DCPAPERS
Proc. Int'l Conf. on Dublin Core and Metadata Applications 2011

Semantic Bibliography Based on Ontology and Linked Data

Haiyan Bai
Institute of Scientific and Technical Information of China, China
bhy@istic.ac.cn

Xiaodong Qiao
Institute of Scientific and Technical Information of China, China
qiaox@istic.ac.cn

Bing Liang
Institute of Scientific and Technical Information of China, China
liangb@istic.ac.cn

Keywords: Semantic bibliography; linked data; bibliography ontology

Our project arises from the need for bibliography organization and integration at the National Science and Technology Library (NSTL[1]) of China. NSTL consists of nine special national libraries serving basic sciences, agricultural sciences, medical sciences and engineering. The union catalog emphasizes on academic resources, including more than 20,000 journals, 100,000 proceedings and 10,000 reports. Traditional bibliography organization methods provide index points and access points based on MARC fields or subfields, but it does not distinguish information objects extracted from bibliography and does not show the hierarchical or related relationships between them. Linked Data principles (Wei, 2010) have already been applied to bibliographic data (Malmsten, 2008) (Zeng, 2009). The goal of our project is to follow them to build a mechanism to identify, describe and organize the characteristics and relationships of all kinds of bibliography objects, so that end users can access and browse them. We focus on the multiplicity of information forms (e.g. the book Harry Potter has forms of printed book in original text or in translation of hard cover/soft cover, audio book, movie & TV and games etc.), the variability in information life circle (e.g. a journal may change its title or ISSN, or change or merge to a new journal) and the complexity of hybrid objects (e.g. a book with its chapters, photos or tables, an article with manuscript, revision, preprint and publication).

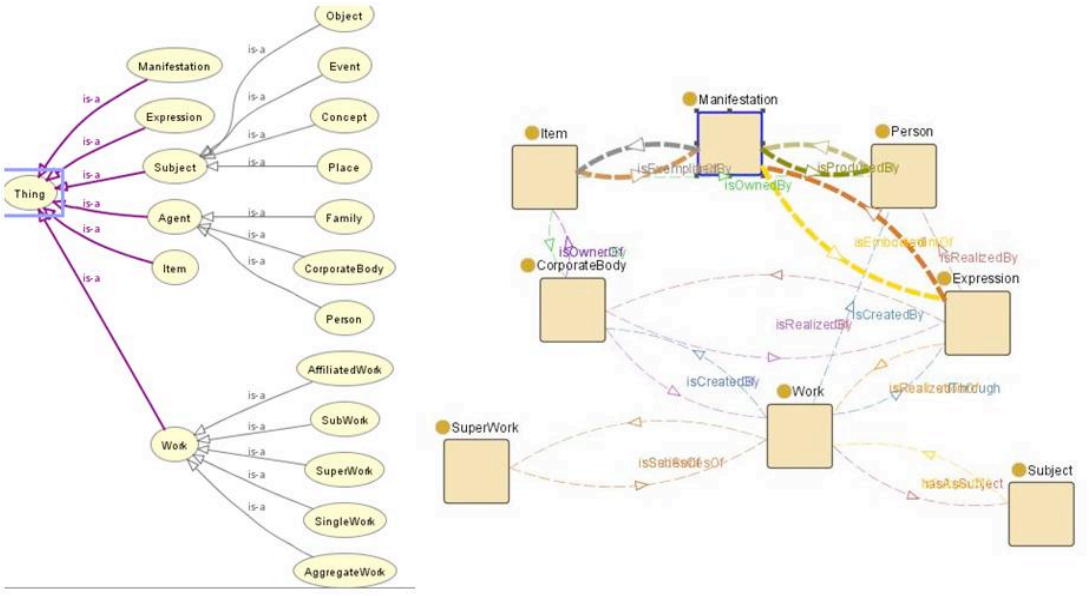
1) Constructing NSTL bibliography ontology.

Our NSTL ontology[2] re-uses common vocabulary sets such as Dublin Core[3]， RDA Entities[4], Bibliographic Ontology[5], MarcOnt[6] etc. to build multidimensional organization pattern. For example, the classes Work, Manifestation, Expression, Item etc. we referenced are from FRBR. We refined the class Work into AggregateWork, AffilliateWork, SubWork, SuperWork and SingleWork. AggregateWork is used to represent series book, conference, multivolume which include more than one single work at least; SuperWork is used to represent journals which include a group of journals may change or succeed to each other (Integration Ontology for Bibliographic Description; Synak & Kruk).

The NSTL bibliography ontology includes 18 classes, 31 object properties and 379 data properties.

[1] http://www.nstl.gov.cn
[2] http://168.160.18.212/nstlontology.owl
[3] http://purl.org/dc/elements/1.1/
[4] http://RDVocab.info/Elements
[5] http://bibotools.googlecode.com/svn/bibo-ontology/trunk/doc/index.html
[6] http://semdl.info/books/2/appendices/G

https://doi.org/10.23106/dcmi.952135832



(a) Class and subclass (b) Class and relationship

2) Transfering organization pattern.

We processed NSTL bibliography data by document type and extracted some entities into different tables in relationship database. For example, we build AggregateWork table which include MARC data of conference, serials books and multivolume books. And other entities such as expression, manifestation and relationships between them are resolved based on MARC subfield and some rules (Displays for Multiple Versions; AquaBrowser Library FAQ). Using open source software D2R, NSTL bibliography ontology is mapped to our database schema. This tool transfers database into RDF file using table names as class names and column names as property names, and generates linkage between entities and their attributes and properties. In advance, we customized the mapping file and add some conditions to generate more no external linkage with terms from our ontology. (Bizer, Cyganiak, & Heath; D2R Server; Bizer & Cyganiak)

3) Publishing linked data.

Using D2R, linked data[7] is published and accessed by HTML browsing, URI referencing, SPARQL endpoint and RDF dump. Linked data web with richly structure provides uniform data model, consistent semantic describing method and standard SPARQL access measures.

As the result of the project, the linked data published by D2R can be used as a demo of semantic OPAC to support the following:

1) Dynamic facet query. Because it is easy for RDF file to extract attributes of a specific class and property, it is available to facet query dynamically based on specific object in any phase of query and browsing. For example, the SPARQL statement to get all object properties and data properties show as following:

```
prefix  istic: <http://www.istic.ac.cn/ontologies/2009/10/5/ontology_bib.owl>
Select  ?n  Where { ?n rdf:type owl:ObjectProperty}
Select ?n Where {?n rdf:type owl:DatatypeProperty}.
```

Figure 1 shows the results of object property search in Protégé with the RDF file exported from our D2R Server.

[7] http://168.160.18.212:2020/

https://doi.org/10.23106/dcmi.952135832

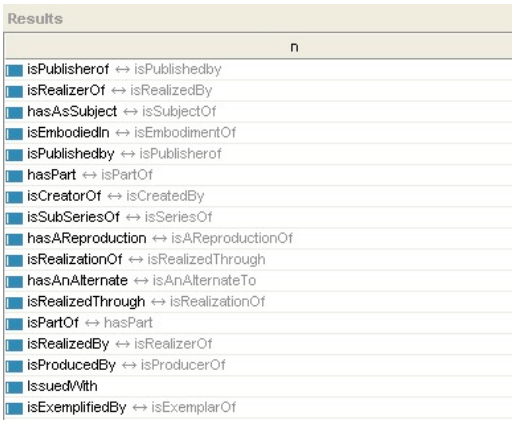
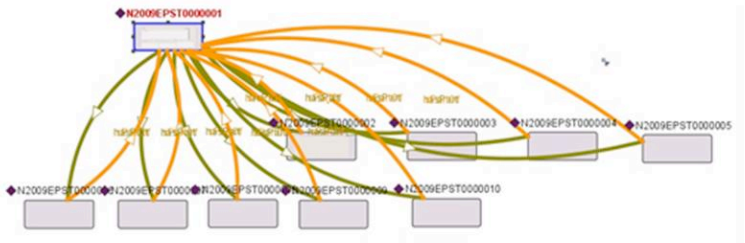


FIG. 1. Result of object property search to support dynamic facet query

2) Semantic relationship browsing and query. User can browse and get an object and related objects by following links between classes (see FIG 2(a)). And user can also query them by using relationship as conditions. For example, we imported RDF file from D2R into Protégé using dump tool and got the visualization search results .In this example, HasPart can be used as query condition to find out collection work journal BBA and its 9 sub-works. (see FIG 2(b))

(a) HTML view about aggregated work (journal BBA) and its sub-work



(b) Visualization of query result of journal BBA and its sub-works

FIG. 2. Semantic relationship browsing and query

3) Complex relationship query. SPARQL is a query language for pattern matching against RDF graphs, so it supports relationship matching and it is helpful to complex relationship query.

In further work we will attempt to apply semantic technology and deeply sequencing methods in wider scope and more information objects such as STM abstract, citation and keywords and subjects to improve the ability of resource recovery.

https://doi.org/10.23106/dcmi.952135832

References

AquaBrowser Library FAQ.[EB/OL].[2009-08-10].http://www.aip.cz/download/20080909_AquaBrowser_FAQ.pdf

Bizer, Chris, Richard Cyganiak and Tom Heath. How to Publish Linked Data on the Web. [EB/OL].[2010-02-22]. http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/20070727/

Bizer, Chris and Richard Cyganiak.The D2RQ Platform v0.7 - Treating Non-RDF Relational Databases as Virtual RDF Graphs. http://www4.wiwiss.fu-berlin.de/bizer/d2rq/spec/#specification

Displays for Multiple Versions from MARC 21 and FRBR.[EB/OL].[2010-03-25].http://www.loc.gov/marc/marc-functional-analysis/multiple-versions.html

D2R server, Publishing Relational Databases on the Semantic Web ,http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/

Integration Ontology for Bibliographic Description Formats. [EB/OL]. [2010-03-25]. http://www.marcont.org/marcont/pdf/DC2005skmskz.pdf.

Malmsten, Martin. (2008) Making a Library Catalogue Part of the Semantic Web. Retrieved April 30, 2011, from http://dcpapers.dublincore.org/ojs/pubs/article/view/927/923.

Synak, Marcin and Sebastian Ryszard Kruk . MarcOnt Initiative the Ontology for the Librarian World. [EB/OL]. [2010-03-25]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.114.3986&rep=rep1&type=pdf

Wei, Liu. (2010). Linked data ABC and recent progress. Retrieved April 30, 2011, from http://www.slideshare.net/keven/abc-5044639

Zeng, Marcia Lei. (2009). Linked library data. Retrieved April 30, 2011, from .http://202.114.9.60/dl6/pdf/26.pdf

https://doi.org/10.23106/dcmi.952135832