to fully resolve the challenges of semantic persistence, machine-processability and decentralization.

A knowledge graph enables data from diverse sources to be linked in a simple triple format, preserving their relationships and meaning and providing semantic access to information precisely when required. It is represented using the Resource Description Framework (RDF), which allows data to be expressed in a machine-readable and linkable form and is particularly well suited to structuring and analyzing large, complex datasets, such as health data.

Responsibilities and (foreseen) tasks

- Design a knowledge graph for personally owned and fully controlled health data, based on the requirements for primary and secondary use.
- Develop a formal description of this knowledge structure using descriptive logic.
- Provide a mathematical justification of the coherence and correctness of the RDF-based knowledge representation.
- Aligning this solution with existing standards such as FHIR and ContSys, as well as with the semantic-interoperability and other tools developed in eMedLab, enhancing RDB2RDF and RDF2RDB transformations where necessary.
- Collaborate with clinicians and healthcare IT experts to gather qualitative feedback and assess the solution’s applicability in clinical workflows
- Participate in research and development seminars
- Publish scientific articles in international journals and present at conferences on NLP and medical AI topics.

Applicants should fulfil the following requirements:

- Master’s degree in information and communications technology

Glowbase Graduate Recruitment Platform - http://www.glowbase.com - © Glowbase GmbH - 2025-06-27 01:09:12

- A clear interest in the position's topic
- Excellent command of English
- Strong, demonstrable writing and analytical skills
- Ability to work both independently and as part of an international team
- Willingness and capacity to assist with organisational tasks relevant to the project

The following experience is beneficial:

- Programming in C# and .NET
- Working knowledge of SQL
- Working knowledge of software engineering
- Working knowledge of healthcare standards
- Working knowledge of Knowledge Graphs and RDF formats

The candidate should submit a research plan on the chosen topic, detailing the overall studies, research and publication strategy. They may expand on the listed research questions and tasks and propose the theoretical foundations to be employed.

We offer:

- A fully funded, 4-year PhD position at eMedLab (https://taltech.ee/en/emedlab), TalTech's interdisciplinary centre for digital health innovation.
- Access to state-of-the-art research infrastructure and collaborative opportunities within the European Federation of Medical Informatics (https://efmi.org/).
- Opportunities for conference travel, international research stays, and networking with leading universities.

About eMedLab

eMedLab is a research group in digital health and medical informatics at Tallinn University of Technology (TalTech). It brings together researchers from the Centre for Digital Health and Business Information Technology research groups of the School of Information Technologies. The eMedLab research group, led by Professor Peeter Ross and Dr Gunnar Piho, consists of around twenty researchers exploring the opportunities and challenges of using human-owned and -controlled health and medical data and developing new methods and technologies to address these challenges.

Additional information

For further information, please contact Dr Gunnar Piho gunnar.piho@taltech.ee or visit https://taltech.ee/en/emedlab

To get more information or to apply online, visit https://taltech.glowbase.com/positions/1036 or scan the the code on the left with your smartphone.

Glowbase Graduate Recruitment Platform - http://www.glowbase.com - © Glowbase GmbH - 2025-06-27 01:09:12