A Semantic Web-Based Framework for Quality Assurance of Electronic Medical Records Data for Secondary Use

Guoqian Jiang, Harold Solbrig, Christopher Chute
Mayo Clinic, Rochester, MN, 55906

Standardized clinical information models have been regarded as the basis for retaining computable meaning when data are exchanged between heterogeneous medical computer systems. Notably, the Clinical Information Modeling Initiative (CIMI), an international collaboration effort, has been actively working on developing a common format for detailed specifications for the representation of health information content. In addition, there is an emerging interest in the use of clinical information models in the context of secondary use of Electronic Medical Records (EMR) data. For example, the Office of the National Coordinator (ONC)'s Strategic Health IT Advanced Research Projects Area 4 (SHARPn) project has adopted Intermountain Healthcare clinical element models (CEMs) for normalizing patient data extracted from EMR systems, in which the XML-based technologies are typically used for data validation.

Preliminary studies have demonstrated that the emerging semantic web technologies would play an important role in enhancing the capability of clinical information models in terms of model validation and consistency checking, and EMR-driven patient cohorts identification. The objective of the position paper is to describe our preliminary work at Mayo Clinic on building a semantic web-based framework for quality assurance (QA) of EMR data for secondary use.

First, we will introduce our effort in developing a CIMI-compliant semantic web representation and services for clinical information models. The efforts include 1) converting the CIMI reference model in the Unified Modeling Language (UML) into the Web Ontology Language (OWL) rendering; 2) developing a CIMI-compliant semantic web representation and services using domain templates generated from the BRIDG model. We consider these efforts would provide standards-based infrastructure and computable semantics for enabling QA mechanisms to ensure data quality.

Second, we will introduce our preliminary work in exploring the role of SPARQL rules in QA of EMR data that are normalized in SHARPn CEM models. We utilized a "SecondaryUseNotedDrug" CEM model and its sample instance data and tested the SPARQL rules in a SPARQL Inference Notation (SPIN) framework in terms of defining inference rules and checking domain-specific constraints for data validation.

Finally, we will discuss the requirements and challenging issues (e.g., data type and valueset binding issue) on building such a semantic web-based QA framework for validating model-driven EMR data for secondary use.