

Thinking like the "Modern Operating Systems": The Omega architecture and the Clavius on the Web project

Angelo Mario **Del Grosso**, Emiliano **Giovannetti**, Simone **Marchi**
Institute for Computational Linguistics "A. Zampolli" ILC-CNR
{nome.cognome}@ilc.cnr.it

Introduction

The current digital turn in studying and analyzing historical documents results in both having machine actionable cultural data and providing software able to process them. However, these data and services often lack in integration strategies among them in order to be reused in other contexts different from the original ones.

As pointed out by Franz Fischer in a worthy of note article: "There is no out-of-the-box software available for creating truly critical and truly digital editions at the same time" [1]. Likewise, Monica Berti stated that is now important to "build a model for representing quotations and text reuses of lost works in a digital environment" [2]. In this vision Bridget Almas is in charge of developing an integrated platform for collaboratively transcribing, editing, and translating historical documents and texts. She claimed that through this platform, called Perseids, students and scholars are able to create open source digital scholarly editions [3].

A number of interesting projects are currently under development to realize general models, digital services, and online tools that can be adopted as part of a long-term infrastructure for managing digital editions. Among Perseids and others, we cite as reference systems (a) the Textual Community project led by P. Robinson and B. Bordalejo, (b) the AustESE project led by the Australian eResearch group, (c) the Tagore Online Variorum "Bichitra" project led by Sukanta Chaudhuri, (d) Homer Multitext led by Neel Smith and Christopher Blackwell, (e) Sharing Ancient Wisdoms founded by the HERA network.

Method: the Omega framework approach

The Literary Computing group of the Institute for Computational Linguistics at the National Research Council of Italy (ILC-CNR) < <http://licolab.ilc.cnr.it> > is carrying on a line of research about designing software models for textual scholarship as well as implementing them by using cutting-edge software engineering approaches and technologies. The research work is aimed at providing a general framework (called Omega) [4] - inherently conceived with the object-oriented paradigm and semantic web technologies - suitable for studying historical and literary documents and texts. Indeed, a rigorous use-case and domain-driven design is at the basis of the framework's high modularity. Omega is organized around a micro-kernel architecture for the sake of extensibility and pluggability like the "Modern Operating Systems" (A.S. Tanenbaum). Moreover, our model puts into effect the best design

patterns in order to reuse general solutions to recurring problems in the domain of historical textual studies. Omega provides a collection of Abstract Data Types (ADTs) which defines the Domain Specific Application Programming Interfaces (DS-APIs), independent from any particular implementation, combining the internal representation of textual data and the available processing on that data in a single unit of concerns. The APIs represent the provided services at different levels of granularity, id est, they are organized in different layers of abstraction: 1. low-level API (Class API), 2. middle-level API (Module API) 3. high-level API (Web API). The ongoing implementation of the platform can be followed on Github < <https://github.com/literarycomputinglab> >.

Case study: the Clavius on the Web project

The aforementioned digital scholarly approach has been already adopted within the Clavius on The Web project, funded by the Registro.it [5]. The aim of this initiative is to preserve and promote the Christophorus Clavius' (1537-1612) correspondence between some important scientists of his time (such as Galileo Galilei and Tycho Brahe). Clavius was a German Jesuit - mathematician and astronomer - famous for his work on Gregorian Calendar and the editing work on the Euclide's Elements and the "Commentarius" about Sacrobosco's "De Sphaera Mundi".

The historical archives of the Pontifical Gregorian University (APUG) own these documents written in Latin and Italian. The project has started in 2012 and it is a collaboration among various research partners, namely the ILC-CNR, APUG and the Institute of Informatics and Telematics (IIT-CNR).

During the project activities a number of web tools have been developed following the method introduced above. These tools allow scholars to transcribe, linguistically process, and annotate - both lexically and semantically - the corpus. One of the initiatives of the project involved several groups of students to test the software in their educational training. This experience also provided interesting insights about the impact of the digital paradigm in educational systems.

The main components of the platform that have been developed are listed below [6][7]:

- The Digital Archive < <http://claviusontheweb.it> >
- The Text Encoder and Annotator < <http://wafi.iit.cnr.it/webvis/tea> >
- The Annotarium search tool < <http://wafi.iit.cnr.it/webvis/annotarium> >
- The Lexical-Ontological resource < <http://licodemo.ilc.cnr.it:8080/clavius> >

Conclusion

This contribution describes some critical issues in the design and implementation of a flexible digital scholarly platform (Omega) for the study of documents, texts and languages. The approach has been illustrated through the Clavius on the Web project case study.

Among the next steps, we are working to integrate the software and the resources within international research infrastructures such as CLARIN and DARIAH.

We believe the Open Conference on Digital Infrastructures for Global Philology would be a perfect context to illustrate the outcomes of our initiatives. Moreover, it would be extremely useful to receive a feedback about the adopted models and techniques.

References

- [1] Franz Fischer, "All texts are equal, but... Textual Plurality and the Critical Text in Digital Scholarly Editions," *Variants* 10 (2013): 77-92.
- [2] Monica Berti et al., "The Linked Fragment: TEI and the Encoding of Text Reuses of Lost Authors", *JTEI* (8), (2014).
- [3] Bridget Almas and Marie-Claire Beaulieu. "Developing a New Integrated Editing Platform for Source Documents in Classics," *LLC* (2013) 28 (4): 493-503.
- [4] Angelo M. Del Grosso, et al., "Defining the Core Entities of an Environment for Textual Processing in Literary Computing," in *DH2016: Kraków*, pp. 771-775.
- [5] Andrea Marchetti et al., "Il progetto Clavius on the Web entra nelle scuole," *Digitalia* (2015): 74-84.
- [6] Silvia Piccini et al., "When Traditional Ontologies are not Enough: Modelling and Visualizing Dynamic Ontologies in Semantic-Based Access to Texts," in *DH2016. Kraków*, pp. 653-655.
- [7] Fabio Valsecchi et al., "Text Encoder and Annotator: an all-in-one editor for transcribing and annotating manuscripts with RDF," in *ESWC 2016*, Springer: 399-407.