Simple visualization of structures of interrelated concepts in the FRBRoo ontology

Krzysztof Sielski, Marcin Werla

Poznań Supercomputing and Networking Center,
ul. Noskowskiego 12/14, 61-704 Poznań, Poland
{sielski,mwerla}@man.poznan.pl

Keywords: Semantic Web, ontology, OWL, RDF, FRBRoo, RDF Unit, knowledge base

The Knowledge Base which was created by Poznań Supercomputing and Networking Center (PSNC) as a part the SYNAT[1] project integrates information from distributed heterogeneous sources such as digital libraries, digital museums, scientific and technical information systems. The gathered knowledge is stored in an RDF semantic database and is represented in FRBRoo ontology with some custom extensions, which had to be introduced in order to represent all the information without any semantic loss.

As of the beginning of 2013, the Knowledge Base contained information from over 3,100,000 metadata records, which were originally encoded in various schemas: PLMET (data obtained from Polish Digital Libraries Federation), MARC 21 XML (from union catalog of Polish research libraries NUKAT), MONA (from the National Museum in Warsaw) or CDWA LITE (from the National Museum in Krakow). These records were converted to FRBRoo ontology using jMet2Ont[1] tool. Some auxiliary data sources such as VIAF, Geonames, KABA Subject Headings and Lexvo have been used to enrich the records with detailed information. Currently, the number of RDF triples building the Knowledge Base is 536M, which includes 235M explicit and 301M implicit triples. The implicit triples have been added by the inference engine with our custom rule set, which is a subset of OWL 2 RL/RDF entailment rules.

Unlike traditional relational databases, data represented as triples does not have a precise schema with strict constraints. Instead, OWL ontologies describe the structure of concepts and relations between them. As FRBRoo is a complex ontology with many classes, this model is often converted to a simpler one when presented to an user. The contents of the Knowledge Base can be explored in a couple different ways:

- a raw SPARQL endpoint, which is aimed at expert users who know the ontology very well and have precisely defined goals;

[1] SYNAT project, financed by Polish National Center for Research and Development (grant number: SP/I/1/77065/10), is aimed to conduct a research task titled "Creation of universal, open, repository platform for hosting and communication of networked resources of knowledge for science, education and open society of knowledge"

- a full text search application, which searches for keywords provided by user in RDF literals from the triplestore and uses the Query Processing Module (QPM) which maps on-the-fly information represented in the FRBRoo ontology to a simplified model, consisting of the following concepts: works, items, persons, places, legal bodies, and subjects;
- a geographical search application, which allows user to select an area on a map to find all objects connected with places contained in that area (e.g. all publications whose subject is a particular city);
- an application to explore semantic database with dynamically fetched portions of data describing particular object from the triplestore, which are presented as interrelated FRBRoo concepts in a legible way understandable by non-experts.

The last named application was built as a proof of concept of RDF Unit[2]. RDF Units are graphs which consist of several ontology objects of different classes that are needed to provide all the essential information about a certain resource. For example, an RDF Unit for a particular instance of Publication Expression from the Knowledge Base would include objects representing its Title, Publication Event and Place of Publishing, but not geographical coordinates of that place. RDF Units are dynamically constructed based on the metaproperties of ontology relations and actual data in the triplestore.

Such graphs are transformed into a tree structure, in which the examined resource becomes a root. Then, the obtained RDF Unit tree is prepared for presentation by replacing names of predicates with more user friendly labels and by flattening some long predicate paths to a single dummy edge in order to provide information in a straightforward way. This transformations are represented as a set of rules which take into consideration a predicate and classes of a subject and an object. Examples of such rules include (here `[?]` stands for any class):

- `[E21_Person] P100_i_died_in [E69_Death] P4_has_time_span [E52_Time-Span] P1_is_identified_by [?]` → *date of death*
- `[F18_Serial_Work] P148_has_component [F14_Individual_Work]` → *series element*
- `[?] P9_consists_of [F28a_Contribution] P14_carried_out_by [?]` → *contributor*
- `[?] P9_consists_of [F28a_Contribution] P2_has_type [?]` → *in the role of*

Figure 1 presents a result of mapping one record in MARC 21 XML schema to FRBRoo. It is a graph of 47 connected FRBRoo objects represented by 108 RDF triples. Figure 2 presents a view in our application that represents an RDF Unit of an Individual Work resource which was created in mentioned mapping. This unit contains all the information from source record except for author's and contributors' dates of life, which can be examined in those resources' view.

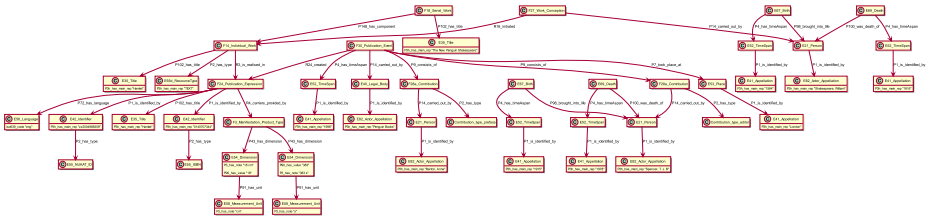

Fig. 1 A result of mapping a single metadata record from MARC XML to FRBRoo represented as a graph. An image in high resolution can be viewed at http://bit.ly/frbroo_ham

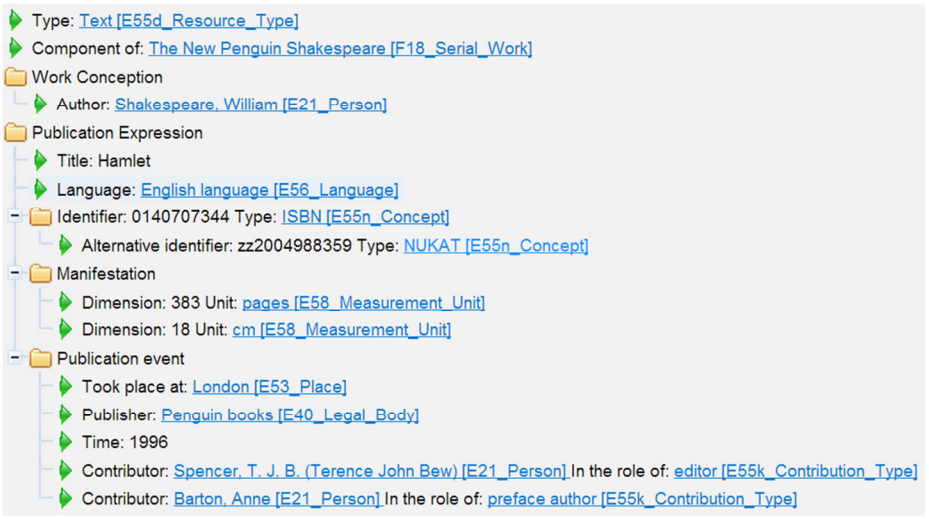

Fig. 2 A representation of F14_Individual_Work object (*Hamlet* by Shakespeare) from Fig 1 converted to a simplified tree compatible with FRBRoo ontology for presentation

The described Knowledge Base browser application was prepared for dynamic viewing of FRBRoo data from the triplestore, but this approach is generic and should work for another ontologies as well. It uses no predefined SPARQL queries and is based only on a relatively small configuration: a set of graph path flattening rules for presentation and a set of single metaproperty for each ontology predicate which is used to build RDF Units.

References

1. Walkowska, J., Werla, M.. (2012). Advanced Automatic Mapping from Flat or Hierarchical Metadata Schemas to a Semantic Web Ontology. TPDL 2012. Lecture Notes in Computer Science, vol. 7489, pp. 260-272.
2. Sielski, K., Walkowska, J., Werla, M.. (2013). Methodology for Dynamic Extraction of Highly Relevant Information Describing Particular Object from Semantic Web Knowledge Base. TPDL 2013. Lecture Notes in Computer Science, vol. 8092, pp. 260-271.