A Multilingual Approach to Technical Manuscripts: 16th and 17th-century Portuguese Shipbuilding Treatises

Carlos Monroy and Richard Furuta
Center for the Study of Digital Libraries and
Department of Computer Science
Texas A&M University
College Station, TX 77843-3112, USA
(979) 845-3839
{cmonroy,furuta}@csdl.tamu.edu

Filipe Castro
Center for Maritime Archaeology and Conservation
Texas A&M University
105 Anthropology Building
College Station, TX 77843-4352, USA
(979) 845-6694
fvcastro@tamu.edu

ABSTRACT
Shipbuilding treatises are technical manuscripts written in a variety of languages and spanning several centuries that describe the construction of ships. Given their technical content, understanding terms, concepts, and construction sequences is a challenging task. In this paper we describe a scalable approach and a multilingual web-based interface for enabling a group of scholars to edit a glossary of nautical terms in multiple languages.

Categories and Subject Descriptors
H.3.7 **[Information Storage and Retrieval]**: Digital Libraries – *system issues*.

General Terms
Design.

Keywords
Nautical Archaeology, multilingual technical manuscripts.

1. INTRODUCTION
The Nautical Archaeology Digital Library (NADL) [5] is a collaborative effort between the Center for the Study of Digital Libraries and the Center for Maritime Archaeology and Conservation at Texas A&M University. One of NADL's main goals is to assist Nautical Archaeologists in the reconstruction of ships. Ship reconstruction is a task where archaeologists have to rely on two main supporting sources: archaeological evidence recovered from other excavations and shipbuilding treatises. These treatises are technical manuscripts that describe the construction of ships. In this paper we discuss a scalable architecture we have developed that enables scholars to edit a multilingual glossary of nautical terms, a preliminary step in understanding the contents of ancient shipbuilding treatises.

Recently archaeologists have started using a variety of computing technologies to make information much easier to access, manipulate, and analyze. For example, the ETANA-DL initiative [4] assists archaeologists in collecting and recording their data, as well as in disseminating their findings. The Alexandria Archive Institute and the University of Chicago's OCHRE project [2] help in documenting surveys and excavations, providing an XML-based schema called ArchaeoML. In terms of written documents, text encoding—such as TEI or XML—allows structuring and indexing documents. Perrow and Barber [3], propose a method for parsing unstructured textual records of a collection of 11th-century manuscripts. Fekete, et al. [1], show the use of Compus, a visualization tool for analyzing a corpus of XML-encoded 16th century French manuscript letters.

JCDL '07, June 18–23, 2007, Vancouver, British Columbia, Canada.

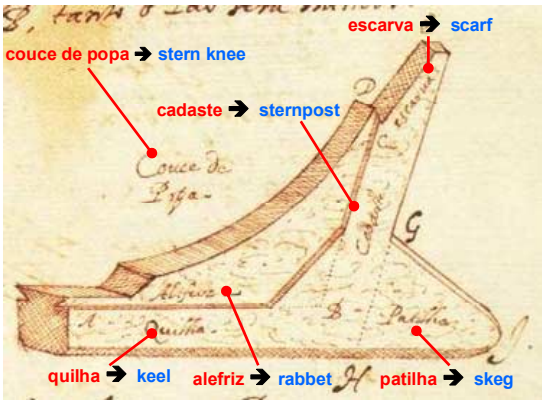

Figure 1. Illustration from a Portuguese treatise showing a composite part of a ship. Original labels are in Portuguese.

2. SHIPBUILDING TREATISES
Shipbuilding treatises are ancient technical manuals that describe properties of the wood and materials used (ship components), as well as the steps to be followed in the construction of a composite object (ship). Figure 1 depicts a composite object in a treatise in which labels and text are written in a different language.

Three of the most significant Portuguese shipbuilding treatises in our collection are: *O Livro da Fabrica das Naus* written in 1580 by Fr. Fernando Oliveira, *Livro Primeiro da Architectura Naval* by Joao Baptista Lavanha, c. 1610, and *Livro de Tracas de Carpinteria* written by Manoel Fernandez (1616). Although their contents varies, in general they describe the characteristics of the wood and properties of other materials used in assembling timbers, as well as geometric descriptions and proportions, and the steps in the construction of vessels. Other features often present in the manuscripts are illustrations, glossaries, curves, designs, as well as geometric algorithms and physics.

3. OUR MULTILINGUAL APPROACH

For archaeologists language makes shipbuilding treatises difficult supporting material for two reasons: a) they are written in different languages, and b) some terms are unknown to scholars. To certain extent our multilingual glossary shares some similarities with an on-line dictionary such as (http://dict.leo.org/) which provides English-German-English term definition, translation, and classification

Our multilingual glossary is based on the concept of an entity with multiple properties and roles. Properties are characteristics, attributes, or features such as languages. Roles on the other hand, describe functions related to a term such as spellings and synonyms. Associating terms with taxa allows the creation of a taxonomy of terms, which in turn will enable temporal, spatial, and functional segmentation of the treatises.

We have implemented a web-browser interface (figure 2) for editing the multilingual glossary, which allows access to scholars dispersed in different geographical locations. The interface is divided in 4 areas: 1) term list, 2) property/role editing area, 3) taxonomy/categories selector area, 4) definitions editing area. In the term list (area labeled 1) each column corresponds to a particular property—languages in this case. Colors indicate the status of each term: a) red, a term without translation, b) blue, a term with translation, and c) green, a term with translations and definition. The term labels can be displayed in any given language. The property/role editing area allows entering information about the properties and roles of a term (area labeled 2). A pop up window allows editing the values (area labeled 6). Multiple values are separated by the character |.

The taxonomy/categories selectors area enables to associate taxa and categories from a multiple-selection list (area labeled 4). The definitions editing area allows entering the definition of terms in multiple languages (area labeled 3). Editors can also preview the information of a particular term (area labeled 5).

4. CONCLUSION

Multilingual documents are source materials that researchers in other disciplines need to access for their scholarly work. Therefore, we expect that our approach can help other digital libraries' repositories with multilingual documents. Our design extends the basic functions of a dictionary by allowing as many languages as necessary. This is an important feature for our collection since treatises, terms, and concepts from diverse naval traditions are constantly added. Our team is currently composed of eighteen archaeologists and naval/maritime experts editing a glossary of about 600 terms in twelve languages.

5. ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IIS-0534314. Special thanks to the Academia de Marinha (Lisbon, Portugal) for granting permission to digitize Portuguese shipbuilding treatises.

6. REFERENCES

[1] Fekete, J, and Dufournaud, N., Compus: Visualization and Analysis of Structured Documents for Understanding Social Life in the 16th Century. IEEE-ACM JCDL 2000, San Antonio, Texas, June 2-7 2000. pp. 47-55.

[2] Kansa, E., A Community Approach to Data Integration: Authorship and Building Meaningful Links Across Diverse Archaeological Data Sets. Geosphere Vol 1, Issue 2, October 2005, pp. 97-109.

[3] Perrow, M., and Barber, D., Tagging of Name Records for Genealogical Data Browsing. Proc of the 6th ACM/IEEE JCDL, Chapel Hill, NC, USA, June 11-15 2006, pp. 316-325.

[4] Ravindranathan, U., Shen,R., Goncalves, M. A., Fan, W., A. Fox, E. and Flanagan, J. W. ETANA-DL: A Digital Library for Integrated Handling of Heterogeneous Archaeological Data. ACM-IEEE JCDL 2004, Tucson, AZ, June 7-11, 2004.

[5] The Nautical Archaeology Digital Library (NADL) http://nadl.tamu.edu accessed on January 2007.

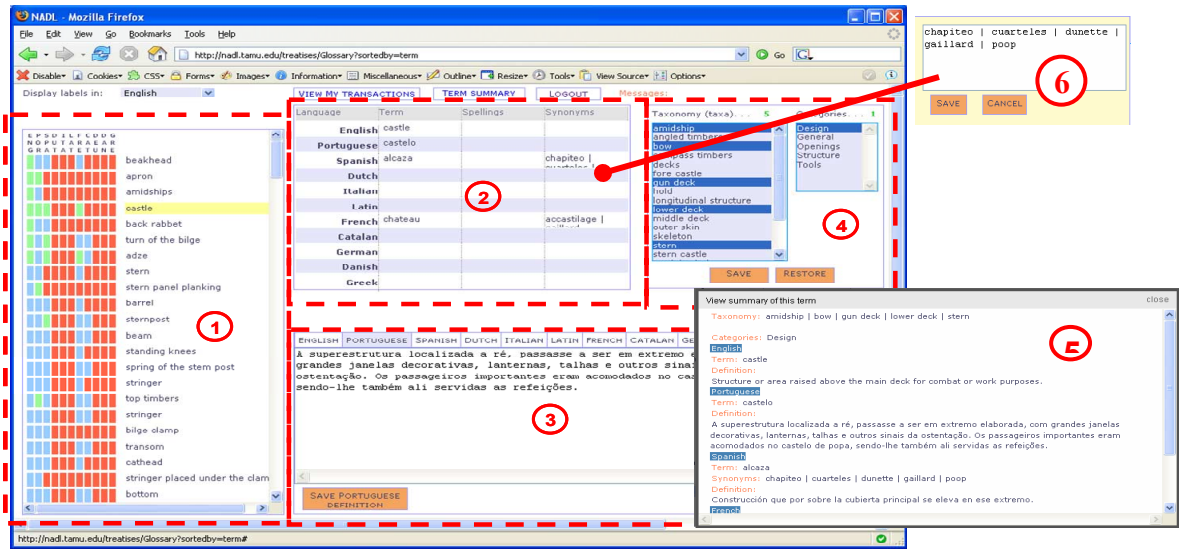

Figure 2. The web-based interface of the multilingual glossary.