Name of Project: Interface Development for Static Multimedia Documents
Institution: Duke University
Project Director: Matt Cohen
Award Amount: $29,857
To support: Building a prototype digital interface for static, mixed-media texts, such as documents containing handwritten annotations on printed text or on images.
Project Website: http://asmodeus.ws/cohenlab/annotations.htm
Keywords: Annotation, marginalia, markup, image annotation, multimedia
Grant/Period: NEH Digital Humanities Start-Up Grant, 09/01/07 - 12/31/08

White Paper: Interface Development for Static Multimedia Documents

By Matt Cohen, Duke University; Erica Fretwell, Duke University; Kevin Webb, Tackle Design

Our project was to build a prototype digital interface for static, mixed-media texts, such as documents containing handwritten annotations on printed text or on images. Using a small set of extraordinarily diverse and sophisticated documents by the poet Walt Whitman as a testing ground, we built software to facilitate the transcription and markup of these texts and to articulate those marked-up transcriptions to digital images so that the locations of manuscript words can be revealed to online readers. We also created prototype interfaces for returning search results for handwritten documents using digital images of original documents, to catalyze user adoption of documents containing handwriting and other graphic forms of representation such as print or images.

In the field of scholarly editing, there are few projects that gather and display an author's manuscript marginalia. Annotations made by a writer in the margins of printed texts or images are crucial sources for analysis in literary, philosophical, and historical study because they are rare evidence of the dialogic interactions between a reader and his or her influences that shape how meaning is made. Marginalia also demonstrate the range of such influences, which often reach far beyond the genres in which the annotator worked. Indeed, they challenge scholars to rethink how "literary influence" itself is defined in the mainstream of literary criticism and history. While a small number of important scholarly studies in literature, history, and bibliography have emerged from the archival study of marginalia, such study has not penetrated the methodologies of the humanities because scholarly editions are available for only a few writers.

Print editions, even simple transcriptions, of documents featuring combinations of printed text, marginalia, and images, are expensive; this cost partly explains the paucity of annotation editions. But even among digital archives such projects are rare (and often use non-open-source software for document delivery). The powerful searching and display capabilities of a digital platform would make such editions far more accessible. But digital presentation of annotated documents is difficult to effect in practice. Annotations depend on more than their linguistic content to make meaning; it makes a difference *where* on the page a comment is made. The significance of spatiality in

relating layers of information pushes current markup approaches to, and perhaps beyond, their descriptive limits. Marginalia, in fact, open a door onto a much larger problem: despite leaps in the automated transcription of printed text, handwritten texts--which include vast numbers of key cultural historical documents--continue to be underrepresented in free online digital resources.

Our project took some first steps towards creating an efficient, open-source tool for transcribing and capturing the locations of handwritten and printed words and pictures *in relationship to* digital images of source documents. The work proceeded in three phases. First, we consulted with Brett Barney, an expert in TEI-based XML markup, to determine a working vocabulary and accompanying tag set for designating different kinds of annotations and document morphologies. Using the resulting strategy, we transcribed and tagged 10 sample documents. In the second phase, software designer Kevin Webb built two prototype interfaces: one for capturing relations between a bitmap image and text markup of that image, and one for searching and browsing documents tagged using this strategy. In the final phase, Brian Pytlik Zillig of the University of Nebraska created methods for web viewing of textual renditions of the XML markup in conjunction with the image-based interface.

Phase One: Markup

The question of how to tag marginal annotations had, in the months leading up to our grant, seen some discussion on the TEI listserv. First, it was clear that the issue involved both TEI markup and stylesheet handling, since the para-hierarchical nature of annotations meant that how a comment was rendered on the screen would impact the degree to which it expressed the intellectual relations in the source text. Second, it was clear that different projects had different solutions to this, of varying complexities. None of the projects posting were, as we are, dealing with three-dimensional documents, and many were not dealing with multimedia documents (usually manuscript-only or print-only ones).

We had hoped to offer an all-purpose solution for this heterogeneity, but we ended up thinking that a variety of approaches is a good thing. Some annotations stand in a comparatively uncomplicated relationship to document hierarchies, and existing approaches handle these well. With the graphical interface design, however (discussed below), we made decisions that we believe will function well for many schemas that handle such documents. We committed to P5 early on, and attempted first to use the <facsimile> element group to relate the spatial specificity of annotations or multiple media presences to TEI-encoded transcription. We found ourselves needing some way to indicate <zone>s that crossed two <surface>s, however--a common feature of three-dimensional documents. After creating a milestone element, <vb> (for “vertical break”) as a kluge for handling this problem, and after consulting with the graphic interface designer, we instead decided to table <facsimile> and extend the existing Whitman Archive schema such that it could handle pages -- surfaces -- as intellectually significant structural units.

Our current implementation utilizes a <page> and model that allows content to be encoded hierarchically. In this system document entities are represented as

s with attributes used to differentiate the various physical and conceptual structures that they encode. Physical ordering is expressed through the hierarchy of the elements with the inner-most element within a stack of elements having the highest or top-most position on the page.

Occasionally, intellectual units will cross surfaces intact, even as the surfaces impact their interpretation or potential field of reference. In these cases, we chose to join s using the @prev and @next attributes and an ID system. This seems clunky, but we ended up liking the way it calls attention to hierarchies that cross each other. We also ended up needing to create rules for establishing the fundamental @type declaration for the document: While we might be choosing a piece of newspaper clipping for encoding because Whitman wrote on it, declaring this a Whitman manuscript seemed wrong when the document contained predominantly text authored by someone else and in the print medium. We ended up thinking about the type declaration in terms of "guests" and "hosts": The "host" determines the @type of the document as a whole, while the "guests" (Whitman annotating a newspaper article in a minor way; or, alternatively, newspaper clippings pasted to a predominantly Whitman-authored manuscript) were indicated by nested elements such as <note>, <hand>, and <text>. Some of these calls are hard to make, but our guiding thought was to make the resulting encodings as portable to other contexts as possible.

These decisions were largely made during a week-long visit to Duke's campus by Brett Barney, who, together with Erica Fretwell, then trained the undergraduate transcriber/encoders in the markup approach. By mid-spring, sample documents were marked up and ready for the software designer and high-quality scans of the originals had been created by Special Collections at Duke University Libraries.

Phase Two: Graphic Interface Design

This phase involved the creation, by Kevin Webb, of two interfaces, both web-based. One helps automate the capture of spatial coordinates in an image file for insertion into an XML-encoded transcription. The other allows for browsing and word-searching of the linked images and encoded transcriptions.

In our original proposal we anticipated using JavaScript as a platform for the graphical interfaces. Just before the grant began, however, Adobe made the Flash / Flex platform available for developers under an open-source license. We liked the superior look, feel and speed with which this solution worked. The coordinate-capture interface allows an encoder to insert a transcription of a document, draw layers on a digital image or images, and then quickly arrange and designate individual word-spaces. Each coordinate space becomes a div, and the tool is extensible to perform more elaborate relating of element-spaces as individual projects desire. For this prototype, we focused principally on easing coordinate capture for individual words and on designing an interface that could handle documents with multiple layers attached to single supports.

The encoding tool maintains a direct association between the image surfaces and the markup expressed by the user, either as the encoding of structural elements or the

transcription of document content. The tool also understands the basic structure of the transcribed text (line and word breaks, paragraphs, etc.), allowing it to build a spatial model of the word positions on the page with minimal user input. This internal model can be saved to and restored from an XML document, allowing the user to work visually while at the same time seeing the encoded output.

The search and browsing interface uses the same basic framework as the capture interface. Unfortunately, our implementation of this part of the project has been slowed by changes in the technological infrastructure and staffing at the *Whitman Archive*'s home institution. Running prototypes of both interfaces, with instructional video walk-throughs describing them, may be found at the Cohen Lab website through the URL listed in the "Resources" section below.

Phase Three: Text-based Interface Design

The generation of a text-based interface, through which researchers can quickly read documents and copy them into other applications for use in publications, is still in progress. Here our scheduling projections were perhaps overly ambitious; until the full capabilities of the graphical interface were in place and the schema extensions finalized, it was difficult to start work on stylesheet rendering without the potential of having to do the work twice. A basic stylesheet rendering of the text content of the XML files, paired with the source images, has been generated. But we have begun work on an implementation that will pass the coordinate information to ImageMagick to allow for a more dynamic visual presentation of the text transcription--one that draws on the spatial relations of the text-chunks in the original document. Through this interface users will be able to "flip" through different components of the document, displaying combinations of text that more closely approximate the reading options of the originals.

Next Steps

The next steps involve testing, schema extension, documentation, and implementation at the *Whitman Archive*. Once the stylesheet is developed, student encoders will be asked to test-drive the coordinate-capture and browsing interfaces to assess usability and learning curve. Fretwell and Cohen have been testing the interface at its different stages of development, and we anticipate few major changes to the interface during testing. We will also contact our partners at the University of London and at Duke University and make the software available for their use. The *Whitman Archive*'s DTD will need to be extended in order to handle the coordinate markup approach we've developed. While this is happening, necessary documentation for installation and adaptation of the interfaces will be generated by the developers (some of which has already been accomplished in Kevin Webb's walk-through video). Draft guidelines to our XML markup approach have been created, and these will be wikified at the *Whitman Archive*, expanded, and made available to the public. We anticipate these steps being finished by the end of May 2009.

Implementation at the *Whitman Archive*, and testing on a larger body of static multimedia documents, will take a bit longer, as it must be coordinated with the University of Nebraska's larger strategy for digital text searching infrastructure. Our aspiration,

however, is to use the prototype as the interface for building and accessing an edition of Whitman's annotations on printed texts and images. The interface may in fact be useful to other major projects at the archive, including the poetry manuscripts and periodical poems sections.

Publicity and Reflections

Since one of our goals all along has been to create a tool that could be broadly adaptable, publicizing our work will be an important part of our "next steps." Fretwell and Cohen will be presenting on the project at the Society for Textual Scholarship conference in March 2009; Fretwell will be presenting a more theoretical talk based on our work for the Digital Humanities conference in June. Fretwell will also be writing an article intended for refereed publication about TEI P5's approach to facsimiles and the implications of our alternative approach to intellectual structures that are simultaneously built on textual and visual logics, particularly as these relate to the temporality of reading. Once testing has been completed, the project will be shared on the TEI list and the software made available for download from the *Walt Whitman Archive*'s server via the *Whitman Archive*, the Cohen Lab website, and other venues to be negotiated.

In the process of developing the project, other future development possibilities emerged. First, Terry Catapano (Columbia Univ.) suggested we look at projects working with layered approaches to TEI. Such an implementation would address many of the fundamental problems we encountered with encoding physical and conceptual layers within the same document. In a layered markup physical and conceptual encoding could be interleaved without one interrupting or impacting the other. This eliminates the need for structures like the @next / @prev attributes in our current schema and opens up many possibilities for expressiveness not currently feasible with traditional XML encoding methodologies. Aside from important work at the University of Kentucky (the ARCHway Project) and a few small working groups, there does not appear to be much active development going on in this area, but it is the direction we feel we need to go in terms of capturing multiauthored, multihierarchical, three-dimensional texts. Too, Doug Reside's work with the *Our Americas Archive Partnership* offers potential for making our coordinate capture and transcription system more robust.

In addition to the reflections above on the challenges of representing text that is non-hierarchical by virtue of its spatial, rather than strictly text-contextual orientation, we found that we had bitten off a bit too much in this project, given its funding levels and time-frame. To coordinate the various elements of the project, some salary replacement for the PI would likely have streamlined progress, but even in that event, we suspect ourselves to have underestimated the significant time documentation and information-sharing required. Another possibility would have been to focus on markup and graphic interface development alone, though this would have postponed our learning important lessons about potential solutions to the challenge of representing complex multifaceted documents on the web.

Resources

Documentation of the project, conference papers, and access to the prototype interfaces are available at the Cohen Lab website:

http://www.asmodeus.ws/cohenlab/annotations.htm

Downloadable versions of the interface software and stylesheet code, as well as software documentation, will be made available at this site when they have been debugged.