Roland[HT]: A Proposal for the First Annual Nebraska Digital Workshop.
Vika Zafrin

I would be thrilled to participate in the Digital Workshop this September, and to present my dissertation work, *Roland[HT]*. It is a continuation of the work I did for my Master's degree in Italian Studies; its development in the coming academic year will be informed by my work as project director of the Virtual Humanities Lab [1] at Brown University.

Who?

Roland may or may not be a historical character. It is generally accepted that the literary/artistic Roland is the fictional descendant of count Hruodlandus, captain of king Charlemagne's Breton legion, mentioned in historian Einhard's *Vita Karoli Magni*. Three hundred years after Hruodlandus' death in a 778 ambush by the Basques, the *Song of Roland* was written down in France, having undoubtedly been part of the oral tradition long before then. He is now regarded as a national hero of France.

The French Roland has easily identifiable personality traits, such as his infamous pride, which cost him and twenty thousand of Charlemagne's best men their lives, as well as his physical prowess. There are also artifacts, such as his enchanted sword and the horn that, when Roland sounds it, is heard for miles. Together, they have allowed me to trace Roland's presence in many different guises throughout most of Europe, the United States, and Brazil. His great popularity as a character is remarkable for two reasons. It has not waned for over a thousand years (the most recent item in the corpus dates from 1995). What's more, the character is not merely borrowed; he is mercurial. Roland's identity has been appropriated by many cultures in different times, and adapted to address the moral and socio-political issues of particular *where*s and *when*s. [2] Twelfth-century Germans made him a martyr; the medieval Welsh, a single-minded, Klingon-like warrior. During the Renaissance, Italians shifted his birthplace to Italy and turned him into a madly devoted, clinging lover. He has been melodramatic comic relief in France, a puppet in Sicily, a comic book hero somewhere between Santa Barbara and Brazil. Roland is a cultural hypertext.

My working thesis is that there exists a Roland corpus, consisting of works in such diverse media as literature, live and puppet theater, opera, sculpture and stone carvings, painting, film, contemporary music, and comics. This corpus has never been considered as a whole. Essays and books are written on individual works about the character, on his role in one particular culture, and even on the migration of his legend to one country from another – usually France. Roland's true polyphony has gone largely undiscussed.

An examination of themes recurrent in it offers insight into the evolution of art and the pace of intercultural transmission. This insight is relevant to the study of literature at the dawn of Western secular writing; it also sheds light on our own literary present, in an age of widely accessible electronic writing, when the processes of composition and transmission are changing again.

A Heuretic Dissertation

The approval by Brown's Graduate council of a special-studies doctoral program in humanities computing presents an opportunity for this experimental dissertation, half of it in prose and the other half in semantic code. In order to show that the body of texts [3] under consideration is indeed a corpus, I am using XML to encode themes, imagery and characters recurrent in it. At the moment, all of the primary texts in use are word-based. I will be adding digital photographs of stone carvings, sculptures and other physical objects depicting Roland, including some sound and video files, and a video capture of the 1995 computer game *The Madness of Roland,* as rights management allows. The recurrent threads are many, and they can only be discovered through close examination of the primary sources. Because of this, I am encoding "freehand," without using a pre-defined Document Type Defnition. This allows me to devote all of my attention to the text directly in front of me and what it actually contains, as opposed to what I imagine it should contain. In this process patterns (an InfoSet) have emerged, and with them expectations. Nevertheless, the absence of a DTD is freeing. Whenever a new pattern seems to be emerging, it can be tagged

immediately using a new element. If later it proves not to be a sufficiently consistent pattern, it can be unencoded just as easily. The process resembles solving a Rubik's cube or assembling a jigsaw puzzle, in that it is thinking by doing. Stephen Ramsay advocated for such an approach at a recent ACH-ALLC conference. "It is, of course, possible," he acknowledged during a panel on reconceiving text analysis, "to go to a literary text armed with a hypothesis, but we do better to go to it with a hunch borne of our collective musings – a sneaking suspicion that looking at it this way will turn up something interesting."

Encoding as a mode of active thinking is a heuretic approach to literary-artistic analysis. In his 1994 *Heuretics: The Logic of Invention*, Gregory Ulmer states that theory is assimilated into the humanities in two principal ways: by critical interpretation and by artistic experiment. Heuretics is the latter – it is a heuristic approach to theory, a reading process that, instead of attempting to theorize "what might be the meaning of an existing work," guides "a generative experiment: Based on a given theory, how might another text be composed?" (5) By marking up the semantic connections as I go along, I propose another text, one composed of fragments from here and there and (at least as regards word-based texts) containing no previously created work in its entirety. Such a text may only be a preamble to a *deep* understanding of any particular work, but it does illustrate the intercultural *breadth* of the Roland corpus and allow the reader to explore the corpus according to her interests.

Considering the number and diversity of existing works about Roland, it seems impractical to perform this work efficiently in a paper-bound book. The book's content would be redundant, quotations long and unwieldy, explication of connections among them too wordy. It would be difficult for the reader to construct a coherent (if multiplicitous) whole; and most importantly, it would be practically impossible to see larger thematic patterns which emerge as individual correlations are encoded. The electronic medium offers more flexibility in terms of multiple presentation possibilities, cheaper production and more efficient dissemintation.

Collaboration Now And In the Future

From the beginning it has been clear that Roland's reach far exceeds the possibilities afforded by a doctoral dissertation. I do not propose to catalog the entire corpus; this task will require a professional lifetime's work. More importantly, it will require collaboration with a well-established community of Roland scholars.

During my tenure directing the two-year Virtual Humanities Lab project, I have found collaboration issues to be at the heart of almost all the tasks we have undertaken. Two of our encoders worked on a large text together, but are based at two different institutions. The latter case necessitated creativity in communication. Neither of the two scholars had done semantic encoding before; both were somewhere between beginning and intermediate computer users. We had some initial sessions wherein they were trained on encoding and using relevant software (oXygen for XML work, a subversion client for the versioning system). After that, they were – at least physically – on their own.

This challenged us on several fronts. Delays in communication – telephone was not always an option, and email sometimes took three or four days to respond to – meant that it was difficult for the encoders to know exactly when they would be able to work. Since they depended on my help when something went wrong with checking in documents or version conflicts, telephone support for Windows machines was about as easy to provide as you might imagine, given that these were the only two Windows computers being used on the project. Furthermore, both scholars had many other obligations, and most of them were in the "real world" – much more tangible than the mostly electronic VHL project. Of course (and rightly), the other tasks took precedence, diverting their efforts further away from the work that just didn't seem as *real*.

We solved the problems described above by drastically increasing our real-time and face-to-face communication. Every Friday the lead of that sub-project drove two hours each way so that we could spend time dealing with larger conceptual difficulties and specific encoding questions. We made a commitment to be

more reliably available over the phone, which became easier once we'd had opportunities for show-don't-tell sessions and had worked out a larger common vocabulary specific to the text. We also made use of instant messaging, giving the encoders greater confidence in their workflow.

All of this, along with recent discussions on mailing lists and in informal gatherings, has led me to conclude that, if *RolandHT* is to become a truly collaborative project, it will need to accomodate many levels of technical expertise. This will likely take the form of clear and concise documentation that aims not just to retain a passing glance but to attract scholars who may have been suspicious of new technologies, or haven't had time to learn new ways of working. After all, formal collaboration is quite new to contemporary humanities work.

That said, I believe that we (the entire planet, albeit at vastly different rates) are on an inexorable path to a Star-Trekkian world, where most information is digitized and you may ask The Computer to find the best layman's description of cyanobacteria and actually get the answer you seek. Our approach to that fantastic world is slow enough, however, that everyone – the humanities academy included – will have a chance to catch up. *RolandHT*'s quite traditional subject matter, which one can study with a basic understanding but no *practice* of semantic encoding, may be a non-threatening gateway to patterns of thought as yet unfamiliar to many humanists.

The encoding behind the Roland project will, of course, be made available for scrutiny by the relevant communities (Roland scholars, comparatists, digital humanists, as well as any other interested party within or outside of academe). Aside from critical input from these sources, the project will eventually require two types of collaboration. Foremost, of course, is gathering additional materials. Contributors may simply point me to an interesting piece of art, indicate connections within it to the existing corpus, or actually encode it in accord with the existing system.

The other type of collaboration I will seek is aesthetic design and programming. While I am reasonably competent in several web-relevant languages and tasks used by digital humanists, there are people far more talented when it comes to visualizations, programming for increased processing efficiency, web design and so on. All major digital humanities projects I have seen so far have been built by highly interdisciplinary groups; my hope is that someday *RolandHT* will grow up to join the ranks of well-known scholarly web resources.

Getting Technical

The current interface to *Roland^{HT}* uses JavaScript, XML, XSLT and CSS. Future editions will use an XML database such as eXist and/or an object database to allow for XQueries to be run against the full corpus and provide a REST/RPC/SOAP API for integration with other humanities computing tools. More information on the project, along with links to some articles and to the latest *RolandHT* interface prototype (a collaboration with Ethan Fremen), may be found at <http://wordsend.org/rht/>.

[1] http://brown.edu/Departments/Italian_Studies/vhl/
[2] I refer here to the formulation of parallel worlds in Stephen King's seven-book series *The Dark Tower* (bibl. info), the protagonist of which is a gunslinger named Roland. It is as yet unclear to me whether this Roland will be considered part of the corpus. However, the premise of the series and several key storytelling elements echo the protagonist of my inquiry as well as the multifaceted nature of the Roland corpus as a whole.
[3] I include all form[at]s of Roland works in the notion of "texts", using the word to mean a cultural artifact when critically examined. Similarly, the term "art" here includes the literary arts, which in turn include creative electronic writing.