

Exhibition: A Problem for Conceptual Modeling in the Humanities

Allen H. Renear (*renear@uiuc.edu*)

GSLIS, University of Illinois at Urbana-Champaign

Jin Ha Lee

GSLIS, University of Illinois at Urbana-Champaign

Yunseon Choi

GSLIS, University of Illinois at Urbana-Champaign

Xin Xiang

GSLIS, University of Illinois at Urbana-Champaign

Contents

1. Introduction
2. Exhibition
3. An Example
4. Why Exhibition is a Problem for Modeling
5. Relevant Work
6. False Resolutions
7. Conclusion

Introduction

There has recently been increased interest within the humanities computing community in formalizing the 'semantics' of document markup (e.g., Sperberg-McQueen et al., Buzzetti, Witt, Renear et al. 2002, Bayerl et al., Dubin et al., Sasaki), or, in an alternative characterization, developing 'conceptual models' that generalize the representation of the textual structures (Cover). We endorse this agenda, which has been a long time in the making (Raymond et al.), but we also wish to draw attention to some difficulties that may be unique to cultural material.

Natural human communication is characterized by what might be termed *plenary semiosis*. Without waiting for formal languages to be provided humans immediately proceed to attempt to say everything they think, and, at least arguably, they generally succeed. The result is that natural communication systems exhibit every imaginable feature that troubles knowledgeable engineers: fuzzy predicates, modal notions, non-extensional contexts, incompleteness, inconsistency, ambiguity, and so on. But there is an additional complexity as well. Human communication takes place within social contexts that, as linguists and philosophers have been telling us for some time, confound efforts to conceptualize it as sets of assertions only. These two aspects of human communication, plenary semiosis and multiple interacting levels of non-assertional representation combine to produce some of the most difficult, and significant, features of communicative artifacts. We describe one of these features and argue that unless current conceptual modeling systems are extended to accommodate this and other related features those systems will be inadequate for the representation of cultural objects.

Exhibition

In ordinary linguistic communication we often use a name to refer to something in order to then go on to attribute some property to that thing. However when we do this we do not naturally construe our linguistic behavior as being at the same time an assertion that the thing in question has that name. We do however have a particular cognitive relationship to this latter state-of-affairs; it is just that this attitude is not one of assertion — we *rely* on, or are *committed* to, or *presuppose* that the thing in question has the name we are using to refer to it, but we are not *asserting* that it does.

We refer to this relationship as *exhibition*. We say that the brief document/utterance "*Moby Dick* was written by Herman Melville" *exhibits* the state of affairs that "the name of the author of *Moby Dick* is 'Herman Melville'", but it does not *assert* that state of affairs. What it does assert is that Melville is the author of *Moby Dick*. Although naming is our prototypical example of exhibition in this paper, we believe that exhibition is a widespread and diverse phenomenon.

An Example

Consider this XML markup, adapted from the *TEI Guidelines* (P4):

```
<bibl>
<author>Edward R. Tufte</author>
<title>Envisioning Information</title>
<pubPlace>Cheshire, Conn.</pubPlace>
```

```
<publisher>Graphics Press</publisher>
</bibl>
```

The *Guidelines* characterize these element types as follows:

- `author`: "... contains the name of the author(s), personal or corporate, of a work ...".
- `title`: "... contains the title of a work ...".
- `publisher`: "...provides the name of the organization responsible for the publication ... of a bibliographic item".
- `pubPlace`: "contains the name of the place where a bibliographic item was published."

Close reading of these definitions reveals that these markup tags convey two quite different sorts of information:

Set A

1. Edward R. Tufte authored *Envisioning Information*.
2. *Envisioning Information* was published by Graphics Press.
3. *Envisioning Information* was published in Cheshire, Connecticut.

Set B

1. The name of the author of [this book] is "Edward R. Tufte".
2. The name of the publisher of [this book] is "Graphics Press".
3. The name of the place where [this book] was published is "Cheshire, Connecticut".

Why this is a Problem for Modeling

First, note that the markup is overloaded. The markup tag `author` is used to say that something is a name, *and* it is also used to say that someone is an author (or the author of a particular book). Consider a representation in any commonly used data modeling language, say, RDF's graph-based representation: nodes for individual entities and arcs for binary relationships between them. We would expect a single arc for the assertion represented by a single element — but here apparently a single element must be unpacked into two arcs. TEI specialists are fond of saying that TEI markup is about the text, not about the world the text is about; but we see plainly this isn't always so. And we also note that this overloading crosses a profound and famously troublesome semantic boundary: that between *using* an expression and *mentioning* it.

But the most revealing feature of this analysis is that when we take the union of assertions in sets A and B we will have a model of possible semantic content that, as a whole, is almost certainly incorrect; at least in this sense: it is unlikely that there is any single communicative object whose semantics is correctly modeled by this set of assertions. There are two cases to consider; we present them using some terminology ('expression', 'work') from the *Functional Requirements for Bibliographic*

Records (IFLA 1997) and distinguish two senses of 'XML document' (Renear et al. 2003).

1. Consider first an XML document that is understood to be a symbolic expression realizing an intellectual work such as, say, a manual about web design. Such a document will be correctly understood as making the assertions in Set A, but not as asserting any of the assertions in Set B.
2. Now consider an XML document that is a transcription of a source text, that is, a document that is an expression realizing a work which is itself a "theory of the text" (Sperberg-McQueen); that text (expression) being the text of the manual (a work). Such a document would generally be understood as making the assertions in Set B, but not as asserting any of the assertions in Set A.

As a consequence a correct graph model for either case cannot represent the assertions (as assertions) in the other case. However a correct representation of Case 1 could represent Case 2 assertions as exhibitions — if specific expressive devices, qualified arcs say, were available for this representation. This is the extension that we are recommending.

Some clarification of these intricacies may be useful. First note that the cases are not isomorphic: Case 1 asserts the propositions in Set A and exhibits those in Set B, but Case 2, although asserting the propositions in Set B, does *not* exhibit the propositions in Set A. While might be plausibly argued that the propositions in Set B logically imply those in Set A, and so any document that asserts Set B asserts Set A, we would resist this for two reasons: first because the intuitive logic of assertion simply does not seem to require that all logical implications of asserted propositions are themselves asserted; and second because we suspect that a completely correct presentation of Set B, one more in line with TEI doctrine on the textual orientation of markup, would eliminate all commitments to books, authors, and authorship, and that paraphrase would block the logical implications in any case. What one could say however is that in Case 2 the Set A propositions occur in *oratio obliqua*.

We also note, as an illustration of the usefulness of the concept of exhibition, that scholarly transcription into TEI markup can be understood as identifying exhibitions and then re-expressing them as assertions.

Relevant Work

The rudiments of this problem have already made an appearance in the Semantic Web and Dublin Core communities. However we do not think its significance, at least for cultural material involving human communication, is fully appreciated. Dan Brickley, chair of the W3C Semantic Web Interest Group has noted that the Dublin Core `dc:creator`

element is defined in a way that encourages a similar confusion between names and things (Brickley), not surprisingly, as the definition of `dc:creator` is similar in logical structure to the ones we cite from the *Guidelines*:

"Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity." (DCMES 2003)

The varying usage of the `dc:creator` code (sometimes for Creators, sometimes for their names) amongst metadata encoders is now recognized as a serious practical problem for the development of an 'abstract model' for Dublin Core. (Powell).

It is of course in the context of efforts to be absolutely precise and formal that the problem is acute. Jeremy Carroll, co-editor of the W3C Recommendation *Resource Description Framework (RDF): Concepts and Abstract Syntax*, writes in a posting on `w3c-rdfcore-wg`:

I have been looking through the (RDF) primer, particularly looking at the Dublin Core examples (throughout the primer). These seem like perfectly fair examples of how Dublin Core is used. Unfortunately, there are many instances where strings are used to represent people and things rather than themselves. This is not in agreement with the model theory...

(Carroll)

Carroll then goes on to note that given the RDF model theory incorrect implications immediately ensue: in our example for instance, that *Moby Dick* was authored by a string rather than by a person.

False Resolutions

Three deflationary perspectives on this problem are possible.

One, anticipated from the TEI community, is that TEI encoding always represents the features of the linguistic text only and 'real-world' assertions are either misunderstandings, mistakes, or anomalies. This may be so, although we are skeptical as to whether this stance can be maintained with respect to the full range of TEI applications. But in any event exhibition remains a common feature of communicative artifacts, characteristic of many XML element sets, and of many other systems of symbolic communication. It must be accommodated.

Another approach, this one anticipated from the Semantic Web community, is simply to insist on an unambiguous corrected conceptual representation: one arc for being named "Herman Melville", one for authoring *Moby Dick*. But this resolution fails for the reasons presented in the preceding section. Although this model would be in some sense an accurate representation of "how the world is" according to the document, it would not represent what is *asserted* by the document. The authorship arc

in the corrected RDF graph model will correspond to relationships of exhibition, not assertion; and there is no accommodation for this distinction in the modeling language.

Finally, it is also natural to feel that the phenomenon of exhibition is similar in some respects to the already noted much studied phenomenon of linguistic *presupposition* and to wonder whether exhibition is simply a special case of presupposition (Levinson). Currently we are undecided on this issue but we note that even if exhibition does turn out to be a form of presupposition that would remove neither the difficulty exhibition creates for conceptual modeling, nor its intellectual significance. In fact it would be a rather substantial finding to determine the matter one way or the other.

Conclusion

The phenomenon of exhibition is not limited to the simple naming examples used above. We believe it is characteristic of communication and communicative cultural artifacts in general. For instance when we title our articles we do not say that the title is a title, although we exhibit it as a title, allowing that inference to be drawn (Renear). Or for a quite different sort of case: consider how morphological distinctions exhibit our commitments to syntactical roles, without actually asserting that the words in question are playing those roles — though indeed we use those words with those particular grammatical and syntactical properties in order to make the assertions we do make.

We conclude that current conceptual modeling projects within the humanities computing community will fail to be adequate for the study of cultural objects if they take the approach of the Semantic Web community and see exhibition as a simple problem of ambiguity or error, rather than defining new constructs to express these distinctive relationships. To be adequate for the humanities, conceptual modeling must be extended to accommodate the data of the humanities.

Bibliography

- Bayerl, P.S., H. Lungen, D. Goecke, A. Witt, and D. Naber. "Methods for the Semantic Analysis of Document Markup." *Proceedings of the 2003 ACM symposium on Document Engineering*. ACM Press, 2003. 161-170.
- Brickley, D. *Using Dublin Core Creator*. FOAF Wicki, July 2003. Accessed 2005-03-21. <<http://rdfweb.org/topic/UsingDublinCoreCreator>>
- Buzzetti, D. "Digital Representation and the Text Model." *New Literary History* 33.1 (2002): 61-88.

Carroll, J. *Dublin Core, the Primer and the Model Theory*. Posting in w3c-rdfcore-wg, May 16, 2002 10:32:42. Accessed 2005-03-21. <<http://lists.w3.org/Archives/Public/w3c-rdfcore-wg/2002May/0040.html>>

Cover, R. *Conceptual Modeling and Markup Languages*. Cover Pages, January 24, 2001. Accessed 2005-03-21. <<http://xml.coverpages.org/conceptualModeling.html>>

Dubin, D., C.M. Sperberg-McQueen, A. Renear, and C. Huitfeldt. "A Logic Programming Environment for Document Semantics and Inference." *Literary and Linguistic Computing* 18.2 (2003): 225-233.

Dublin Core Metadata Element Set. Version 1.1 Reference Description. DCMI, 2003. Accessed 2005-03-21. <<http://dublincore.org/documents/dces/>>

International Federation of Library Associations (IFLA). Functional Requirements for Bibliographic Records: Final Report. UBCIM Publications-New Series. Munchen: K.G.Saur, 1998.

Levinson, S.C. "Chapter 4: Presupposition." *Pragmatics*. Cambridge: Cambridge University Press, 1983. 167-225.

Powell, A. *DOAP*. Posting in "Creative Commons Metadata", July 16, 2004:33:48 EDT. Accessed 2005-03-21. <<http://lists.ibiblio.org/pipermail/cc-metadata/2004-July/000421.html>>

Raymond, D.R., and F.W. Tompa. "Markup Reconsidered." *Technical Report 356*. Department of Computer Science, The University of Western Ontario, 1993. Presented at the First International Workshop on the Principles of Document Processing, Washington DC, October 21-23 1992; an earlier version was circulated privately as "Markup Considered Harmful" in the late 1980s.

Renear, A. "The Descriptive/Procedural Distinction is Flawed." *Markup Languages: Theory and Practice* 2.4 (2001): 411-420.

Renear, A., D. Dubin, C. M. Sperberg-McQueen, and C. Huitfeldt. "Towards a Semantics for XML Markup." *Proceedings of the 2002 ACM Symposium on Document Engineering*. Ed. R. Furuta, J.I. Maletic and E. Munson. McLean, VA, November 2002. 119-126.

Renear, A., H.C. Phillippe, P. Lawton, and D. Dubin. "An XML Document Corresponds To Which FRBR Group 1 entity?" *Proceedings of Extreme Markup Languages 2003*. Ed. B.T Usdin and S.R. Newcomb. Montreal, Canada, August 2003.

Sasaki, F. "Combining Markup Semantics and Semantic Markup: A Secret Marriage." *Proceedings of ALLC/ACH 2004*. Goteborg Sweden, 2004. 122-125.

Sperberg-McQueen, C.M. "Text in the Electronic Age: Textual Study and Text Encoding, With Examples from Medieval Texts." *Literary and Linguistic Computing* 6 (1991): 34-46.

Sperberg-McQueen, C.M., A. Renear, and C. Huitfeldt. "Meaning and Interpretation of Markup." *Markup Languages: Theory and Practice* 2.3 (2000): 215-234.

Witt, A. "Meaning and Interpretation of Concurrent Markup." *Proceedings of ALLC/ACH 2002*. Tuebingen, 2002.