Challenges and solutions towards creating a semantic network of historical maritime data

Pavlos Fafalios*[1], Korina Doerr[1], Athina Kritsotaki[1], Kostas Petrakis[1], Giorgos Samaritakis[1], Anastasia Axaridou[1], George Bruseker[2], Yannis Tzitzikas[1], and Martin Doerr[1]

[1]Institute of Computer Science, FORTH-ICS – Greece
[2]Takin.solutions – Bulgaria

Abstract

Descriptive and empirical sciences, such as History, Archaeology and Cultural Heritage, are the sciences that collect, observe and describe phenomena of the past in order to explain them and draw interpretative conclusions about their behaviour and their relationships under given circumstances. A vast area of research in these sciences concerns the quantitative analysis of empirical facts, their description and interpretation of possible causes, influences and evolution trends.

SeaLiT [1] is a History research project in this context that explores the transition from sail to steam navigation and its effects on seafaring populations in the Mediterranean and the Black Sea between the 1850s and the 1920s. Historians and researchers in this project investigate, besides others, the maritime labour market, the evolving relations among shipowners, captain, crew and local societies, and the development of new business strategies, trade routes and navigation patterns, during the transitional period from sail to steam. The information management challenge that SeaLit faces is the ability to faithfully catalog historical data sources as a primary source for research and for long-term preservation, while integrating this data into a common form from which historical analysis and questions can be carried out efficiently.

In this work, we present the methodology we followed and the challenges we faced for supporting the historians of SeaLit in *digitizing* and *exploring* their unique information sources. These information sources range from ship logbooks, crew lists, payrolls and student registers, to civil registers, business records, account books and consulate reports, written in different languages (Spanish, Italian, French, Russian, and Greek). We specifically describe the process of building a semantic network of integrated maritime data, which includes i) developing tools that allow historians digitizing and curating the sources by maintaining at the same time the important information about the provenance of the curated data, ii) modeling and integrating the data using established standards, in particular CIDOC-CRM [2] and its compatible models, for supporting data exchange, interoperability and long-term validity, and iii) offering exploration and visualization tools on top of the semantic network that enable historians take advantage of the digitization and integration of the original information sources. We further discuss the difficulties we faced at each step of the process and related implementation issues, as well as our overall ambition for a standardized digitization and integration workflow.

---

*Speaker