Digital Humanities 2009

Conference Abstracts

University of Maryland, College Park
June 22 – 25, 2009

The 21st Joint International Conference of the Association for Literary and Linguistic Computing, and the Association for Computers and Humanities

and

The 2nd Joint International Conference of the Association for Literary and Linguistic Computing, the Association for Computers and Humanities, and the Society for Digital Humanities — Société pour l'étude des médias interactifs
Ask Not What Your Text Can do For You. Ask What You Can do For Your Text (a Dictionary’s perspective)

Carlos Monroy  
Texas A&M University  
cmonroy@csdl.tamu.edu

Richard Furuta  
Texas A&M University  
furuta@csdl.tamu.edu

Filipe Castro  
Texas A&M University  
fvcastro@tamu.edu

1. Introduction

This paper’s title, a modified version of President Kennedy’s well known quote, in a metaphorical sense suggests shifting the role texts play in a collection. This shift is based on the growing number of available tools that can be used to enhance textual analysis. We are investigating the effect of tools as generators of what can be done with the texts. We are not advocating a conceptual change in the role and nature of the texts from the literary or textual studies perspective. Rather we suggest a pragmatic role change. In doing so, we believe that new hypothesis about the content of the texts can be posited, or at least their use can be augmented.

Our motivation is based on our experience in the creation and use of a multilingual glossary of nautical terms for the Nautical Archaeology Digital Library (Monroy et al, 2006). In this context, we have seen two major benefits from the glossary that affect both the scholarly practices and the collection. First, it has enabled collaboration among scholars and researchers geographically scattered. Second, how it has broadened the possibilities in the use and understanding of the textual materials—shipbuilding treatises in our case.

2. Dictionaries

Dictionaries have been used extensively in numerous digital humanities initiatives. The Perseus Project (Crane 2002) provides a good example of incorporating dictionaries in a classics collection. The idea behind a dictionary is very simple: an alphabetical list of words with definitions. Yet it has great potential and usefulness when used simultaneously with the contents of a digital collection. Dictionaries also come in various flavors—bilingual or multilingual, thesaurus, specialized, illustrated, and encyclopedic—to name a few. With the use of information technology and the Internet, it has been possible to expand not only their use, but also the way they are created and edited.

Searching for a term in the on-line dictionary of the Real Academia de la Lengua Española (RAE, 2008), for example, presents users with occurrences of that term in all the digitized dictionaries where it has ever been edited; see Fig. 1. Because all editions of the printed version of the dictionary have been digitized, it is possible to visualize the evolution of the definitions of a given term. Although this electronic version of the dictionary in itself is a great resource, one can imagine what could be accomplished if used in combination with a corpus of texts.

Arachne (Foertsch 2006) is an electronic repository (database) of the German Institute for Archaeology. Because archaeological objects are scattered across the world, Arachne provides multilingual access and thesaurus. The Getty Thesaurus of Geographical Names (Baca 2004) is another good example of an external tool that can be incorporated into existing textual materials, enhancing searching and browsing.

The LEO on-line dictionary (LEO, 2008) was originally launched as a German-English dictionary in 1995. At present it includes German translations into French, Spanish, Chinese, and Italian. Since its beginning, two of the most remarkable accomplishments have been the
integration of a larger and linguistic diverse editorial team. And the creation of new environments for searching, using, and learning. For instance, the current version allows users to join groups and work together to learn the language; it also enables teachers to organize lessons, see Fig. 2.

Fig. 2 A screen shot of a bilingual dictionary—LEO—depicting translations and definitions

3. Motivation
Although the use of tools for textual analysis is not a new concept in digital humanities, their emergence is reshaping not only the use of textual materials, but also the mere notion of the texts themselves. Impacting in the end how scholarly practices are conducted.

Traditionally, textual scholars can be seen as “consumers” of the texts. They analyze, compare, and study their contents, how they relate to each other, and their historical and cultural contexts to name a few. Rockwell (2003) commenting on a well known discussion about two approaches to texts: as a hierarchical objects advanced by Renear, and as a performance proposed by McGann, states:

If we are to take McGann’s public performance of a reading as an analogue for what we wish to achieve with these tools, we have to think not only about how we represent the text but also about the performance of analysis and the tools that are used to perform this analysis with a computer.

In this paper, following Rockwell’s statement, we describe a change in the role of textual scholars, from consumers (users of the texts) into producers (augmenting the texts) with the use of tools. Our observations are based on the function external tools can play in augmenting the use of the texts; how they can be used; and what can be learned. Therefore, our goal is neither to propose a right approach, nor to compare approaches.

We take this approach for two reasons. First, at a recent textual studies conference (CASTA, 2008) a participant asked one of the presenters regarding the numerous tools available to text scholars: “But don’t you think that quite often the problem with ‘tools,’ is precisely that there are too many of them; and we don’t know what to do?” This is an interesting question because—although coming from a literary scholar with strong background and expertise in using technology for textual studies—it shows the marked prominence of the role texts play in digital humanities, or at least in how humanists perceive their role.

The second reason is based in our experience working in the creation of a multilingual glossary of nautical terms. The glossary has allowed the incorporation of a new layer to the original transcriptions. For example it would be possible to search for a given term in one language and retrieve occurrences in the transcriptions in multiple languages. Also categories associated to the terms can be used for retrieving occurrences in various contexts.

Used in the context of the Willa Cather Archive (The Willa Cather Archive, 2008), Evince—a non-invasive text analysis tool that mediates the integration of analytical data with the text—shows how a tool can enhance the textual materials. What is interesting in this case, is the fact that the tool is being used to augment the study of the texts, hence improving what can be learned from them. Discussing the use of Evince, (Jewell et al., 2008) state:

We posit that integration of analytical data with the reading text will create new possibilities for interpretation informed by textual data, as it will eliminate the need to enter a specialized environment.

4. Our Collection of Shipbuilding Treatises
Shipbuilding treatises are ancient technical texts, both printed or manuscript, that describe the conception and construction of ships, establish the required types and properties of the wood and building materials utilized, and sometimes describe the steps to be followed in their construction. Given their characteristics, these texts can be properly considered as ancient technical manuals. Our collection was started with three Portuguese treatises obtained with permission from the Portuguese Academia de Marinha and National Library. At present our collection has grown to eleven copies; and includes ma-
tersials in Portuguese, French, Italian, and Dutch, spanning a period from the late 16th to the early 18th centuries. Additionally, an English book is already digitized and ready to be added.

In terms of naval and seafaring dissemination, shipbuilding treatises are priceless sources for scholars working in ship reconstruction and studying the evolution of shipbuilding techniques. Moreover, the development of underwater archaeology in the last 50 years propitiated the growth of the archaeological data corpus, which can now be tested against the textual evidence pertaining to the conception and construction of these complex machines. Nautical Archaeology students, on the other hand, study ship treatises as part of their curriculum. Finally, for the general public they are a great source of historical and cultural contexts in which seafaring flourished.

5. The Tool—A Multilingual Glossary of Nautical Terms

The need for the creation of our glossary goes back to an English illustrated glossary included as appendix in an underwater archaeology book (Steffy, 1994). Tied to one language and a printed medium, the glossary's limitations were evident. But the most pressing reason was the various languages in which the texts in our collection were written. Further, the glossary is essential because nautical archaeology is a highly specialized domain where technical terms need to be explained in order to understand their meaning and context.

Our model uses term as the atomic element. Each term in turn has an associated matrix where columns correspond to roles and rows to properties. Because we are working on a multilingual glossary, we decided to use properties to map languages, while roles map synonyms and spellings respectively, see Fig 3. Each cell at the intersection of role and language can contain zero, one, or more values separated by the symbol |. This implies that each cell can be represented as a vector of values.

Our approach allows scalability and flexibility. For example, we had to add a new language—Venetian—since it was not originally considered, and was requested by one of the scholars. Adding the new language was a straightforward process. Similarly, adding new roles entails the addition of more columns. In both cases both the architecture and the interface scale easily. From the implementation standpoint, we use a relational database for storing terms, synonyms, spellings, and definitions in multiple languages.

Using Lucene—an open source full-text retrieval software (The Apache Lucene Project, 2008)—we are parsing texts and automatically creating links to the glossary. Fig. 4 depicts a screen shot of the treatises interface showing the image on the left, and the transcription on the right, with linked terms underlined in blue. Although this process might seem a simple one, implementing it turned out more complicated than expected. The two main reasons were multiple-word entries and the limitations on Lucene’s stemmer to handle 17th-century Portuguese.

6. Conclusion

Our Web-based interface has enabled scholars to work remotely in editing the glossary, expanding its contents and attracting other scholars. This collaboration goes beyond merely the editing of materials remotely. It has allowed us to obtain materials from other libraries and also to engage the special collections library at Texas A&M in the acquisition of original materials.

Profiting from the rich illustrations nautical treatises provide and the numerous ship models in our collection, we want, in the near future, to create multilingual illustrated
dictionaries, linking them to the texts. As stated earlier, our goal is not to redefine the role of texts in the humanities. But as our experience with the introduction of the multilingual glossary in NADL indicates, tools are shifting the way texts are perceived.

7. Acknowledgements
This material is based upon work supported by the National Science Foundation under Grant No. IIS-0534314.

8. References


The Willa Cather Archive, University of Nebraska Lincoln http://cather.unl.edu (accessed 1 November 2008)