On the Characteristics of Scholarly Annotations

Richard Furuta and Eduardo Urbina
TEES Center for the Study of Digital Libraries
Texas A&M University
College Station, TX 77843-3112
{furuta,urbina}@csdl.tamu.edu

ABSTRACT

We report on our observations of annotations for use in scholarly communication, rather than for use as personal artifact. Scholarly annotations reflect uses that predate digital representations and benefit from formalized structure. Scholarly annotations may originate from a broader set of sources than personal annotations, and their association with texts may result from inferences rather than from explicit specifications.

Categories and Subject Descriptors

H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia; I.7.2 [**Document and Text Processing**]: Document Preparation—*Hypertext/hypermedia*

General Terms

Design, Theory

Keywords

Annotations, Digital libraries, Cervantes Project, Electronic Variorum Edition

1. SCHOLARLY ANNOTATIONS

Annotations provide a significant mechanism for communication, with commonly-seen uses that range from the informal and personal to the formal and public. Many recent publications, focusing primarily on personal annotations, illustrate their importance. A few selected examples include discussion of use and form [3], strategies for presentation [6], and influence on readers [5]. Our recent work, requiring more formal annotations in support of scholarly communication, underscores to us that the structure of the annotation is also a factor of significance. Moreover, our experience suggests that annotations optimized for our scholarly use environment benefit from pre-defined structures while the work focused on personal annotations suggests that those applications benefit from discovery of emergent structures.

The observations in this paper arise from our ongoing projects centered around the writings of and about the Spanish author Miguel de Cervantes Saavedra (1547–1616). One of our current activities [2] is the preparation of multiple "Electronic Variorum Editions" (EVE) of Cervantes' best-known work, *Don Quixote de la Mancha*, based on facsimile images obtained from microfilms of the textually significant editions published around Cervantes' time. An EVE contains and interlinks facsimiles, associated textual transcriptions, any number of virtual editions resulting from editors' resolution of variances among the editions, justifications of editorial decisions, and general commentary.

Don Quixote was published in two parts. Our current tasks center on producing an EVE from eight copies of the first printing (the 1605 *princeps*) of the first part and a separate EVE from the princeps along with five other significant early printings. The resulting EVEs will be significant for use in the study of *Don Quixote*—for example, only 18 copies of the princeps are known to have survived to modern times and of those, only 12 are accessible. Our collection of eight copies already is more than has been available to Cervantes scholars, and is particularly interesting since the literature suggests that no two of the known copies are the same (e.g., because of changes made in press) [1].

There are two main uses of annotations in the EVE. The first use is in conjunction with an editor's evaluations of differences among editions. Here the editor must determine the importance of the differences, must select among the alternatives (or propose a new alternative), and must provide justification for the selection. The second use is in relating the text to the broader environment—for example to cultural aspects, to expanded writings, etc.

Also mandating formality in our application's annotation structure is the preexisting scholarly culture that distinguishes fairly precisely among annotation uses. As illustration, a 1998 critical edition of *Don Quixote*, issued in paper and CD-ROM format, directed by Francisco Rico and edited by an extensive team of over 100 collaborators, uses symbols to categorize the identified cause of variants, and separates other annotations into four levels (footnotes, complementary, critical, and readings) with notes at one level containing cross references to notes at other levels using a series of markings. As a further example of the importance of formal structure in scholarly annotations, the TEI specification [4] reflects a distinction between notes and corrections, each defining its own set of attributes to encode the metadata that defines the entries' structure/characteristics.

We now turn to some of our observations about the structure of annotations in our application domain, focusing particularly on the characteristics that seem to differ from annotations intended for personal use.

HT'02, June 11-15, 2002, College Park, Maryland, USA.

Scholarly annotations require strongly-typed metadata. In conjunction with an editor's choices among variants, a justification must be provided stating the decision basis. Our goal is to make editor's decisions open to the reader; to allow the reader to evaluate the editor's choices by providing access to the source materials available to the editor, by providing visualizations of the decisions made by the editor, and by permitting comparison of multiple editors' choices. In support of this, therefore, the metadata accompanying the justification also includes a classification of the variant and an indication of its importance. Furthermore, in keeping with scholarly conventions relating to authority, the identity of the editor must be associated with the justification, since multiple editors may have created virtual editions. And in order to enable readers' visualization and filtering of decisions, classifications and identities must be constrained to be from a previously-defined set of values.

The metadata types required vary with different classes of annotations. As our project developed, we realized a need for a second form of general annotation. Some elements of the metadata are similar in both forms of annotation (e.g., the editor's identity). Others, however differ; for example the indication of importance is not relevant to the general annotation, but instead a categorization of purpose (e.g., historical, geographical, cultural, etc.) is useful.

Scholarly annotation metadata items can be interrelated, for example in a hierarchy. Furthermore, not all traversals of the hierarchy are meaningful. For example, in an early prototype of our reader's interface, we allowed the reader to choose whether or not to show emendations based first on importance level and secondly on editor identity. This conflicted with common practice, since generally an editor's decisions are made in concert, not independently for each importance category. The current design inverts the choices, with the editor's selection being the primary filter.

Attribution of a scholarly annotation is not necessarily to the editor who enters the annotation. We find that in some cases, the editor desires to carry forward an observation or justification made by an earlier scholar in a well-respected study. Consequently, our annotation records carry a separate field for "reference" to signify properly the authority for the annotation. We note further that such annotations may be associated with materials not included in the EVE.

Multiple forms of the annotation's text may be required. Scholarly annotations may have diverse readers, and consequently the annotation's presentation may be required to reflect that diversity. In our application, there is no "natural" language in which annotations should be expressed. Consequently, at a minimum, we must support alternate presentations of the annotation text in English and in Spanish. We note that in addition to scholarly uses, similar requirements may arise for regulatory reasons; for example in supporting the six official languages of the United Nations or the European Union's eleven official languages.

An annotation may be associated with multiple anchors, which may be inferred rather than explicitly specified. An EVE creates an implicit association among the editions that it contains, relating together the corresponding parts of the included texts. Consequently, an annotation anchored to one of the editions in an EVE can potentially be considered to be anchored to others through the implicit associations. Moreover, multiple EVEs with common constituent editions suggest definition of a form of "closure" in which annotations anchored within one of the EVEs may be inferred to be extended to editions within another EVE, with the specific mapping defined through the intermediary of the commonly-contained constituent.

A practical problem to which this can be applied is an editor's desire to reuse some of the annotations entered in one EVE (e.g., the princeps EVE; see above) in another (e.g., the multiple edition EVE; also see above). However, the same editor is unlikely to have a desire to reuse annotations discussing the specific relationships among editions in an EVE in the context of other EVEs; indeed the annotation may not make sense in the different context.

In our current model, annotations can be anchored in multiple contexts. All annotations are anchored in the context of the EVE in which they are defined. Additionally, the annotation can be associated, as well, with zero or more of the editions contained in the EVE. When any of these editions is used in another EVE, the editor can specify that the associated annotations are to be anchored in the context of the new EVE as well.

In conclusion, we note that support for scholarly annotations seems to require attention to issues that are not as critical when supporting personal annotations. Perhaps a reason for this is that scholarly annotations are intended to be public and archival, and consequently the role of the annotation in supporting communication is explicit and primary in scholarly annotations. An implication of this observation is that the design of systems to support scholarly annotation will not achieve their full potential if they simply adopt the parameters of systems that have been designed to support personal annotation.

2. ACKNOWLEDGMENTS

Furuta also is affiliated with the Department of Computer Science at Texas A&M University and Urbina also is affiliated with the Department of Modern and Classical Languages.

This material is based upon work supported by the National Science Foundation under Grant No. IIS-0081420. Support for this work was provided by the Interdisciplinary Research Initiative Program, administered by the Office of the Vice President for Research, Texas A&M University. Further information about the Cervantes Project is at http://www.csdl.tamu.edu/cervantes/.

3. REFERENCES

[1] J. M. Casasayas. *Ensayo de una guía de bibliografia Cervantina*. Tomo V. Mallorca, 1995.

[2] R. Furuta, S. S. Kalasapur, R. Kochumman, E. Urbina, and R. Vivancos-Pérez. The Cervantes Project: Steps to a customizable and interlinked on-line electronic variorum edition supporting scholarship. In *Research and Advanced Technology for Digital Libraries: 5th European Conference, ECDL 2001*, pages 71–82. Springer, 2001.

[3] C. C. Marshall, M. N. Price, G. Golovchinsky, and B. N. Schilit. Designing e-books for legal research. In *Proceedings of the first ACM/IEEE-CS joint conference on Digital Libraries*, pages 41–48, 2001.

[4] C. M. Sperberg-McQueen and L. Burnard. TEI P4: Guidelines for electronic text encoding and interchange, 2001. http://www.tei-c.org/P4X.

[5] J. L. Wolfe. Effects of annotations on student readers and writers. In *Proceedings of the fifth ACM conference on Digital Libraries*, pages 19–26, 2000.

[6] P. T. Zellweger, N. O. Bouvin, H. Jehøj, and J. D. Mackinlay. Fluid annotations in an open world. In *Proceedings of the twelfth ACM conference on Hypertext and Hypermedia*, pages 9–18, 2001.