
Play and Code in Humanist Research

Vika Zafrin (zafrin@brown.edu)
Brown University

1. VHL: An Introduction

The Virtual Humanities Lab is a new humanities computing project at Brown University. We focus on two areas of research. First, we are designing and building a web-based engine for the presentation of semantically encoded primary texts, and for further annotation of these texts by invited scholars. Together with this engine we will be publishing several annotated texts. This engine is complemented by a weblog and by a discussion forum; both of these invite input from anyone interested.

We are in the process of semantically encoding, annotating, and publishing online three early modern Italian texts. The first and largest, Giovanni Boccaccio's *Esposizioni sulla Comedia di Dante* is the vernacular, originally oral text of Boccaccio's unfinished lecture series on Dante's *Commedia*. Giovanni Villani's *Nuova Cronica* (of which we are publishing a part) is an extensive account of Florentine history up to 1348. It is written in lively Italian, valuable not only for its record of events but also for its historiographical methods and political commentary. Finally, Giovanni Pico della Mirandola's *Conclusiones Nongentae* is an aphoristic Humanist text currently being developed as part of the Pico Project (<http://www.brown.edu/pico/>). All three of these will be presented electronically for the first time.

The sheer amount of information present in these texts — as well as their size, relative obscurity and general importance for the humanities — lend themselves to semantic encoding, collaborative annotation and electronic dissemination.

The number and variety of electronic tools being built for humanities research is ever-increasing. So is the learning curve for taking full advantage of these tools. Since semantic markup plays an increasingly important role in electronic humanities scholarship, having an idea of what it looks like and — broadly — how it functions seems to be an advantage for academics in the humanities.

Scholars who have never practiced semantic encoding of texts as a research tool, or performed complicated searches on semantically encoded texts, may find themselves reluctant to

spend time learning an unfamiliar way of working, even if the result of such learning may prove useful to them. Such researchers are a large part of our intended audience, and we are putting significant effort into writing clear, concise documentation and hands-on tutorials. The documentation is aimed at academics relatively new to humanities computing and, as such, will include a brief overview of the principles of semantic encoding as well as a guided tour of the VHL toolset. Our goal is to make these supplementary materials enjoyable and concise: we want scholars to receive just enough technical information to enable them to play with their texts.

2. Playing and Modeling

Michael Mahoney says that a sufficiently complex idea for a piece of machinery cannot be described; the thing must be made or modeled. In order to be understood, complex texts should also be modeled.

A usable model of a text need not be comprehensive, but may rather address one or more specific issues. VHL researchers read the texts and use semantic encoding to arrange their parts (linguistic entities, recurrent themes and imagery, rhetorical devices etc.) in sets of metadata. These sets may overlap and intersect, and function as scholarly arguments. Encoding once does not preclude a division of the same text into a *different* set of parts, with another purpose or from another angle, or in response to an argument made through previous encoding. Each variant model contributes to a deeper understanding of the text at hand.

2.1 Collaboration

At last year's joint conference, Siemens et al. reported: "In terms of mark-up, respondents appear to be a bipolar group with half expecting to acquire text with no mark-up and half with rich XML." This no-middle-ground report seems to imply that once a user of electronic humanities resources is at all familiar with semantic encoding, rich markup becomes preferable to weak markup. Marking up large texts and corpora, common units of literary study, is a challenge both in terms of resources and required expertise. Such work calls for collaboration. VHL's toolset for presenting and working with primary texts (in development) provides several ways to contribute. A complex annotation engine and an opportunity to view the encoding behind any given segment of text are in place. In development is a tool for suggesting corrections to our encoding (intended to replace it), or submitting variations on it (intended to be viewed as alternate encodings of the same text).

While providing increased potential for new forms of communication, this toolset does not force scholars to change their preference for working mostly in solitude: Siemens et al.

do warn us that most of the humanists they surveyed "do not [currently] see the need for collaborating with other scholars."

2.2 Atomic Approach to Research

Freehand semantic encoding allows us to construct our own set of elements, based on prior knowledge of sources both primary and secondary, modifiable at will. Eventually this set must be regularized, perhaps later transcribed into a standardized form. But in the beginning stages such constraint would be detrimental, limiting the scope of analysis at the outset.

So we have begun to model without these constraints, permitting ourselves the spontaneity of a ludic approach. In doing this, we adapt Edward Hall's 1976 objective in examining culture — "look at the way things are actually put together" (13) — to text analysis. The encoding structure emerges bit by bit out of the primary source itself, which frees the researcher's critical eye to note interesting aspects of the text that might have eluded a pre-existing DTD.

Combining such an atomic approach to gathering research results with a web-based presentation implies a lot of flexibility for participating scholars: work may be done in smaller segments by individuals who live far apart. Here lies a strong driving force behind our work: similarly to already-successful electronic means of communication (email, weblogs, discussion lists), VHL allows small information packets to be published and discussed. Being unsuitable for the essay format because of their seemingly incomplete, fragmentary nature, these bits of information might not otherwise be expressed at all. Reducing the minimum size of a contribution to the knowledge base from an article to a paragraph or sentence, provided a review process is still employed, increases the net amount of useful knowledge available for discussion. We hope that it will actively encourage researchers to branch out and participate in more conversations, perhaps creating a distributed version of the editing process.

Stripping critical expression down to the essentials as expressed through semantic tagging will either highlight or address (or perhaps both) the difficulty Willard McCarty sees humanists having "with any intellectual culture whose cognitive activity is expressed in things rather than in words" (168). Thinking about a text by encoding criticism directly into it bridges the gap between the two, allowing multi-media corpora (literature, sculpture, films, drawings) to be encoded within the same electronic framework. Emphasis is shifted from the prose that delivers ideas (which consumes time and energy and often dilutes the argument) to precision in presenting the argument itself.

2.3 Humanists and Code

The encoding process requires considerable resources; writing up separate documentation is a significant enough amount of

additional work that it isn't often done well. For humanist academics, it is absolutely necessary to be able to look at semantic encoding and more or less understand it.

Mahoney, and Henry Ford before him, are right: the masses are not mechanics. Yet, these days a certain amount of common knowledge about how machines work is necessary. Since semantically encoded electronic texts will only multiply as time goes on, humanists must know what code is and understand how it works. Knowledge of the underlying principles of encoding is not yet widespread, and VHL has taken it as a goal to present these principles in such a way that they become tacit knowledge for the humanist. We are making all of our XML code transparent -- any unit of text is viewable with all its code, and the XML itself is easily human-readable and well documented. Thus code remains an argument meant to be discussed and challenged as necessary, not an implicit, uncontested premise.

Learning to read code may require non-trivial effort, but carries with it an important additional benefit: it opens the door to a format of academic expression markedly different from the essay. Both have their uses, but mastering the basics of semantic encoding is a learnable and improvable skill that is likely to become tacit knowledge more readily than the much more difficult natural-language rhetorical approach of essay writing. Sentence structure, flow and finding the right word are essential to the humanist; but encoding makes it easier to learn and practice critical, in-depth analysis of texts.

3. Summary

Putting small bits of information together and hoping that a larger picture will emerge is arguably risky. There is no guarantee that the results will be interesting or useful. That said, this risk is inherent in all academic discussion, and recent experience indicates a movement (back?) toward tinkering with primary sources directly. Stephen Ramsay's call to go in "with a hunch borne of our collective musings" (171) encourages play, frightening though it may be to dedicate extremely scarce resources to the endeavor. This is where a playground like VHL shines. It is a tool for collaborating, community building and education that does not require a significant commitment of finances or time from its participants. In fact, for it to function, there need only be interest in the subject matter, and the willingness to record a single thought.

Bibliography

Decameron Web: A Growing Multimedia Archive of Materials Dedicated to Boccaccio's Masterpiece. Accessed 2005-03-04. <<http://www.brown.edu/decameron/>>

Hall, Edward. *Beyond Culture*. Garden City, NY: Anchor Press, 1976.

Mahoney, Michael. "Keeping In Touch With the World." Commencement Address delivered at Brevard College. 16 May 1998. Accessed 2005-03-04. <<http://www.princeton.edu/~mike/brevard.html>>

McCarty, Willard. "As It Almost Was: Historiography of Recent Things." *Literary and Linguistic Computing* 19.2 (2004): 161-80.

Pico Project. Accessed 2005-03-04. <http://www.brown.edu/Departments/Italian_Studies/pico/>

Ramsay, Stephen. "Toward an Algorithmic Criticism." *Literary and Linguistic Computing* 18.2 (2003): 167-174.

Siemens, Ray, Elaine Toms, Stéfan Sinclair, Geoffrey Rockwell, and Lynne Siemens. "The Humanities Scholar in the Twenty-First Century: How Research is Done and What Support is Needed." Paper presented at ALLC/ACH 2004, Gothenburg. 2004.

Virtual Humanities Lab. Accessed 2005-03-04. <http://www.brown.edu/Departments/Italian_Studies/vhl/>