Twin multi-lingual ontologies for heritage data localisation and retrieval

Jean-Yves Blaise[1], Iwona Dudek[1]

[1] UMR CNRS/MCC 694 MAP-gamsau
EAML 184, av. de Luminy 13288 Marseille Cedex 09 France
jyb(idu)@gamsau.map.archi.fr
http://www.map.archi.fr

Abstract. Understanding the evolution of historic artefacts, may they be individual edifices or complex urban fabrics, requires a cross-examination of various sources: specific pieces of data (what remains of the edifice itself today), and generic pieces of knowledge. As an answer, we establish a triangular relation between specific information ("what is known about object O"), generic knowledge ("what is known about objects like O", and a virtual artefact thereby acting as a data integrator ("what morphology O may have had during period P"). The paper introduces an implementation of this framework in which we have developed twin multi-lingual , multi-scale ontologies, one dedicated to terminology, one dedicated to toponymy. Sources are connected to a term (that identifies a concept) and to a toponym (that locates an instance). Our contribution details the principles behind this triangular relation, and its usability for navigation inside data sets.

Keywords: Information retrieval, visual interfaces, architectural heritage, vocabulary analysis, toponymy.

1 Introduction

Due the amount of sources on the architectural heritage, people in charge of providing an easy and attractive access to documents on the cultural heritage face a challenge to structure the information they handle with regards to what this information is about: in our case architecture. As an answer, we propose a methodological framework that lets users to organise sources in a triangular relation associating documents, toponyms and general terms. Documents are the pieces of information to structure. Toponyms and general terms are twin multi-lingual , multi-scale ontologies used to "cross-localise" them :

- toponyms localise documents in time and space;
- general terms localise documents with regards to the "theory" of architecture.

Specific information (i.e sources, "what is known about object O") is linked to generic knowledge ("what is know about objects like O") by stating that object O and objects like O "have the same name". Specific information is also positioned in time and space ("what is known about the object located in position P between dates D[1] and D[2]). Finally, because of the spatial nature of architectural objects, specific information can be accessible thanks to virtual artefacts, either in 2D or in 3D (depending on the amount of data we can gather, or simply on the "scale" of the object).

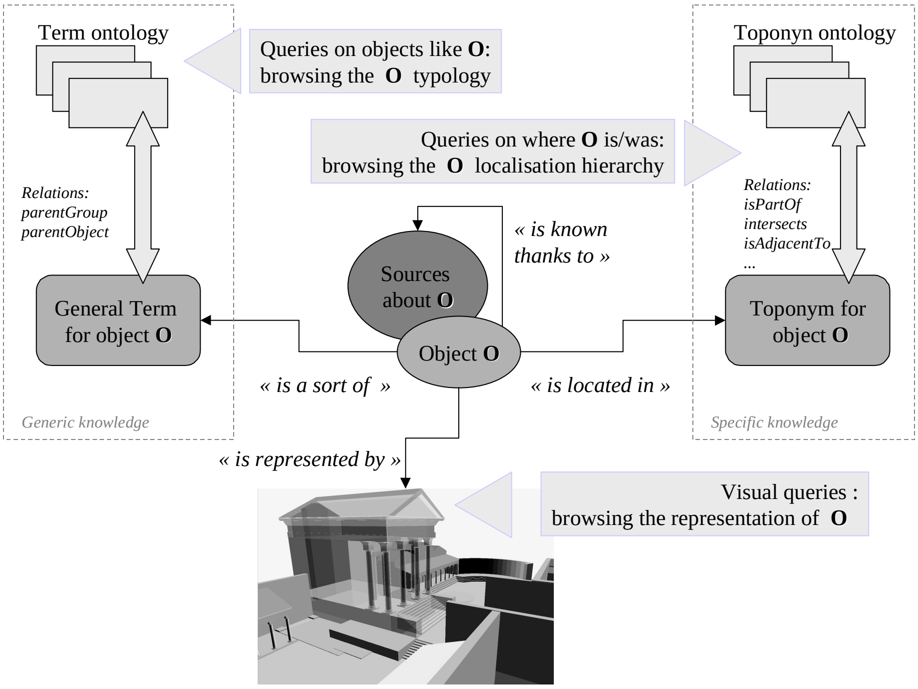

Fig 1 Putting object O in relation with a general term and a specific location. If object O is the pronaos of Douggha's roman capitol, then possible relations are:

O is a sort of Pronaos, a Pronaos' parent Group is a Temple

O is located in the toponimical hierarchy ../Douggha/Forum/Capitol

The 3D representation of O allows a visual querying of the sources about O and of the twin ontologies

The ideas and hypothesis behind this approach are the result of various previous experiments such as those reported in [1], [2], [3]. In those experiments we have investigated the *informative value* of graphics in the field of the architectural heritage. We have identified major constraints one must obey to if graphics are to become a cognitive tool in the research process about architectural evolutions (these constraints have consequences at various levels: on the modelling effort, the documentation handling, the representation itself, etc.). As a result of these experiments a methodological framework called informative modelling (see http://en.wikipedia.org/wiki/Informative_Modelling) could be proposed. The basic idea behind informative modelling is that the representation of artefacts should not necessarily claim veracity, but should support dynamic information retrieval and visualisation. In this paper a specific application, developed for the STRABON European programme, of this generic methodological framework. It has however to be stressed here that we make no claim concerning the reusability of the resulting ontologies. We only think our contribution can help better delineate the issues one has to deal with when addressing the specific question of how to use urban or architectural ontologies in the context of historic sites and therefore of evolutive data sets.

2 Context of the research

The STRABON European programme is presented as a multimedia, multilingual information system for the enhancement of the Euro-Mediterranean cultural heritage. In short, this programme is about making available and attractive on the Internet documents about the cultural heritage- (from paintings to archaeological sites) of the various partners (collection holders). Inside this programme, we have been given the task to investigate the possible benefits of using 2D or 3D web interfaces as navigation. Considering the broadness of the task, we have focused on two issues:

- A problem of representation (*i.e.* representation of the objects the documents are about, from 2D geographical maps to detailed architectural modelling).
- A problem of localisation (*i.e.* localisation of the documents themselves, localisation of what the documents show, link between localisation and documents, link between localisation and 2D or 3D graphics (SVG/VRML/Virtools™)

As an answer, we have developed a document description method in which the document (XML-formatted) is linked to two families of descriptors,

- Terms of the architectural vocabulary that relate a given document's content to conceptual definitions of objects.
- Toponyms, *i.e.* names of places as left to us by history, that relate a given document's content to a particular place, and structured as a hierarchy (matching the idea of scale).

The method can be summed up by three statements:

- For a given vocabulary item, a catalogue of "real objects" - that match the vocabulary item's definition"- is created. Each "real object" is in relation with a toponym, so that objects can be displayed within a 2D/3D representation. Therefore the 2D/3D representation can be used to query the vocabulary items, *i.e.* generic information.
- For a given toponym, a list of sources is created. Therefore the 2D/3D representation can also be used to query specific information.
- A triangular relation is created between three notions, with the visual display (2D/3D representation) exploited for navigation purposes.

In the following sections we introduce with more details the ontological structures developed in order to represent on one hand general terms and on the other hand toponyms. Both ontologies are multilingual, and the browsing supposedly visual. The label of a given term or toponym should therefore remain hidden.

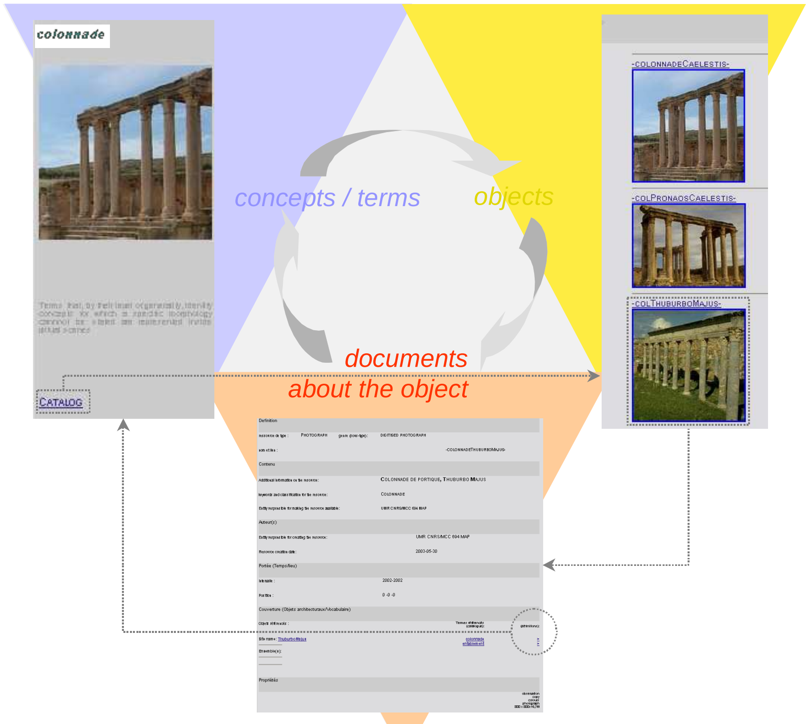

Fig. 2: Left, a triangular relation between general terms, specific locations (topoynms) and documents. Right, alternative visual displays used to filter and query either specific information or general terms associated with the toponym ontology.

3 General terms: the architectural vocabulary analysis

Terminology analysis, when applied to our field, as developed in [Pérouse De Montclos, 1988], provides a methodological framework aiming at isolating terminological items. These items fit various needs:

- terminological items are used in describing without ambiguity the architectural artefact itself.
- terminological items are used in comparative evaluations of main architectural trends.
- terminological items act as symbols to which data collections can be attached to in an effort to develop semantic-based searches inside data collections.
- terminological items identify not only individuals objects but also families: the terminology defines items that clearly range from the general to the specific.

Isolating the items that define the artefact's form is naturally crucial to its understanding (see [4]). It also opens opportunities to use the artefact's representation as an intermediary between the user and information not only on the artefact as a whole but also on each morphological item considered relevant in the architectural analysis. It is clear, however, that geometric modelling has up to now mainly given birth to still-life images of artefacts in which the semantics used do not appear (see for instance [7] or [8]). In the field of the architectural heritage, researchers and practitioners call for the emergence of modelling methods in which geometry illustrates and interfaces knowledge, as defended by [9][10].

We are dealing here with terms that identify concepts used in architecture, and we are dealing with sources that illustrate those concepts. This observation is the very heart of our approach: terms identify concepts that are materialised in this or that edifice, and sources figuring that edifice therefore illustrate the term. This is the basic statement on which we have built the DIVA experiment, a visual vocabulary tool exploited here in the above mentioned context.
We deal with two type of data sheets (formalised in XML):

1. Vocabulary items that contain a user-chosen number of definitions in various languages and a user-chosen number of translations. Both definitions and translations are linked with their bibliographic references. Each term is positioned inside two hierarchies. The “parent object” hierarchy states which term the current term refines (complicates) (example: the corinthian capital refines the term capital). The “parent group hierarchy states to which group of elements the current term can belong to (example: a capital is part of a column).
2. Sources descriptions that contain Dublin-Core based documentary elements but also a section that we entitled “architectural reference”. In this section the source is identified as an illustration of one or several terms, and as figuring one specific piece of architecture.

Vocabulary items identify a general concept (example: a capital). They are implemented as classes, each instance contains a free number of definitions and translations, each of which accompanied by the relevant bibliographical reference.

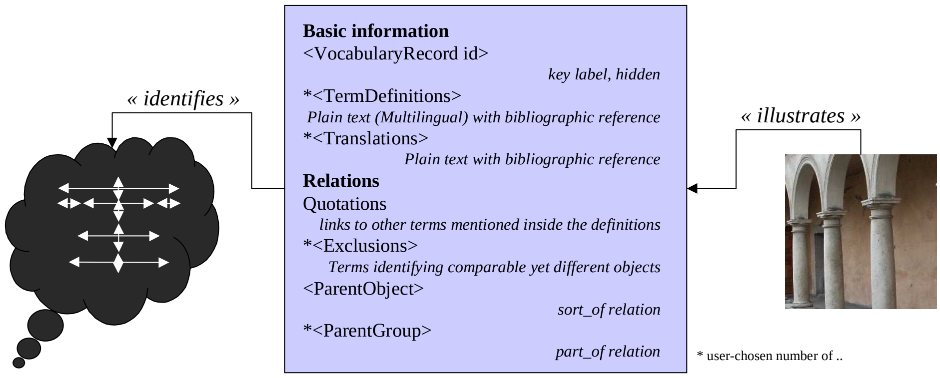

Fig. 3 Center: content of a vocabulary item sheet, and relations inside the term ontology, left, the vocabulary item identifies a concept, right, a source illustrates the vocabulary item.

4 Toponymy

Whereas toponymy is used to give names to places, we use it to *locate information*. The *architectural toponyms* we have developed are supposed to offer two complementary services :

- Further narrow and circle the information in a knowledge acquisition process where the information is marked by uncertainty, incompleteness. (for example , if one cannot attach a source to a given edifice because the location mentioned in the source is imprecise, then the source can be attached to the city where the edifice was erected).
- Segregate the information with regards to a geography-like hierarchy inside which the higher the toponym is, the more objects it is put in relation with (this is used in information retrieval phases).

We use toponyms in order to provide for a given location (may it be a whole region or an archaeological site, or even a single edifice) a unique identifier to which we can attach pieces of information. But names of places evolve through time, as well as their physical boundaries Therefore we have implemented a model of toponym that includes a series of variable features. Toponyms are formalised as classes (in the sense of OOP) which contain attributes or nested objects for each of the above mentioned features. The following list details their roles.

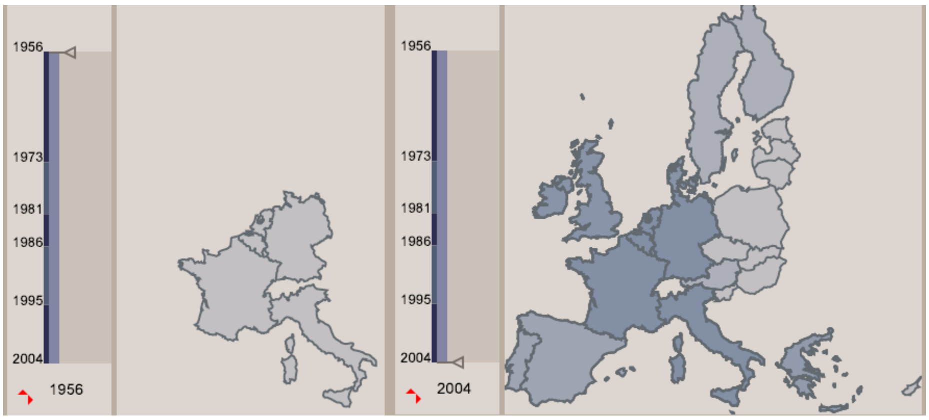

Fig. 4 The origin/contour/time formalism, toponym *europeanUnion*, representation for period 1956 and 2004.

Identity: Each toponym is given a label (name in mother tongue) added to its hierarchy path in order to act as its identity feature (ex: europe/france/rhoneAlpes/rhone/lyon/…).

Changes over time: May the town of Lyon be a French major town of the roman *Lugdunum*, it has to be for us one and only one „semantic object". The implementation we propose considers *Lugdunum* and Lyon as one and unique toponym. The name may change over time, it anyway defines one unique location.

Each toponym can contain an array called HistoricalNames that stores objects containing a label (associated with a language) and a time validity (an interval defined by two dates, called yearInterval). For instance, the toponym "…./…./rhone/lyon" would have a historical name *Lugdunum* for the period -200 to +500. Each toponym also contains an array of "historical boundaries". These are objects storing polygons and a yearInterval. Changes over time are therefore handled on two independant levels: name changes, boundary changes. If the toponym corresponds to an object that we want to represent in 3D, then the successive 3D geometries are not defined inside the toponym object but inside specific architectural classes. The toponym is linked with a "part-of" relation to the architectural classes.

Finally, an array of " timeValidities" stores the historical periods during which the use of the toponym makes sense : the toponym "Lugdunensis", part of the roman empire where *Lugdunum* was located, ceases to make sense for the localisation of toponym Lyon during the VIth century, when the western part of the roman empire ceases to exist. Yet its parent toponym, "imperiumRomanum" continues to exist until the fall of Constantinople, during the middle ages.

Multilingual aspects : The name of a place may vary at a given period depending on the language used, it anyway defines one unique location. A unique label (a simplified version of nowadays name, local language) can be created to manipulate the toponym, but it remains hidden for the user. Each toponym is given a basic name (today's name in mother tongue) and contains an array in charge of storing translated names. The number of translations is free. Moreover, an array of "alternative names" is created in order to store local variations of the name (names in regional languages for instance).

Scale issue : Toponyms are used to provide a location for human artefacts. However each toponym is placed in a hierarchy of toponyms. The top levels do not identify artefacts, but geographical concepts. A different toponym is therefore needed at each scale. The root levels of a toponym's hierarchy "place it on a map", the lower levels "draw it in 3D". Depending on the toponym's scale, its representation uses the geographical information (position, contour) or the architectural information (referential, 3D morphology exploited as contour in 2D or as volumes in 3D). In that sense, architectural toponyms act as integrators between geographical and architectural information, and between the scales corresponding to the various urban concepts our documentation mentions ("old town" -> geographical/urban information, "main square" -> architectural information). One can say that they are re-interpreted here in order to bridge the gap between GIS technologies (where the graphic sign gives few hints on the actual thing it represents [11]) and architectural investigations (where only the result, 3D model, is today given attention [12]).

Architectural / Geographical information sets : Each toponym can optionally contain pieces of quantitative information corresponding to its scale inside an an array of position/time period doublets. Position is expressed alternatively as longitude/latitude or as XYZ/αβγ; time periods are expressed as two dates (yearInterval). Besides, a reference is created to a term categorisation connecting a particular object to general knowledge (temple A -> the notion of temple, *i.e.* the vocabulary item temple)

Alternative geometries : A toponym can be represented, depending on the scale of the 2D representation, either by a contour or by a symbol. Typically, a toponym corresponding to an edifice will be represented by a dot on a map, with if needed more analytical symbols (see fig.5).

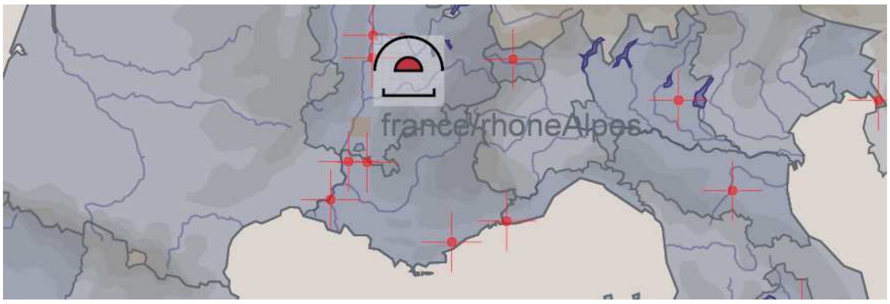

Fig. 5 Alternative representation of toponyms: the roman theatre in Lyon (Lugdunum) represented by an analytical symbol on a map featuring toponyms represented by a contour (region France/rhoneAlpes) or by a dot (other theatres).

Basic relations between toponyms: Each toponym has a relation to a hierarchy of toponyms (region/city/site/...). This relation uses the semantics of the "isPartOf" relationship, it is implemented in a very straightforward manner by an attribute inside the class. A toponym may also have a relation to toponyms that have ceased to exist. For instance Lyon is related to the roman Province Lugdunensis. An array called historicalParents stores this relation.

As mentioned above, each toponym's description includes relation to a hierarchy of toponyms (region/city/site/...). But this mechanism remains limited to the isPartOf semantics. We have in recent experiments introduced a specific "model of relation" that helps specifying for instance that "toponym t1 is adjacent to toponym t2" and in addition proposes relations between toponyms and topographic entities. In the section below we detail the possible relations.

"Lexical" relations between toponyms :

These relations are called lexical because they do not include quantitative information, but only a label for the relation. It is inspired by ideas from [13]; with here the aim of storing correspondences between instances as XML, with as many storage structures as there are tags of relations. The LexicalSpatialRelation object is a generic object with an attribute that allows the choice of a given tag such as "isAdjacentTo" or "Intersects, etc.. In addition, this object specifies a time validity (interval of dates). Such a relation could for instance state that toponym "Lugdunensis isAdjacentTo Narbonensis timeValidity1st to 6th century".

Relations between toponyms and Topographic Entities:

Whereas Toponyms provide a location for artefacts, topographic entities provide a context for these artefacts. In other word, topographic entities are "natural" elements that need to be taken into consideration in order to better understand the evolution of architectural sites. It is for instance worth mentioning that *Arelate*, *Vienna*

Allobrogium and *Lugdunum* are sites erected along the same major river, *i.e.* a supply line at the roman period. Each toponym can be associated to a set of instances of TopographicEntity. Each toponym and each Topographic entity have quantitative information used to produce SVG graphics that one can understand as the visual exploitation of the underlying toponym Ontological structure. These graphics are customized cartographic representations of toponyms, produced at query time, that include user-side choices on:

- Hierachy levels in toponym t1's tree (upwards and downwards);
- Related toponyms (exploiting part or all of the LexicalSpatialRelations instances associated to toponym t1);
- Related Topographic entities (exploiting part or all of the Relations instances associated to of toponym t1).

5 Implementation and limitations

The concepts presented in this contribution, may they concern terminology or toponymy, are implemented using the same principles. A hierarchy of classes (in the sens of OOP) produces instances of the various concepts (vocabulary items, documents descriptions, toponyms, topographic entities, relations,...). These instances produce XML sheets that are transformed (XSLT) in textual interfaces. Visual interfaces are produced at query time by methods of the above mentioned classes after parsing the XML sheets. It has to be made clear that various sets of data are manipulated, not necessarily in parallel. The vocabulary items/documents experiment started without the toponymy, the toponymy without the relations. As a consequence, a lot remains to be done to systematize links between the above mentioned elements. We think that at this stage the principles we have developed appear promising, but no other claim can be made before a deeper evaluation of the work is carried out. Each of the concepts implemented inside the two main ontologies has reached a point of stability. However we need to stress that the linking of those concepts remains a methodological concern. One has to keep in mind that our initial objective is to localise documents in a generic knowledge space (terminology) and in a specific knowledge space (toponymy). An evaluation of the method we propose therefore requires an evaluation not of the ontologies themselves (although this task also is needed) but on their usefulness in the interfacing of sources about architectural evolutions.

6 Conclusion

The methodological framework we have experimented lets users to organise documents in a triangular relation associating documents, toponyms and general terms. Toponyms and general terms play here the role of twin ontologies in charge of localising the documents with regards to specific information (a given place) and to general information (a given piece of knowledge). It is far too early to weigh the benefits of such an approach beyond the limits of the research context we have presented. Moreover, the platform remains at this stage experimental and the development of appropriate graphics requires more questioning on for example the user-friendliness of the interface (choice of level of detail, document availability visualisation; ...) and the switching between scales (2D/3D; geographical->urban->architectural, ...). Still the application of the method to real-case problems has shown that a gain of readability is at hand if documents or information one needs to deliver or retrieve are exploited using ontological structures.

Acknowledgments. STRABON programme is part of the EU's EUMEDIS Initiative.

References

1. Blaise, J.Y, Dudek, I., De Domenico, F, De Luca, L: Architectural Modelling and Information Interfacing : Learning from Three Case Studies proc. ITI 2004 (Information Technoloy Interfaces), Cavtat, Croatia, 7 - 10 VI 2004 (2004)
2. Blaise, J.Y, Dudek, I.: Using abstraction levels in the visual exploitation of a knowledge acquisition process, JUCS Journal of Universal Computer Science, proceedings International Conference on Knowledge Management - Knowledge and Information Visualisation I-Know 05 (2005)
3. Blaise, J.Y, Dudek, I.: Modélisation informationnelle : un cadre méthodologique pour représenter des connaissances évolutives spatialisables RNTI – E – 6 (Revue des Nouvelles Technologies de l'Information) Volume 1, pp 347-358, (actes 6èmes journées d'Extraction et de Gestion des Connaissances EGC 06) (2006)
4. Tajchman J Stropy drewniane w polsce. Propozyca systematyki Ośrodek dokumentacji zabytków, Warszawa (1989)
5. Pérouse De Montclos JM : Architecture vocabulaire - Principe d'analyse scientifique, Imprimerie Nationale (1972-88)
6. S. Pajares Tosca A pragmatics of links, Journal of Digital Information, volume 1 issue 6, (2000)
7. Perkins, A. 2003 The Cone Sisters' Apartments: Creating A Real-Time, Interactive Virtual Tour Proceedings ICHIM 03 Paris France (2003)
8. Von Lehn, D, Heath,C :Displacing the object: mobile technologies and interpretive resources Proceedings ICHIM 03 Paris France (2003)
9. Alkhoven, P. : The changing image of the city. A study of the transformation of the townscape using Computer assisted Design and visualisation techniques PHD Thesis, University of Utrecht (1993)
10. Kantner, J : Realism vs Reality: creating virtual reconstructions of prehistoric architecture, in J.A Barcelo, M.Forte, D.H Sanders (Ed.) Virtual reality in archaeology, (Oxford: Archeopress 2000) (2000)
11. Sebillo,M. : A Web-GIS for promoting archaeological assets Proceedings ICHIM 03 Paris France (2003)
12. Huber, B : Adding dimensions to GIS with VRML, issues Proceedings of Pennsylvania GIS Conf., Hershey, USA, (2000)
13. Zeitouni, K., Yeh, L, Aufaure, M.A., Join indices as a tool for spatial data mining, Int. Workshop on Temporal, Spatial and Spatio-Temporal Data Mining,, In Ain2007, Springer, 102-114, Lyon, France, Sept 12-16 2000 (2000)