i2b2 Ontology representation based on SKOS

VS Thiemann[1], R Röhrig[1], RW Majeed[1]

[1]Carl von Ossietzky University Oldenburg, Oldenburg

Background

In the AKTIN Project we use i2b2 databases to establish local data warehouses in the participating clinics[1]. The concepts of the AKTIN datasets are imported from standardized HL7 CDA documents[1-4] so that at this stage in i2b2 no further data mapping is required. However, we need a description language for the i2b2 ontology to be able to maintain it, distribute it to the local i2b2 installations and produce documentation.

In order to create or update the i2b2 ontology from a script we need SQL statements to manipulate the i2b2 database tables. It is not feasible to create and maintain these manually. The complete AKTIN ontology for the base module consists of 2102 concepts which translate to about twice that number in insert statements in i2b2. Most of these come from the ICD10GM (n=1795) and CEDIS (n=188) terminologies. The other concepts refer to LOINC if possible or use proprietary AKTIN codes.

Methods

Since SQL is not a good way to maintain ontology information we looked for relevant standard formats and found Simple Knowledge Organization System (SKOS). It is the W3C standard to represent knowledge and a variety of tools for processing is available. Using XSL transformations we were able to create a script to transform the XML release document of ICD10GM to SKOS. The CEDIS classification is available in a less structured tabular format that can also be converted into SKOS. The resulting SKOS ontology can be loaded in any Resource Description Framework (RDF) engine (e.g. Apache Jena, OpenRDF). The RDF graph is then traversed to obtain the necessary SQL insert statements for i2b2. We used XML-FO (via Apache FOP) to create a PDF documentation of the AKTIN ontology.

Results

The AKTIN ontology information is available in a machine readable form and can easily be used by different systems, which we have demonstrated by generating a human readable PDF overview from the ontology as well as an automated i2b2 import. The developed transformations are generic and can be reused in similar projects. It is feasible and efficient to manage bio-medical ontologies for clinical data warehouses using only existing international standards SKOS, XSLT, XML-FO.

Conclusion

Our approach shows that a careful application of standard formats and tools leads to a coherent workflow and less maintenance work. All proprietary or modified contents of the ontology need to be maintained in one place only: The SKOS ontology representation. It is also possible to process the ontology with external standard tools like Protégé without further transformation.

The standard catalogs (e.g. ICD) do not need any manual maintenance – if a new release is available it can be transformed to SKOS and the whole ontology can be deployed to multiple i2b2 installations, documentations, reports etc. This significantly reduces the work to maintain the ontology for AKTIN and also increases the quality because there are as few as possible manual steps which effectively prevents copy and paste errors and ensures the changes are deployed to all relevant parts of the project.

Acknowledgement: Funding by BMBF No. 01KX1319B

[1] Ahlbrandt J, Brammen D, Majeed RW, Lefering R, Semler SC, Thun S, Walcher F, Röhrig R.: Balancing the need for big data and patient data privacy--an IT infrastructure for a decentralized emergency care research database.Stud Health Technol Inform. 2014;205:750-4.

[2] Kulla M, Brammen D, Greiner F, Hörster A, Lefering R, Somasundarama R, Wrede C, Röhrig R, Erdmann B, Walcher F: Vom Protokoll zum Register - Entwicklungen für ein bundesweites Qualitätsmanagement in deutschen

Notaufnahmen. DIVI 2016; 7: 12–20
[3] AKTIN Art-Decor Webpage. Available from:http://aktin.art-decor.org/. (Last access March 11 2016).
[4] Dolin RH, Alschuler L, Beebe C, Biron PV, Boyer SL, Essin D et al. The HL7 Clinical Document Architecture. Journal of the American Medical Informatics Association : JAMIA 2001; 8(6):552–69.

Keywords: Data mapping, Health Data Management and networking, Knowledge representation and processing, Medical ontologies, Research IT infrastructures and EHR data reuse