Porphyry & Steatite: Software layers for sense makers in humanities

Aurélien Bénel[1], Andrea Iacovella[2] and Sylvie Calabretto[3]

[1] Laboratoire Tech-CICO (Institut Charles Delaunay), CNRS / Université de technologie de Troyes,
<http://www.porphyry.org/Members/abenel>
[2] Centre d'édition numérique scientifique du CNRS, Lyon,
<http://www.porphyry.org/Members/aiacovella>
[3] Laboratoire LIRIS, CNRS / Institut national des sciences appliquées de Lyon,
<http://www.porphyry.org/artcadhi/membres/page?id=scalabretto>

Abstract

The aim of ARTCADHi, the transdisciplinary research network which designed *Porphyry* and *Steatite* software, is to experiment the feasibility of a digital hermeneutics. The main idea is to offer a digital place where scholars could build their interpretation about document corpora and confront it with the viewpoints of other scholars. In order to be easily confronted the viewpoints are modeled through a network of topics and document fragments inspired by the traditional methods for text criticism.

Keywords: Viewpoint, Topic, Document fragment, Source, Digital library, Knowledge modeling.

1 Overview

Our researches deal with the "sense making" process. To experiment our theories, we design and build new kinds of software for people whose jobs go beyond a given procedure (lawyers, doctors, engineers… and especially historians and archaeologists). One of the common features of these kinds of jobs is that their practitioners build their own interpretation over a corpus of documents.

1.1 Features

We provide two complementary software prototypes.

The first software layer, called *Steatite*, is used for:

- Making accessible copies from original contents,
- Tidying up a corpus,
- Giving situational clues (metadata),
- Archiving revisions.

On top of *Steatite*, stands a second software layer, called *Porphyry*, which is used for:

- Selecting fragments in document sources,
- Gathering fragments (or sources) in collections,
- Organizing collections into an outline,
- Writing a new document from the outline,
- Submitting the document to peers or superiors,
- Publishing the document (making it public).

1.2 Aims

The tasks listed earlier match exactly what scholars do since the middle ages [11]. If that means it really corresponds to what people do with a pen and paper, it means also that we could wonder whether it is necessary to go digital or not. In our opinion, this "simple" change of medium could drastically modify practices and disciplines.

First, it enhances the way scholars make explicit the links they find between documents.

Secondly, it is far easier for them to share their primary sources and their interpretations (from narrow groups to world wide publishing).

Thirdly, it is now possible to trace every action done on the digital media and then to follow the interpretation trail leading to a research result.

2 Inside look

The original *Porphyry/Steatite* model [2] has been recently adapted to be compatible with the model of *Agorae* [4]. In both of these models, users can create multiple viewpoints on shared entities and these viewpoints are modeled with a network of topics.

A common model (cf. Fig. 1) and protocol, called *HyperTopic*, has been defined and implemented in the *Argos server* to be used by *Agorae*, *Porphyry/Steatite* and similar clients.

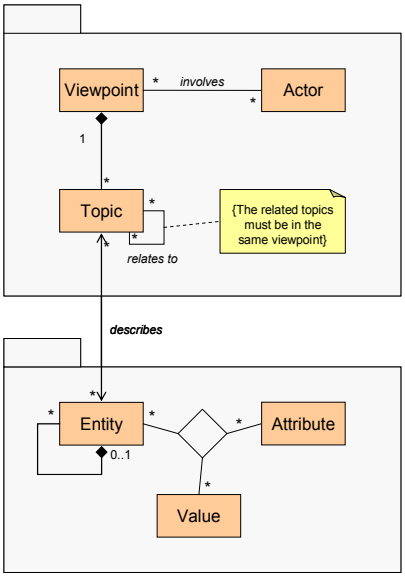


Fig. 1 – The *HyperTopic* model implemented in the *Argos* server (UML class diagram)

In addition to this common model, the *Porphyry/Steatite* framework has a few specificities.

First, because topic networks in *Porphyry* are done by the users themselves rather than by computer scientists, we decided to keep the relations between topics simple and use only one kind of them: "topic-inclusion" (which forms a "directed acyclic graph"). Despite its simplicity, this relation type can be used for hyponymy (e.g. "philosopher is a kind of human"), meronymy (e.g. "head is a part of human"), instantiation (e.g. "Socrates is a philosopher"), and even for the most advanced patterns (associations, association links, association roles, etc.).

Secondly, because we are more interested in the "sense making" process than in its results, *Porphyry* traces every situation when a context between two information objects is tied or untied by someone.

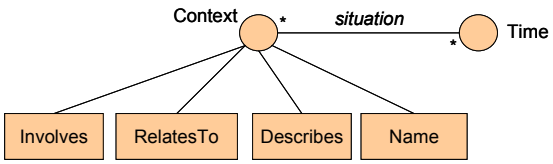


Fig. 2 – Traceability in *Porphyry* (UML class diagram)

Last but not least, because entities must be shared by different author groups and even accessible to readers to follow authors' interpretation trails, these entities must be available *in* the system. Therefore, the entities handled by *Porphyry*/*Steatite* are documentary fragments, sources (images and plain texts) and folders (Fig.2).

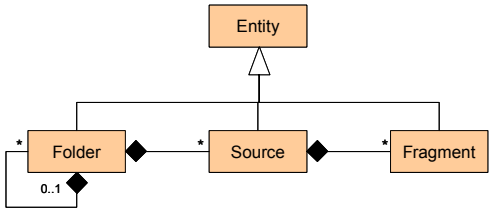


Fig. 3 – Documentary entities handled in *Porphyry/Steatite* (UML class diagram)

3 Genesis

Porphyry and *Steatite* would not exist without the transdisciplinary seminar of ARTCADHi[1]. Researchers in history, historiography, archaeology, and art history came with their document corpora and their related research topics. The transdisciplinary study of corpora and research practices lead to the design of theoretical models.

First, archaeologists had been early adopters of databases and expert systems. They had experienced the limits of impersonal "data" and "facts". The archaeologists in our team proposed instead the notion of "source" [6]: from ancient texts, to artifacts photographs and even to researchers' articles, all of those are sources (i.e. authored "discourses").

Secondly, "viewpoints" was a way to bring intersubjectivity (interpretations conflicts), whose need in archaeological information systems had arisen for long without being addressed [8].

Thirdly, the evolving topic network built on top of document fragments was inspired by the semiotic analysis of excavations reports [9] partially based on the "interpretative semantics" [14, 13].

Lastly, the idea to consider historical knowledge as a process is the root of historiography [5].

Indeed, through these discussions, the knowledge engineers of our team had to move from the "ontological approach" they followed before [12, 16] to a new one they called an "hermeneutical approach" [15, 3].

4 Evaluation process

Each time it is feasible, the computational models which underlies *Porphyry* and *Steatite* are updated to conform to the theoretical models discussed in the seminar. The new prototypes are then distributed with an open-source license so that every scholar can use and modify them[2].

For a deeper evaluation process, we also support several "expertise platforms" settled around a researcher and her

[1] *ARTCADHi : Atelier de Recherches Transdisciplinaires sur la Construction du sens en Archéologie et autres Disciplines Historiques*. http://www.porphyry.org/ [FR]

[2] *Prototypes*. http://www.porphyry.org/prototypes/ [FR]

corpus in her own laboratory (cf. Fig. 4). While the researcher interprets her corpus, the software she use and the underlain theoretical models are tested. The results are then discussed in the seminar in order to improve the theoretical models.

5 Beyond Steatite and Porphyry

As we saw earlier, *Porphyry* allows researchers to confront different viewpoints on the same document corpora. But could we confront viewpoints dealing with the same area of interest but on different corpora?

Our first attempt was to compare viewpoints on archaeological time [1]. More generally [7, 9], such a confrontation needs a community with both its consensual knowledge (e.g. chronology) and its methodology (stratigraphicy, stylistic studies, etc.). Then an "editor" involved in building the community (conference organizer, professor, etc.) can link together topics from different viewpoints (with different names, sites…) or from the consensual knowledge. By constraint propagation, it is possible to detect what is a formal contradiction and which could potentially falsify the consensual knowledge. Ongoing works aim at offering a third software layer on top of Steatite and Porphyry for these community features [7].

References

[1] T. Accary-Barbier, S. Calabretto, Temporal reasoning on chronological annotation, *Proceedings of the fourth Mexican International Conference on Artificial Intelligence, Monterrey, November 14-18, 2005. Lecture Notes in Artificial Intelligence,* #3789, Berlin: Springer-Verlag. p.51-60.

[2] A. Bénel, S. Calabretto, A. Iacovella, J.-M. Pinon, Porphyry 2001: Semantics for scholarly publications retrieval, *Proceedings of the thirteenth International Symposium on Methodologies for Intelligent Systems, Lyon, June 26-29, 2002*, *Lecture Notes in Artificial Intelligence*, #2366, p. 351-361, Berlin: Springer-Verlag, 2002.

[3] A. Bénel, E. Egyed-Zsigmond, Y. Prié, S. Calabretto, A. Mille, Truth in the Digital Library: From Ontological to Hermeneutical Systems, *Proceedings of the fifth European Conference on Research and Advanced Technology for Digital Libraries, Darmstadt, September 4-9, 2001, Lecture Notes in Computer Science*, #2163, p. 366-377, Berlin: Springer-Verlag, 2001.

[4] J.-P. Cahier, L'H. Zaher. J.-P. Leboeuf, X. Pétard, C. Guittard, Experimentation of a socially constructed "Topic Map" by the OSS community, *Proceedings of the IJCAI-05 workshop on Knowledge Management and Ontology Management, Edinburgh, August 1, 2005.* Available on: <http://www.limsi.fr/~xpetard/articles/cah-et-al-OM2005-Final.pdf>

[5] M. de Certeau, L'écriture de l'histoire, 358 p., Paris: Gallimard, 1975.

[6] P. Desfarges, B. Helly, L'archéologie, système d'information scientifique. *Aplicaciones Informaticás en Arqueologia : Teorias y sistemas, Saint-Germain-en-Laye, 1991.*

[7] S. Gesche, S. Calabretto, G. Caplat, Confrontation de points de vue dans le système Porphyry, *Actes des sixièmes journées Extraction et Gestion des Connaissances, Lille, 17-20 janvier 2006,* p.725-726, Toulouse: Editions Cépaduès.

[8] R. Ginouvès, A.-M. Guimier-Sorbets, L*a constitution des données en archéologie classique*, 161 p., Paris: CNRS Editions, 1978.

[9] A. Iacovella, Modèle opératoire de navigation pour les experts : Appropriation sémantique et délimitation de l'espace documentaire, *La navigation (Numéro spécial), Les cahiers du Numérique,* vol.3, n°3, Paris : Hermès, 2002. Available on: <http://www.porphyry.org/Members/aiacovella/iacovella_LCN_02.pdf>

[10] A. Iacovella, A. Bénel, X. Pétard, B. Helly, Corpus scientifiques numérisés : Savoirs de référence et points de vue des experts, To be published in: J. Charlet (Ed.), *Documents et contenu : création, indexation, navigation*, Paris: Hermès.

[11] G. Lortal, M. Lewkowicz, A. Todiracu - Courtier, Modélisation de l'activité d'annotation discursive pour la conception d'un collecticiel support à l'herméneutique, *Actes des 16ème journées francophones d'ingénierie des connaissances*, p. 169-180, Grenoble: PUG, 2005. Available on: <http://casanuestra.free.fr/Lortal_IC05.pdf>

[12] J.-M. Pinon, S. Calabretto, L. Poullet, Document semantic model: An experiment with patient medical record, Proceedings of the conference on Electronic Publishing, Canterbury, April 14-16, 1997, p.262-272, Washington: ICCC Press.

[13] Y. Prié, *Contribution à une clarification des rapports entre Sémantique Interprétative et Informatique*, Master thesis in Computer Science, Université Rennes 1, 1995. Available on: <http://revue-texto.net/Inedits/Prie95.pdf >

[14] F. Rastier, Sens et signification, *Protée, printemps 1998*. p. 7-18.

[15] P. Ricœur, *Du texte à l'action : Essais d'herméneutique II*, Paris: Seuil, 1986.

[16] C. Roussey, S. Calabretto, J.-M. Pinon, A new conceptual graph formalism adapted for multilingual information retrieval purposes, *Proceedings of the twelfth international conference on Database and Expert Systems Applications, Lecture Notes in Computer Science,* #2113, p.92-101, Berlin: Springer-Verlag, 2001.

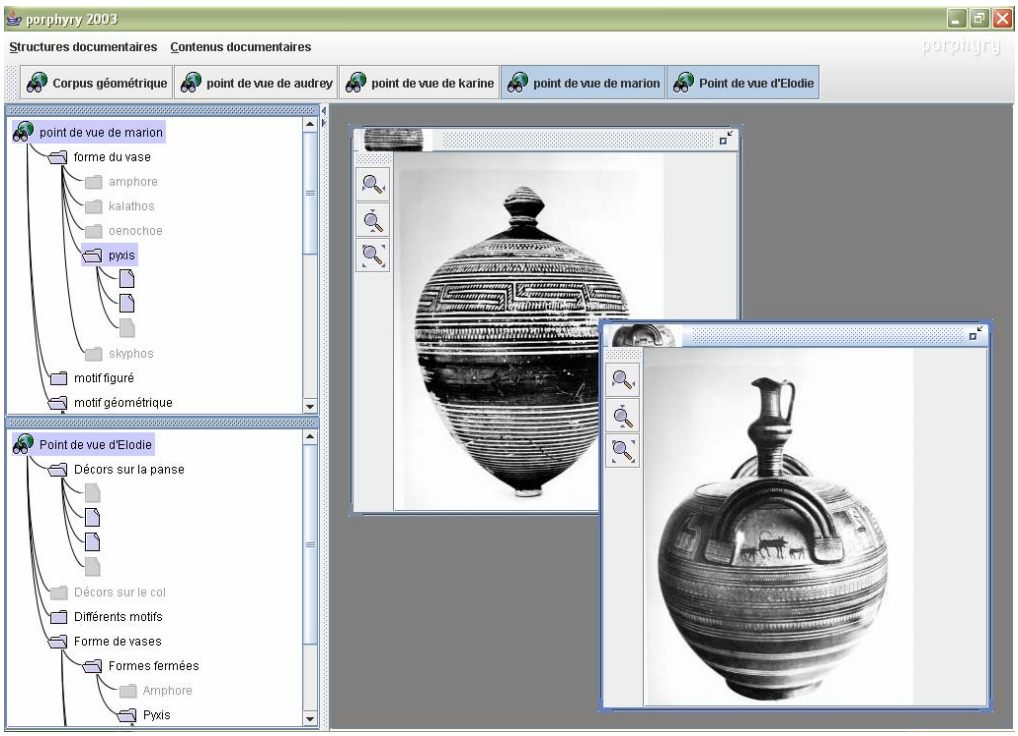


Fig. 4 – Art history students using *Porphyry* (screenshot)