Linking clinicians to biomedical researchers: An application of the ISF ontology at Stony Brook Medicine

Stony Brook Medicine

Janos G. Hajagos

Viviktesh Agwan

From CSV to ISF triples

Anonymized data on billed encounters were obtained from an insurance company partially owned by the medical center. This data was in a CSV (Comma Separated Value) text file format.

The CSV file was transformed with a Python script into n-triples, a serialization format of RDF (Resource Description Format). The RDF was consistent with the ISF ontology.

The n-triples file can be viewed as a graph. This helps visualize how the different class instances are inter-related.

The ISF ontology aligning eagle-i and VIVO

The ISF or Integrated Semantic Framework is an OWL based ontology to connect two semantic web enabled applications. The ISF added two new modules for clinical expertise and for representing encounters. Instances of these classes are used in this work.

The effort to align the eagle-i and VIVO ontologies was funded through NIH's CTSA (Clinical & Translational Science Awards). The campuses involved in the alignment effort included OHSU, Cornell, UF, Harvard and Stony Brook University.

Web Ontology Language or OWL is a W3C semantic web standard for knowledge representation. Protege is a tool for authoring an OWL based ontology.

Computing clinical diagnosis experience

Using a simple proportion a weight was computed for each faculty member's clinical experience (see paper for the formula). A minimum of 10 patients with ICD9CM codes needed to be associated with a single faculty member. Faculty members had to score at least a measure of 0.1 to be counted as having experience in a clinical diagnosis subject.

MeSH: Provider Counts
Hypertension: 10
Lipid Metabolism Disorders: 7
Arthralgia: 6
Chest Pain: 6
Diabetes Mellitus: 5
Myocardial Ischemia: 5
Respiration Disorders: 5
Digestive System Diseases: 3
Arrhythmias, Cardiac: 2
Abdominal Pain: 1

MeSH associated with provider counts for the adult male population.

A SPARQL query showing the computed weights for experience.

Computing biomedical research subject area

Publication details of an article in the VIVO web application.

The MeSH associated with article to the left.

The VIVO site for SUNY REACH, which covers health science campuses, contains faculty publications from PubMed. PubMed articles in Medline have MeSH associated with each scientific publication.

RDF data from the REACH VIVO site was extracted to a triple store with a public querying endpoint.

The SKOS version of MeSH allows the MeSH to be expressed in a standard way.

The RDF version of PubMed showing articles with MeSH.

Representing vocabularies in SKOS

SKOS (Simple Knowledge Object System) is a W3C standard for thesauri, classification systems and taxonomies. SKOS representations of ICD9CM and MeSH were created. The source of the vocabularies was from the UMLS.

ICD9CM codes are used to represent diagnosis in the context of billing.

MeSH is a vocabulary that is used by PubMed/Medline to categorize published articles in the biomedical and life sciences.

Unified Medical Language System® (UMLS®)

The central concept in the UMLS is the CUI (Concept Unique Identifier). The CUI allows synonyms to be linked within a single vocabulary and across different vocabularies. The CUI will be utilized to connect the ICD9CM codes to MeSH.

SPARQL, a graph pattern language, is a W3C semantic web standard. By loading the computed experience, the RDF profiles from VIVO and SKOS based vocabularies it is possible to query across different disparate data sources. In the example to the right the clinical experience of Stony Brook medicine faculty were connected to the published subject area of biomedical researchers in the state wide SUNY REACH network.

A SPARQL query connecting clinicians to biomedical researchers.

Bridging ICD9CM codes to MeSH

Using the shared CUI between two vocabularies a hierarchical pathway between two different vocabularies can be defined. This allows ICD9CM codes to be mapped to a vocabulary that is used to categorize biomedical publications.

The figure to the right shows how the rdf:seAlso property allows connection from SKOS vocabulary to the rich concepts tied around the relationship between cancer and genes from the NCIT ontology as expressed in the UMLS.

FULL ACKNOWLEDGMENT: Moises Eisenberg was the campus principal investigator for the CTSAConnect project and the SUNY REACH project. Erich Bremer was the technical lead for semantic web technology at Stony Brook Medicine. Jizu Zhi was responsible for the PubMed searches associated with Stony Brook Medicine faculty. Tammy DiPrima was responsible for the data loading into the SUNY REACH VIVO site. Code for converting encounter data was programmed by Yuchun Cui, Amit Sahasrabudhe, and Onkar Pednekar of the CS department. Carlo Torniai and Shahim Essaid of OHSU provided guidance on mapping encounter data to the ISF. Margaret Morris assisted in obtaining the encounter data. The principal investigator of the CTSAConnect project which funded the expertise computation was Melissa Haendel of OHSU. Ann Gardner and Barbara Chandler were the local SUNY REACH administrative support. Elaine Collier of NIH was the program director for the CTSA contract.

Future directions and improvements

One of the goals of this work was to demonstrate that using the ISF ontology, SKOS based vocabularies and standards based semantic web technologies would allow database querying across traditional institutional boundaries.

The weakest part of this work is the simplistic algorithms used to compute this experience. The algorithm could be improved significantly and integrated into this pipeline. It was not the goal of the sponsored research to develop sophisticated experience calculation methodology but to demonstrate that the data could be linked. In this study, a second issue is the limited number of faculty that experience could be calculated. This was partially due to the approved IRB (Institutional Review Board) protocol.

The examples developed here could be expanded to other medical coding systems like CPT for procedures. Experience here could highlight a faculty's member experience in performing advanced surgical techniques.

A SKOS representation of CPT codes could be used to compute experience based on billed CPT codes.