

Logical Expressiveness of Semantic Web Languages for Bibliographic Information Modeling

Karen M. Wickett

Graduate School of Library and Information Science, University of Illinois Urbana-Champaign
501 E. Daniel Street, Champaign, IL 61820, USA
wickett2@illinois.edu

ABSTRACT

The Semantic Web promises powerful new functionality for bibliographic databases by combining bibliographic information with knowledge about the world. However, the kinds of representation that are possible in Semantic Web languages are not widely understood in the LIS community. In particular, the W3C Semantic Web ontology languages RDFS and OWL are not sufficiently expressive to deliver the full range of inferences anticipated for bibliographic applications, and will require formalized rules from other specifications.

1. INTRODUCTION

The LIS community has taken notice of the potential of Semantic Web technologies for bibliographic information modeling. In contrast with the characteristic focus of the library automation movement on the bibliographic record as a single unit of description, Semantic Web technologies such as the Resource Description Framework (RDF) take a modular approach, where descriptive information is represented as assertions relating individual things. The reconciliation of these distinct approaches to modeling bibliographic information is underway and has required building a shared understanding what bibliographic entities there are and what relationships obtain between those entities. This movement introduces exciting possibilities for bibliographic systems that make use of knowledge about related entities and the world at large to enhance discovery and access. However, the capabilities and limitations of Semantic Web languages for modeling and inferencing are an important aspect of the movement that is still not well understood by the LIS community.

2. LIBRARIES AND THE SEMANTIC WEB

Considerable progress has been made in developing the shared understanding necessary for distributed bibliographic information modeling and inferencing. IFLA's *Functional Requirements for Bibliographic Records* (FRBR) and the lively discussions around it have had a profound effect. A Dublin Core Metadata Initiative community is currently working to

unite library cataloging practices with Semantic Web technologies by publishing descriptive elements from *Resource Description and Access* (RDA), the latest Anglo-American cataloging code, as properties in the Resource Description Framework Schema language (RDFS) [1].

However, the real promise of using Semantic Web technologies for bibliographic information modeling lies in more than the representation of library cataloging data in RDF triplestores. What is particularly exciting is not the RDF representation itself, but the kind of systems for description and access that can be built to exploit such a representation. The Semantic Web vision is so captivating because of the potential for using information about related objects to perform inferences that lead to new knowledge about an object. This suggests search and retrieval systems that go far beyond the restricted capabilities of keyword search and Boolean combinations. Several authors have noted the possibility of using ontologies to perform inferencing tasks with respect to descriptive bibliographic information [1][2]. They propose, for example, a situation where classification terms are assigned automatically based on some set of related descriptive information.

In addition to a shared vision of the bibliographic domain, we must establish a solid understanding of the kinds of inferencing and representation possible within the various Semantic Web languages. This is a simple task compared to developing consensus about the conceptual makeup of the bibliographic universe. Unfortunately the logical capabilities and limitations of these languages remain poorly understood in the LIS community. It is essential now to move this discussion forward and address the kind of Semantic Web knowledge representation that would be necessary to perform intelligent inferencing on bibliographic data. We take an example from a project to model one kind of bibliographic information: the relationships between metadata assigned to collections and metadata assigned to items in those collections. Our analysis leads to the conclusion that ontologies of the sort that can be articulated with OWL may not be sufficiently expressive on their own to represent the knowledge necessary for generating inferences of the kind mentioned above.

3. MODELING A BIBLIOGRAPHIC RELATIONSHIP

Three categories of relationships make up the main body of the Collection/Item Metadata Relationships (CIMR) frame-

work: attribute/value propagation, value propagation, and value constraint [6]. The categories are defined in first order logic with formulas which characterize attributes as two-place predicates where the first argument is the object of description and the second argument is the value of the attribute for that object. For example, the informal definition for value propagation –

Def v-p 1: an attribute A v-propagates to an attribute $B =_{df}$
if a collection has the value z for A , then every item in the collection has the value z for B

is expressed in first order logic as:

Def v-p 2: an attribute A v-propagates to an attribute $B =_{df}$
 $\forall x \forall y \forall z [(IsGatheredInto(x, y) \& A(y, z)) \supset B(x, z)].$

In order to make these descriptions widely useful in the design of search and discovery tools, the project architects suggest using the Web Ontology Language (OWL) to encode categories and inference rules. The inferences that would be supported by value-propagation are in line with proposals for systems that make use of information about related objects to create new facts. In this case, information about a collection and the fact that some item is a member of that collection are used to infer information about the item.

OWL is a powerful Semantic Web language, equipped to allow the description of complex concepts using logical connectives and the specification of a property as the inverse of another property [5]. OWL is also known to be expressively equivalent to a description logic with appealing computational properties [4]. As such, it is possible to express both facts about individuals and facts about classes and relationships between classes. OWL is well known for incorporating class constructors that correspond to the full set of connectives from first order logic. This gives the impression that the language (or at least OWL Full, the most expressive level) can represent anything that can be said with first order logic.

However, this is not the case. Despite the high expressive power of OWL, especially in comparison to other Semantic Web languages like RDF and RDFS, it is not possible to express definitions like the one given above. Inferences using value-propagation will depend on the use of the same value for some collection-level property and some item-level property. The essential feature here is that while this value is not known, the case can be described generally using an individual variable (shown in this example as "z").

OWL allows for the description of classes based around particular individuals, but does not allow the use of an unknown individual (i.e. a variable) to describe a class [4]. The problem comes down to the fact that the information in the definition of value propagation is in the form of a general inference rule, and is therefore out of the scope of OWL. In order to incorporate this information to build a Semantic

Web system for bibliographic services, it would be necessary to use a formalism designed around expressing rules, such as the Semantic Web Rule Language (SWRL) or RuleML.

Rule languages operate on a basis drawn from logic programming and can express conditional statements that make use of individual variables. There is a known division between what can be expressed in a description logic and what can be said with a system equipped to express rules [7] and major consequences with respect to the computational requirements for using a rule language. These issues need to be understood in order to build systems that fully exploit bibliographic information in a semantic web environment.

In some sense, considering the form and function of the definition for value propagation as shown above, it is not at all surprising that expressing the full CIMR framework would require a rule language. However, the LIS community has maintained that problems related to rules, and the proof and logic layers of the "Semantic Web Layer Cake" were far off in the future [2]. Now it seems that they are just around the next bend.

4. CONCLUSION

Scholars and system builders in LIS need to understand the expressive requirements for creating advanced bibliographic management systems that fully utilize the information and technology at our disposal. Many issues with important consequences for such systems remain poorly understood within the LIS community. In addition, Semantic Web languages such as OWL and SWRL are recent recommendations and still undergoing revisions. We cannot afford to miss the opportunity to inform the development and refinement of these languages and bring about a kind of collaboration that will be very beneficial for both communities [3].

5. REFERENCES

- [1] K. Coyle. Meaning, Technology, and the Semantic Web. *The Journal of Academic Librarianship*, 34(3):263–264, 2008.
- [2] S. Gradmann. rdfs:frbr—Towards an Implementation Model for Library Catalogs Using Semantic Web Technology. *Cataloging & Classification Quarterly*, 39(3-4):63–75, 2005.
- [3] J. Greenberg. Advancing Semantic Web via Library Functions. *Cataloging & Classification Quarterly*, 43(3):203–225, 2007.
- [4] I. Horrocks, P. Patel-Schneider, and F. van Harmelen. From SHIQ and RDF to OWL: the making of a Web Ontology Language. *Web Semantics: Science, Services and Agents on the World Wide Web*, 1(1):7–26, 2003.
- [5] D. McGuinness, F. van Harmelen, et al. *OWL Web Ontology Language Overview*, 2004. W3C Recommendation.
- [6] A. Renear, K. Wickett, R. Urban, D. Dubin, and S. Shreeves. Collection/Item Metadata Relationships. In *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 2008.
- [7] N. Shadbolt, T. Berners-Lee, and W. Hall. The Semantic Web Revisited. *IEEE Intelligent Systems*, 21(3):96–101, 2006.