Expression and share of temporal knowledge in archaeological digital documentation

Tiphaine Accary-Barbier[1], Sylvie Calabretto[1], and Aurélien Bénel[2]

[1] LIRIS UMR CNRS 5205, INSA de Lyon,
Bat. Blaise Pascal 7, avenue Jean Capelle
69621 Villeurbanne cedex, France
{tiphaine.accary, sylvie.calabretto}@liris.cnrs.fr
http://liris.cnrs.fr

[2] Equipe Tech-CICO, Laboratoire ISTIT (UTT-CNRS)
Université de Technologie de Troyes
12 rue Marie Curie, BP 2060, 10010 Troyes CEDEX
aurelien.benel@utt.fr

Abstract. Our previous work on specialized digital libraries showed that each expert was brought to enrich the documentary corpus with his/her point of view (expressed by a model of knowledge). These points of view can be contradictory and therefore, it would be interesting to provide expert communities with tools, enabling them to point out meaning dissensions. In Archaeology, comparison between different temporal models of knowledge then makes it possible to publish relative models of knowledge and tend towards the building of a community knowledge.

1 Introduction

Today, the concept of digital libraries exceeds that of document warehouses: it also includes a social dimension. Libraries can be considered as "social places" in which each document can have multiple meanings depending on its reading conditions. This construction of meaning will be realized through a corpus reorganization according to each reader's "point of view" [4].

1.1 Context: archiving and modeling of archaeological documents

Our works are related to a digitalization project by the French school of archaeology in Athens[3] wich aims was to give online access to one of its main periodical publications: "La Chronique des fouilles", an archeological excavations and findings yearly survey. This corpus contains about 12.000 pages and concerns 80 years of archaeological activity in Greece and Cyprus. The result of this first project was the setting on line of these collections in an easily consultable numerical form on the Web portal CEFAEL[4].

[3] Ecole française d'Athènes (http://www.efa.gr)
[4] http://cefael.efa.gr

However, the setting in line of such collections raises the problem of information retrieval into these structures. Concerning the "Chronique", this problem is not new and not related to digitalization. To help consultation, Publishers have tried for 80 years to structure the contents of each publication wich is made of small independent parts (about 50,000): These parts are hierarchically structured and indexed according to artifact location, dating and typology. The access to the numerical version of the document makes it possible to plan other methods of indexing in order to optimize the search for information. However, none of the retrieval systems using thesauri tried by archaeologists satisfied them: those based on automatic indexing are inadequate because they deals with out-of-context terms and those based on manual indexing are to hard to manage over time (they needs periodic updates of both thesauri and indexes in order to reflect science progress). This second project led to the developpement of the scholarly publication retrieval system PORPHYRY[5] currently in open-source. This system proposes a framework for free descriptors created either by machines (words or phrases occurrences) or by human (categories, annotations). Most of all, owing to collaboration and dynamics, it can be used as a debate media. Every annotation is dated and authored, so that it can be interpreted, contradicted by another annotation, or considered as obsolete.

The report of these differences between the "points of view" expressed by the experts proves to us that it would be interesting to provide to the expert communities tools, enabling them to specify dissensions of significance. In Archaeology, time can be considered as one of the main documentary organization dimension. We thus were interested into uses of temporal dimensions of the document as "dynamic writing" support between documents [3]. Our aim in this project is to propose some tools allowing the comparison between different temporal models of knowledge. Such tools then make it possible to publish relative temporal models of knowledge and to tend towards the building of a community knowledge.

1.2 Background: temporal indexing of archaeological documents

Manual temporalisation of a corpus consists in associating documents (or fragments of these) with annotations having temporal contents. Such annotations can possibly result from the organization of knowledge (thesaurus or ontologies) [6]. All these annotations can then form a temporal structure which will enrich knowledge related to the document. This structure will also allow a "temporalized" exploitation of the corpus.

If many systems of document management integrate the concept of date in order to temporize documents, these dates are essencialy descriptive information or document management information. The dates then constitute a category of the profile of the document. Data we wish to exploit in the archaeological field concern the time of the documentary contents. Interpretations and additional

[5] http://www.porphyry.org

temporal knowledge that the reader will associate to a document during his indexing constitute a particular class of annotations on the documentary corpus. In order to be able to deal with these annotations, we necessarily need a system by wich they can be ordered. Some digital library languages dedicated to historical disciplines such as RAMEAU [12], PACTOLS [13] or PORPHYRY [5] make it possible to express this type of ordered annotations. However they remain far too limited in terms of temporal exploitation and therefore the question of processing of this data remains. Indeed, thanks to these systems one can find a document by formulating a request, but not for working on the knowledge contained around this information.

1.3 Contribution

"Publication" of points of view shared by each expert works within a virtual agora. Therefore, in order to transform debates raised by this work into references for the expert, an "editing" stage is necessary. In this respect, we may set up community regulation functions which allow to detect similarities between points of view and dissenting opinions between experts.

The idea we develop here concerns assistance tools based on the exploitation of temporal characteristics of annotations. In parallel with documentary operations of annotations made by researchers, these tools will make possible the detection of temporal inconsistencies contained among different points of view. They could then validate or invalidate temporal assumptions of experts.

It is advisable to point out importance of the archaeological time and of its mechanism of construction so as to apprehend all the elements necessary to this annotation exploitation. This approach allows to understand that the archaeological model is based on a numerous relative temporal schedules inter-connected in a very flexible way (section 2). It was necessary to take it into account to adopt a data-processing model of time representation more compliant with the archaeological model. After a short comparison of the mains families of representation used in other fields, it appeared that the constraints propagation model presents a double advantage: (1) to permit a rather faithful temporal representation and (2) to offer possibilities of development of tools for new connections emergence(section 3.1). In order to validate our model we then show how it could be integrated in the indexing system PORPHYRY. Whith this aim in mind, we initially propose to provide the network of descriptors with transverse relations wich can explain the reports/ratios of connexity and simultaneity and then devellop a set of translation rules beetween models (section 3.2).

The expression of points of view comes from the graph of descriptors produced by experts. We show why pointing out differences between expert's points of view remains very interesting in disciplines like Archaeology in which both mechanisms of hypothesis refutation and construction build the knowledge (section 3.3). We present a short example of those mechanisms of knowledge revaluation (section 4) and we briefly conclude.

2 Time in archaeological corpora

When a reader works on a corpus, he/she restructures documentary units through a network composed of semi-formal object descriptors. Insofar as archaeological documentary collections are made up of various objects (photographs, plans, text, diagrams) it is advisable to highlight common dimensions of reorganization more semantic than formal characteristics. In Archaeology, the three main dimensions of documentary object description are time, space and styles. We then worked on the description, the construction, and the modelling of archaeological time.

2.1 Search for clues

Contrary to an opinion well anchored in some archaeologists mind, time is not a data but a result. It results from the combination of spatial, stylistic or natural clues relying on more or less reliable heuristics.

One of the method used to date a manufactured object is to identify its style. We suppose that two similar objects in their form or their technique are contemporary. We can then establish a relative order between all observed objects. Evolutionist heuristic are also used to give chronology of styles. For example, an *abstract* style was often followed by a *figurative* style and later by a *declining* style. This kind of heuristics has to be carefully chosen because resulted chronologies were often refuted. Stylistic indications depend very often on spatial characteristics of objects. Indeed, in the diffusionist theory, we will consider that a style can appear with a certain delay in a close region. Moreover, two styles can sometimes be similar in spite of chronological and space large variations. The most current spatial indication is the relative depth of the vestiges. The method, called "stratigraphy", is based on the principle that in a succession of layers, higher units of stratification are youngest and lower are oldest [10]. This rule of temporal sequences admits obviously counterexamples due to natural or human rehandlings. The method can however take into account the more current cases, for example pits and walls. If depth is the spatial indication the most used for dating, geographical position of a vestige could also be used. A style will appear in a place and be diffused gradually in close areas. The distance to the original place could then reveal the time necessary to the diffusion. Still, this type of indication remains very hazardous. Thanks to physicists and chemists we have quantitative theories on the evolution of certain properties of vestiges. The dating with carbon-14, for example, relies upon the assumption that a constant share of the carbon dioxide of the atmosphere is made radioactive by the cosmic rays. However, we can note that for historical times as old Egypt, archaeologists obtained incoherent radiocarbon dating compared with old texts. Therefore the physicists had had to refute their assumption according to which cosmic radiations had been constant in time. It was thus necessary to spread out dating with radiocarbon according to other methods.

2.2 Development of knowledge

In order to obtain a result, we need to cross the different indications characterizing an object. It is what occurs during an identification of stratigraphic unit: experts cross the sequence given by the stacking of layers with the style of artefacts contained in these layers. Thus, on figure 1, the identified object is characterized by its style, the sector of excavation where it was found, and a chemical estimation of its dating.

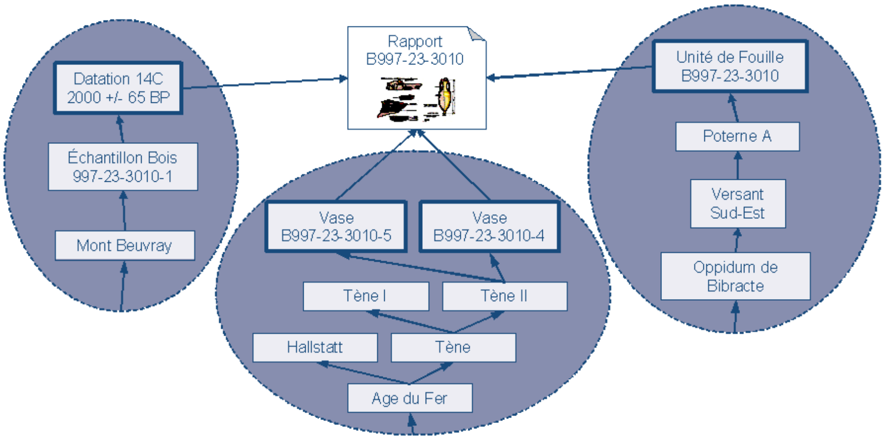

Fig. 1. Crossing of clues

We often wrongly assume that a vestige dates a layer. Actually, it allows to define a *terminus post quem* for the layer. Thus, when the archaeologists of the future will discover the Louvre museum, they will not conclude that such house dates from the Egyptian or Babylonian time, but that the palace is posterior to Egyptian and Babylonian civilization.

We notice here all the complexity of the archaeological time. This is the result of comparisons which validity decreases with the distance between the two compared places. In Archaeology there is not a single chronology but valid chronologies in a given geographical area with sometimes possible correspondences between them [7].

3 Proposals

3.1 Reference model

Time representation applies to many fields such as planning, natural language processing and supervision [8]. Among the temporal models used in these fields, approaches based on constraint propagation seems to be the more adapted to the archaeological field. As we have shown before, archaeological time is relative. We thus automatically excluded from our choice all models of representation founded on absolute dating. Elementary representations using scheduling of dates [2] and the PERT models cannot be appropriate. We also need to handle events corresponding to periods of time. This assumption discredits models which primitives are moments (specific events). For this reason models of Dean and McDermott [9], which are based on the mesurability of events, cannot be used.

Basic models using constraints propagation as that defined by Vilain and Kautz [11] could answer better to the relativity of archaeological representations. However it permit to express strong relations between points ($<$,$>$,$=$) and is too much precise to have an interest in incomplete knowledge construction. The model described by Allen [2] then constitutes a solid base for a valid temporal model for archaeology.

Relations and Notations	Signification	End-points relations			
		A^-B^-	A^-B^+	A^+B^-	A^+B^+
before | after $A\{b\}B \mid B\{bi\}A$	A B	$<$	$<$	$<$	$<$
meets | met by $A\{m\}B \mid B\{mi\}A$	A B	$<$	$<$	$=$	$<$
equals $A\{eq\}B$	A B	$=$	$<$	$>$	$=$
during | contains $A\{d\}B \mid B\{di\}A$	A B	$>$	$<$	$>$	$<$
start | started by $A\{s\}B \mid B\{si\}A$	A B	$=$	$<$	$>$	$<$
finish | finished by $A\{f\}B \mid B\{fi\}A$	A B	$>$	$<$	$>$	$=$
overlaps | overlaped by $A\{o\}B \mid B\{oi\}A$	A B	$<$	$<$	$>$	$<$

Table 1. Allen's basic relations and end-points conversion.

This model uses thirteen primitive relations, mutually exclusive, to describe possible relative positions between two temporal intervals (Table 1). Each of these relations can also be expressed with the three ordinary relations $<$,$>$,$=$ between limits of intervals we want to compare. Then, temporal reasoning will be reduced to the calculation of the transitive closing of the set of relations between limits.

The main reason which has motivated this choice of model is that it allows, using disjunctions of primitive relations, to express uncertainties specific to knowledge of archaeological field. Thus, when we have any ordered information between two events, we can label it with the complete disjunction (OR) of all existing relations. In the same way, if there is an uncertainty, we can characterize an arc with a disjunction of the supposed relations.

The second reason of our choice holds in the principle of constraints propagation. Thanks to a set of rules describing the composition of these thirteen relations (for example the relations $A < D$ and $D < B$ induce $A < B$), the graph is maintained complete at any moment and each inconsistency is detected. Then we will be able to maintain the temporal graph complete at any moment and to detect inconsistencies induced by the addition of new information. This property of the model is completely adapted to the work of archaeology researcher whose essential task is the refutation of assumptions.

3.2 Models of translation

In some usual indexing systems, the document description is made throught insulated descriptors. The main quality of PORPHYRY system is the using of descriptors organized in a partial order in accordance with specialization relations. Each user can define his own descriptors and organize them as he wishes.

The model of knowledge is built on all temporal annotations which were attached to documents. We work on a wide version of the PORPHYRY graph, in which the model of knowledge is expressed by a graph of descriptors connected with relations of inclusion, overlapping and connexity. Even if we can already identify on the network some coherence rules, these detections are too specific and insufficient.

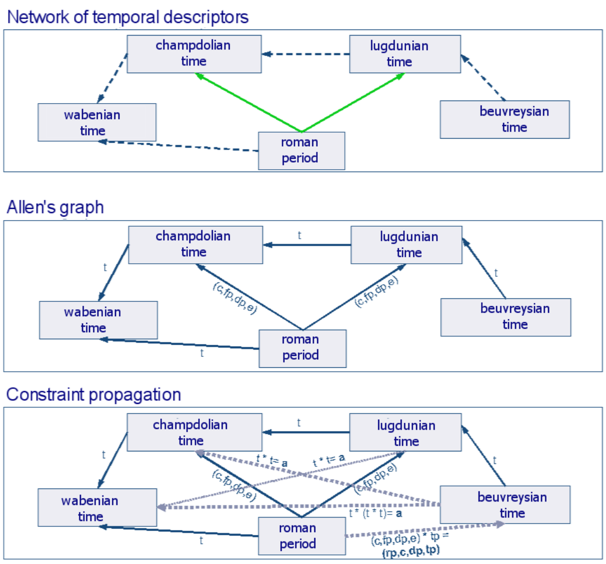

Fig. 2. Transformation of the network into Allen's graph and constraint propagation

In order to apply constraint propagation, we need to transform the temporal descriptor network into an Allen's graph. For this purpose, we propose a set of rules which transform the network's relations into disjunctions of Allen's relations (Table 2).

Network relations	Translation into Allen's disjunction
A meets B	$meets(A, B)$
A fuzzy-before B	$before(A, B) \vee meets(A, B)$
A fuzzy-during B	$egal(A, B) \vee during(A, B) \vee starts(A, B) \vee finishes(A, B)$

Table 2. Set of rules describe into [1] which allows to transform the description network into an Allen's graph .

Figure 2 shows an example of network "translated" into an Allen's graph thanks to this set of rules. Once the translation carried out, we will be able to use the constraint propagation on the corresponding graph. The system will then calculate all the relations between each descriptor. Thus, over the example, the relation between *roman period* and *wabenian time* will be calculated as made up of the relation *roman period / champdolian time* with the relation *champdolian time / wabenian time.*

3.3 Work around points of view

Checking the coherence of different models of the discipline would have an exorbitant algorithmic cost and would not be very usefull. However it is significant to take into account the sharing of works which have common runs. When a researcher publishes in the community his/her corpus reorganization structures, he/she permits other researchers to join his point of view.

In practice, each expert have an inter-subjective workspace in which he/she can import the fragments of the points of view he is interested in. In his private space, the researcher can use fragments as he want to connect them together. He/She establishes his own model of knowledge. However, each imported point of view remains in relation with its original point of view. Thus, if the original point of view changes, this evolution is also proposed to the expert in his private space.

Therefore this dynamic confrontation of points of view becomes the "engine" of the construction of knowledge. Thus, on figure 3, the point of view named 2 has evolved by adding a connexity relation between periods under studdy. This point of view remains very consistent. On the other hand its consequences in the private workspace enters in conflict with some relations calculated thanks to the propagation of relations added by the expert. There is an emergence of divergence of current of thought between two models. Following this observation, the expert must choose to refute his proper work or to break with this other point of view. In all the cases, he/she will have been informed of the existence of another current of thought.

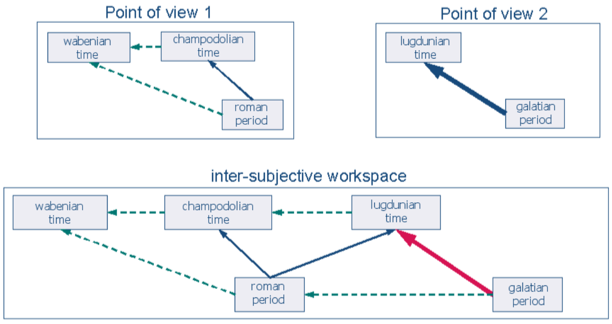

Fig. 3. Examination of points of view and inconsistencies

4 Example

Constraints propagation allows the detection of conflicts in the information network. This detection is of real interest in a field such as Archaeology, in which debates generated by dissensions allow knowledge evolutions. On figure 4, we can note that the general model shared by the community holds that Age of Bronze follows Neolithic era directly. In his workspace, the researcher dated layer 3 from the Neolithic time and layer 1 from the Age of Bronze. The local model related to excavation proposes, thanks to the study of the stratigraphic cut, a relative chronology of the identified layers.

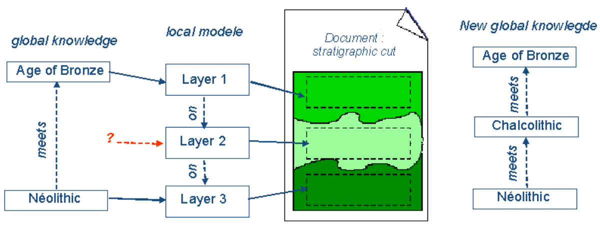

Fig. 4. Setting in doubt of knowledge

The confrontation of the two models highlights an inconsistency: layer 2 is clearly identified and belongs therefore neither to Neolithic era, nor to Age of Bronze, but the general model does not consider the possibility of an intermediate period. The system will then announce this inconsistency to the researcher who can then point out the problem, identify a solution and eventually correct it. On this example, the researcher puts an assumption refuting the general model: the existence of an Age of Copper, located between Neolithic era and Age of Bronze, unknown to the community. He/She could also choose to invalidate his local model and to reject the excavation results.

5 Conclusion

It is particularly interesting to allow communities to build their own models of knowledge. However, it is even more useful to enable them to confront these models with those built by communities working on similar themes.

We have presented here an innovating approach of the temporal data processing. Thanks to both mechanisms of importation of points of view and detection of inconsistencies, our model allows: (1) to permit interactions between different research teams; (2) to obtain coherence among the experts into a team.

In our future work, we plan to compare possible visualization means for the results. We are indeed convinced that an adequate mode, in addition to facilitate the data processing, could lead the researcher to the construction of new assumptions.

References

1. T. Accary, A. Bénel, S. Calabretto, A. Iacovella: Confrontation de points de vue sur des corpus documentaires : Le cas de la modélisation du temps archéologique. In: Actes du 14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle [RFIA], Toulouse, 28-30 janvier 2004, p.197-205.
2. J.F. Allen: Time and Time Again: The Many Ways to Represent Time. In: International Journal of Intelligent Systems, vol. 6, nř4, 1991, p. 341-355.
3. B. Bachimont: L'intelligence artificielle comme écriture dynamique : de la raison graphique à la raison computationnelle,*Au nom du sens*. J. Petitot, Paris, Grasset (ed), 1999.
4. A. Bénel: Expression du point de vue des lecteurs dans les bibliothèques numériques spécialisées. In: Actes du Colloque International sur le Document Numérique, "Approches sémantiques sur le document numérique", La Rochelle, 22-25 juin 2004.
5. A. Bénel, S. Calabretto, A. Iacovella, JM. Pinon: Porphyry 2001: Semantics for scholarly publications retrieval. Proceedings of the thirteenth International Symposium on Methodologies for Intelligent Systems [ISMIS'2002]. Lecture Notes in Artificial Intelligence 2366, Lyon, Berlin : Springer-Verlag, 26-29 juin 2002, p.351-361.
6. S. Calabretto, G. Lallich-Boidin, F. Sèdes: Les temps du document numérique : Rapport de l'action spécifique 95, CNRS/STIC, 2004.
7. JP. Demoule: Projet de bibliographie automatique en préhistoire et protohistoire européenne. Banques de données archéologiques, Marseille, 12-14 juin 1972.
8. J. Euzenat, C. Bessière, et al.: Dossier Raisonnement Temporel et spatial. In: Bulletin de l'AFIA, volume 29, p. 2-13, Avril 1997
9. JP. Haton, N. Bouzid, F. Charpillet, MC. Haton: Le Raisonnement en Intelligence Artificielle : Modèles, techniques et architectures pour les systèmes à bases de connaissances. Inter Edition (ed), 1991, p.183-229.
10. P. Jockey: L'archéologie. Paris : Belin (ed), 1999.
11. M. Vilain, H. Kautz: Constraint Propagation Algorithms for Temporal Reasoning. In: Proceedings of the National Conference of the American Association for Artificial Intelligence (AAAI-86), 1986.
12. Guide d'indexation RAMEAU, Bibliothèque Nationale de France, 2000, ISBN 2-7177-2131-2.
13. Les thésaurus PACTOLS de FRANTIQ, 1989, http://frantiq.adbs.fr/prodserv/jetude/html/PACTOLS/PACTOLS.html.