How to Infer an Ontology of the Data Documentation Initiative

Thomas Bosch[1] and Brigitte Mathiak[2]

Abstract

In close collaboration with experts from the social, behavioral and economic sciences, ontology engineers have designed an ontology, describing the DDI data model's conceptual components which are most relevant for the discovery of microdata. XML Schemas express the vocabulary and the syntactic structure of XML documents representing both DDI data and metadata. There is a huge amount of XML document instances, which can be mapped to the RDF representation of the developed DDI ontology, in order to reuse already existing DDI (meta)data and to profit from the benefits associated with the RDF format (publish in the Linked Open Data world and link with other RDF datasets, retrieval, …).

A manual mapping, however, requires a lot of time and effort, is very error-prone, and therefore not applicable. The authors devised a generic approach which converts unexceptionally any XML Schemas to generated ontologies automatically. Domain ontologies, such as an ontology of the DDI, can be derived on the basis of the generated ontologies using a Semantic Web rule language. As a consequence, all the information, located in the underlying DDI XML Schemas and associated XML documents, is also expressed in the DDI ontology and its RDF representation. Ontology engineers and domain experts can then add supplementary domain-specific semantic information, not represented in the XML Schemas, to the DDI ontology in a subsequent step.

In this paper, multiple, representative, and for non-information scientists understandable examples show in a step by step manner how to derive excerpts of the DDI ontology out of the underlying DDI XML Schemas using the proposed domain-independent approach. The authors compare the manual approach mapping XML documents to the DDI ontology's RDF representation with the devised automatic approach and address the (dis)advantages of each attempt.

Keywords

DDI, DDI-RDF, Semantic Web, Ontology Design, XML Schema, SWRL, OWL, RDF

[1] Thomas Bosch | GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany |E-Mail: thomas.bosch@gesis.org

[2] Brigitte Mathiak | GESIS - Leibniz Institute for the Social Sciences, Köln, Germany |E-Mail: brigitte.mathiak@gesis.org