The Return of the Trivial: Problems Formalizing Collection/Item Metadata Relationships

Allen H. Renear renear@uiuc.edu
Karen M. Wickett wickett2@uiuc.edu
Richard J. Urban rjurban@uiuc.edu
David Dubin ddubin@uiuc.edu
Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, IL 61820

ABSTRACT
Formalizing collection-level/item-level metadata relationships encounters the problem of trivial satisfaction. We offer a solution related to current work in IR and ontology evaluation.

Categories and Subject Descriptors
H.3.7 [Digital Libraries] – collection, standards, systems issues.

General Terms
Management, Documentation, Design, Standardization.

Keywords: Metadata, Collections, Logic, Dublin Core.

1. INTRODUCTION
Contemporary retrieval systems which search across collections ignore collection-level metadata; an obvious problem[1][4]. We are developing a logic-based framework for classifying collection/item metadata relationships and formalizing inference rules. An example is *attribute/value propagation*: whoever owns a collection owns each item (cf. *marcrel:OWN*). Another is *value-propagation*: if a collection's *cld:itemType* is "image", all items have *dc:type* "image". The formalization of simple propagation seems easy. However a known problem representing conditionals, trivial satisfaction, proves surprisingly troublesome.

2. CONJECTURES AND REFUTATIONS
A simple formalization of attribute/value propagation would be:

D1: An attribute A *a/v-propagates* =df
$\forall x \forall y \forall z$ [(IsGatheredInto(x,y) & A(y,z)) $\supset$ A(x,z)]

(*IsGatheredInto*, is from the DCMI DC AP; we assume: [IsGatheredInto(x,y) $\supset$ (Member(x) & Collection(y))]).

D1 is too broad: attributes with no value for any collection count as a/v-propagating although they are not. E.g., a particular collection identifier attribute (*acme:collIdentifier*) counts as a/v-propagating if no identifier has been assigned to a collection.

The standard move here is to modalize the conditional:

D2: An attribute A *a/v-propagates* =df
$\Box\ \forall x \forall y \forall z$ [(IsGatheredInto(x,y) & A(y,z)) $\supset$ A(x,z)]

That is, an attribute **A** a/v-propagates if and only if it is impossible for a collection to have *v* for **A** and its items not have *v* for **A**. But although D2 no longer incorrectly counts our unassigned collection identifier as a/v-propagating, five sorts of attributes still generate similar counterintuitive results:

C1) Impossible attributes: $\Box\ \forall y \forall z \sim$A($y$,$z$)

C2) Necessary attributes: $\Box\ \forall x \forall z$ A(x,z)

C3) Impossible for collections: $\Box\ \forall y \forall z \sim$[Collection($y$) & A($y$,$z$)]

C4) Necessary for members: $\Box\ \forall x \forall z \sim$[Member($x$) & ~A($x$,$z$)]

C5) Domain universal: $\Box \sim \exists x \exists y \exists z$ [A(y,z) & ~A(x,z)]

JCDL'08, June 15–20, 2008, Pittsburgh, Pennsylvania, USA.
ACM 978-1-59593-998-2/08/06.

3. SOLUTION
We address this problem with preemptive modal restrictions. The union of the counterexample classes is a subset of the union of C3,C4,C5; so we exclude all five classes by excluding those three:

D3: An attribute A *a/v-propagates* =df

I. a) $\Diamond\ \exists y \exists z$ [Collection(y) & A(y,z)] &

b) $\Diamond\ \exists x \exists z$ [Member(x) & ~A(x,z)] &

c) $\Diamond\ \exists x \exists y \exists z$ [A(x,z) & ~A(y,z)] &

II. $\Box\ \forall x \forall y \forall z$ [(IsGatheredInto(x,y) & A(y,z)) $\supset$ A(x,z)]

A corresponding definition of *v-propagation* would be:

D4: An attribute A *v-propagates* to an attribute B =df

I. a) $\Diamond\ \exists y \exists z$ [Collection(y) & A(y,z)] &

b) $\Diamond\ \exists x \exists z$ [Member(x) & ~B(x,z)] &

c) $\Diamond\ \exists x \exists y \exists z$ [A(x,z) & ~B(y,z)] &

II. $\Box\ \forall x \forall y \forall z$ [(IsGatheredInto(x,y) & A(y,z)) $\supset$ B(x,z)]

[a/v-propagation can now be identified as a special case of v-propagation: **A** a/v-propagates if and only if **A** v-propagates to **A**.]

4. SIGNIFICANCE
The problem is real. D1 and D2 classify non-propagating attributes as propagating, supporting invalid inferences. There are similar arguments in IR for the inadequacy of non-modal analyses of relevance[5][3]. The view that material implication is "harmless"[5] ignores the fact that nested conditionals occur routinely in definitions and system specifications.

Connections with strategies in other areas. The Barcan and converse Barcan formulas establish relations between our analysis of propagation using *de dicto* necessity and metaproperties based on *de re* necessity which are used in ontology evaluation[2].

5. ACKNOWLEDGMENTS
Supported by a 2007 Institute for Museum and Library Services (IMLS) grant hosted by the GSLIS Center for Informatics Research in Science and Scholarship (CIRSS). Other contributing project members include: Timothy W. Cole, Thomas Dousa, Myung-Ja Han, Amy Jackson, Mark Newton, Carole L. Palmer, Sarah L. Shreeves, Michael Twidale, Oksana Zavalina.

6. REFERENCES
[1] Foulonneau, M., Cole, T. W., Habing, T. G., & Shreeves, S. L. 2005. Using collection descriptions to enhance an aggregation of harvested item-level metadata. In Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries. ACM Press, New York, NY.

[2] Guarino, N. and Welty, C. 2004. An overview of OntoClean. In S. Staab and R. Studer, eds, The Handbook on Ontologies. Springer.

[3] Lalmas, M. 1998. Logical models in information retrieval. Information Processing and Management. 34, 1.

[4] Palmer, C. 2004. Thematic research collections. In S. Schreibman, R. Siemens, & J. Unsworth (eds.). Companion to Digital Humanities. Oxford: Blackwell.

[5] Sebastiani, F. 1998. On the role of logic in information retrieval, Information Processing and Management 34, 1.

[6] Van Rijsbergen, C. J., 1986. A non-classical logic for information retrieval. The Computer Journal 29, 6