

A Distributed Semantic Web Approach for Cohort Identification

Joseph Teagno¹, Richard C. Kiefer², Jyotishman Pathak, Ph.D.², Guo-Qiang Zhang, Ph.D.¹, Satya S. Sahoo, Ph.D.¹

¹Division of Medical Informatics, Case Western Reserve University, Cleveland, OH,
²Department of Health Sciences Research, Mayo Clinic, Rochester, MN

Abstract

Patient cohort identification for clinical trials and research studies often requires querying distributed data resources and subsequent data integration. In this work, we describe the use of Semantic Web technologies, specifically Resource Description Framework (RDF) and the SPARQL query language to support seamless querying and integration of distributed data sources over the Web. In addition, a key novelty of our work is the use of an intuitive ontology-driven visual query interface—Visage—that allows non-informatics experts to compose and execute complex SPARQL queries over heterogeneous data sources.

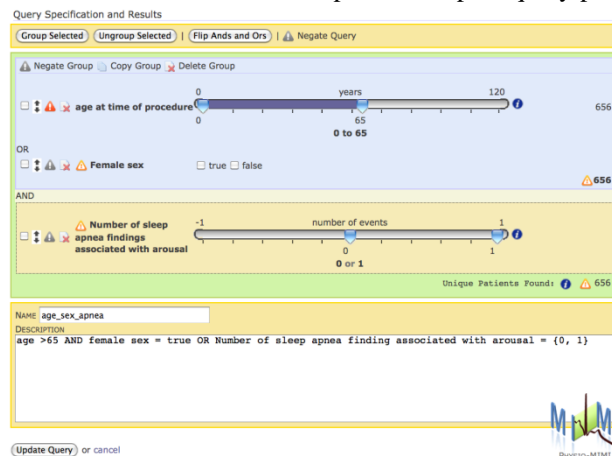
INTRODUCTION and BACKGROUND

Systematic study of clinical phenotypes is important to better understand the genetic basis of human diseases and more effective gene-based disease management. The Linked Clinical Data (LCD) project at Mayo Clinic aims to develop a semantics-driven framework for high-throughput phenotype extraction, representation, integration, and querying from electronic health records using emerging Semantic Web technologies, including RDF and SPARQL. In previous studies [1], we have demonstrated how such an approach can be used for querying public-private data sources via federated SPARQL queries, although there was no systematic way to visualize and compose the queries graphically. To address this limitation, in this work, we explore VISAGE—an ontology-driven visual query interface that was developed by the Multi-modality, Multi-resource Environment for Physiological and Clinical Research (Physio-MIMI [2]) project—a collaborative effort between four CTSA's. VISAGE maximizes its usability by using a large group of visual widgets: data source list and query term browser. It also allows comprehensive query management and exploration.

METHODS and DISCUSSION

We use an exemplar scenario for cohort identification of patients on the drug Prandin, a brand name for Repaglinide, which is used for treatment of type 2 diabetes. In our case, we first identify the side effects of Prandin using SIDER—a public repository containing information about drug adverse events—by querying its SPARQL endpoint. The Mayo patient record use RxNorm codes for representing drugs, and ICD-9-CM for diagnosis. Hence, we query the RxNorm SPARQL endpoint to get the RxNorm CUIs (concept unique identifiers), and use an internal database look up to find the relevant ICD-9 codes for a given side effect [1].

VISAGE enables users to compose a complex query pattern using terms from a “plug-and-play” domain ontology.



For example, Physio-MIMI currently uses the Sleep Domain Ontology (SDO), which categorizes all variable terms into two types, *continuous* (e.g. age 40 – 55 years) or *categorical* (e.g. gender as male and female). VISAGE uses this information to use a “slider” widget for the age variable and a radio-button for the gender variable. (Figure 1 illustrates the use of VISAGE for creating a query for sleep research.)

As part of this project, VISAGE is being implemented to translate a user-composed query into a SPARQL query. Subsequently, using a distributed SPARQL query protocol; VISAGE will execute appropriate subsections of the query against multiple data sources including Linked Open Data. The synergistic use of VISAGE in the LCD project will provide an effective

platform for leveraging Semantic Web technologies for clinical research over distributed data sources.

Acknowledgement. NIH/NCRR UL1-RR024989 and Mayo Clinic Early Career Development Award (FP00058504)

References

- [1] Pathak et al. Applying linked data principles to represent patient's electronic health records at Mayo clinic: a case report, Proceedings of the 2nd ACM SIGHT IHI, 455-464, 2012 Miami, USA
- [2] GQ Zhang et al. VISAGE: A Query Interface for Clinical Research. AMIA CRI 2010, pp. 76-80.