ArchaeoKM: Managing Archaeological data through Archaeological Knowledge

Ashish Karmacharya, Christophe Cruz, Frank Boochs, Franck Marzani

ArchaeoKM: Managing Archaeological data through Archaeological Knowledge

A. Karmacharya[1,2], C. Cruz[2], F. Boochs[1], F. Marzani[2]

[1] Institut i3mainz, am Fachbereich 1 - Geoinformatik und Vermessung
Fachhochschule Mainz, Holzstrasse 36, 55116 Mainz, Germany
{ashish, boochs}@geoinform.fh-mainz.de
[2] Laboratoire Le2i, UFR Sciences et Techniques, Université de Bourgogne
B.P. 47870, 21078 Dijon Cedex, France
{christophe.cruz, franck.marzani}@u-bourgogne.fr

1. Introduction

Nowadays, the use of ontologies in the field of the archaeology is a new direction of research that is not fully tested. Some results are very encouraging and challenge the vision of archaeological sites which cannot be modeled semantically. Most of the researches based on ontology technologies in the archaeology domain are based on the research of finding on a common ground for the interoperability and the integration of data. Papers like (LANG 2009), (KOLLIAS 2008) cover these research. The project ArchaeoKM is a shift from such researches. It focuses on using the knowledge possessed by the archaeologists to model the data of an archaeological project, and to define domain rules in order to enrich the knowledge contained in the modeled data system. This system aims at the definition of a comprehensive archaeological knowledge base system. The initial concept of the application has already been presented in CAA09 (KARMACHARYA, et al. 2009).

As it has been discussed, the ArchaeoKM platform is based on Semantic Web technologies and a domain ontology defined by the specialists of the domain in collaboration with specialists of ontologies. Hence, it deals with knowledge generation and knowledge management through the expertise of the archaeologists. These knowledge could be both semantics as well as spatial. In addition, the ArchaeoKM platform is capable of handling both the knowledge as a single solution. This extended abstract paper presents the final concept and the process of developments since.

The paper is organized in the following manner: section 2 covers the background behind the platform. Section 3 discusses different components within the ArchaeoKM platform. Finally, section 4 concludes by covering the current status of the platform.

2. Background

The main concept behind ArchaeoKM is to use knowledge that archaeologists possess to manage data from the excavated site. With the advancement of data capture technologies, we get widely more precise information. In contrast to the advantage of data interpretations and analyses, we have to deals with the issue of the data management. This problem is even more inconvenient in the case of industrial archaeology where the archaeological sites are available for very limited amount of time. All that can be collected during that time frame should be managed later. After what, the knowledge of experts is exploited to manage the data. The ArchaeoKM platform facilitates the work of archaeologists concerning the information management. This is done using the archaeologists knowledge even long after the data is collected. To do so, the geo-localization of the excavated objects of the site is defined into the system. Then, the semantic definition of the finding is added into the system through the use of a common definition materialized by an ontology of domain. This step is called this enrichment process which consists to enrich the knowledge already presents in the system. From this point, the system is able to take advantage of the semantic definition in order to semantically annotate the identified objects that can be found in the corpus of documents. Archaeologists benefit from this system because the dataset can be mapped to the relevant objects irrespective to their structures and file formats. In addition, it provides a semantic view on datasets materialized by the shared ontology of domain allowing archaeologists to build a global schema between data sources.

2.1 Demonstration Site and Data Pattern

The project ArchaeoKM uses the Industrial Archaeology context as a case study to demonstrate the concept. Actually, the principle of ArchaeoKM could be applied to other branches of archaeology. The demonstration site is the site of Krupp in Essen belt, Germany. The 200 hectares area of the Krupp site was used for the production of steel during early 19th century and was destroyed during Second World War. Most of the area is never rebuilt making it an ideal site for industrial archaeological excavation. The problem is that the area would be used as a park for the ThyssenKrupp main building within this year. So time is running out for the data collection process. In addition to the limited time frame, the amount and the pattern of data collected are very huge and diverse. The dataset and documents are ranged from simple images taken by non calibrated digital camera to highly sophisticated ortho-rectified aerial images. The geometries of the excavated objects are represented through point clouds obtained with the help of various terrestrial laser scanners. There are other types of documents such as archives and site plans. All of these data and documents are collected and stored in a repository using their own format without structuring due to the time constraint. Now, The ArchaeoKM platform should facilitate archaeologists to map these data and documents to the respective objects through semantic annotations.

2.2 The Architecture

Figure 1 shows the platform architecture of the system ArchaeoKM. As it could be seen the system is composed of three distinct but interrelated levels. The lowest level is the Syntactic level. Within this level, all the data and the documents are stored in their native format. The next level is the Semantic level. This level is the heart of ArchaeoKM as it represents the actual excavation site through different components of Description Logics (DLs). Archaeologists use their knowledge to build the description of the excavation site, and on this description the ontology of domain and its components are defined. The components are the categories of objects, theirs attributes and theirs semantic relationship which could be for instance transitive. The ontology is expressed using the Ontology Web Language OWL (McGUINNESS und HARMELEN 2004). The highest level is the Knowledge level. It is the face of the platform because it provides to the user a graphical interface for the user interactions. Additionally, this level provides interfaces for the knowledge visualization through web interfaces based on semantic wiki. Besides these three levels, a parallel structure to facilitate them in their spatial operations could be seen in the system architecture. The details on how the facilitator facilitates the different levels would be discussed in next section.

3. Process in ArchaeoKM

The ArchaeoKM platform considers two processes which are the Knowledge Generation and the Knowledge Management. Both semantic and spatial components contribute to the both processes for proper knowledge manipulation. This section discusses these processes in terms of semantic and spatial components.

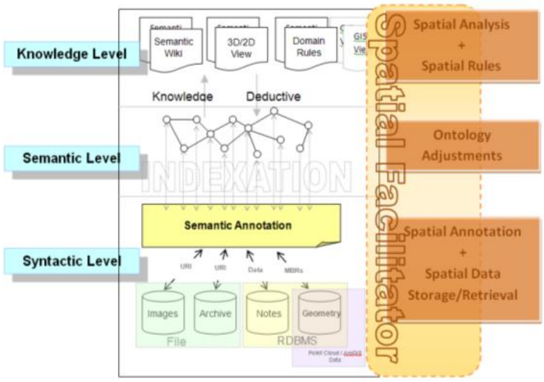

Figure 1: The ArchaeoKM *platform architecture*

3.1 Knowledge Generation

The schema of the ontology which describes the categories of objects is the base from which knowledge is generated within the platform ArchaeoKM. At the beginning, this schema is a collection of DLs axioms which describes the kind of features of the excavation site. The semantic modeling through the domain ontology within the ArchaeoKM platform is shown in figure 2.

Figure 2: *Semantic model through Domain Ontology*

As it could be seen, the ontology is a hierarchical structure of the object categories that could be possibly excavated. The relationships among them do exist but they are managed and visualized through other interfaces. Archaeologists are engaged in the process of enrichment of the ontology by defining new categories or by defining more specific categories. In addition, archaeologists are engaged as well in the ontology population process which consists to map the identified objects to a category. It should be noted that the categories and the relationship which define the semantic model is not yet complete. Efforts are being made to make this ontology as complete as possible in collaboration with the archaeologists working in this area. Concerning the population process, there is three steps, the identification of objects, the semantic annotations process, and the spatial annotation process.

Identification of Objects: the ArchaeoKM platform provides a graphical interface that allows the archaeologists to identify and to tag objects. The tagging process is undertaken by creating bounding boxes in a specified area of the site orthophoto. This could be seen in figure 3.

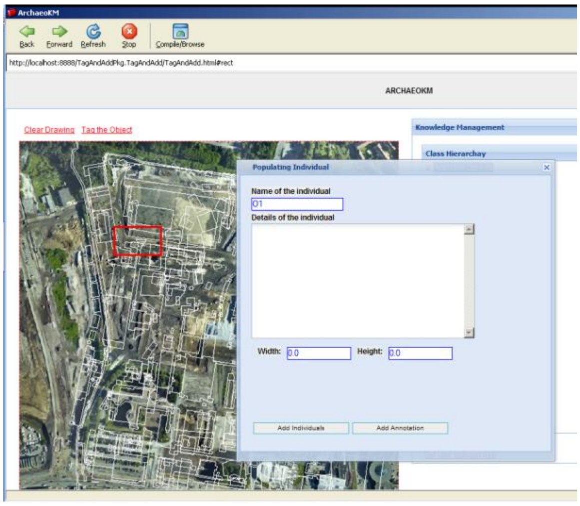

Figure 3: *The tagging process of an orthophoto*

When the objects are added in the domain ontology, it becomes a knowledge base. Now the object has been tagged and added to the ontology, the semantic information is added. This is done by providing detailed semantic knowledge to it like what category it belongs to, what other objects it is related to and through what relationships it is related to them, what are the properties it carries and many others. The ArchaeoKM facilitates the input of information for archaeologists. The bounding boxes, through which the objects are geo-located, are stored in ontology for future visualizations. Actually, this information is also stored in a spatial database system in order to make it possible to use spatial function on semantic object.

Semantic Annotations: An identified object could be semantically annotated and linked to relevant documents in order to provide a semantic index for documents concerning the object. It has been mentioned that huge and diverse documents are collected during the excavation process and are stored in their native formats. The ArchaeoKM platform provides interfaces that allow the indexation and the retrieval of these documents through the semantic definition of objects.

Spatial Annotations: Point Clouds are the base of geometrical information of the object in the ArchaeoKM platform. Thus, it uses PostgreSQL to store the point clouds. The points are stored as simple attributive data within the database with proper indexing system in order to their efficient retrieval. Archaeologists are allowed to feed the spatial annotation through either manually entering coordinates of Minimum Bounding Rectangle (MBR) of the object or choose the coordinates while locating the object as the MBR. In either case, ArchaeoKM uses conventional SQL to query the point cloud. It then extracts and stores the points in ASCII format so that it could be visualized through any visualizing tools. Additionally, the coordinates of MBR is stored as spatial data type within PostGIS – a spatial extension of PostgreSQL for future spatial functions and operations.

3.2 Knowledge Management

The knowledge that was generated through the identification process is managed through different functionalities within the ArchaeoKM platform. This section discusses these functionalities.

Semantic and Spatial Knowledge: As already discussed, the semantic knowledge are fed through semantic relationships with other objects during the identification process. However spatial information are fed through the spatial operations and functions provided by the database system. The domain ontology is adjusted in order to permit spatial operations and functions provided by the database system on the object defined in the domain ontology. The spatial functions are divided into two broader categories, Spatial Processing functions and Spatial Relationship functions. The former return the geometries when they are executed while the later are binary functions. A set of concepts are introduced in the ontology with proper relationships to the excavated objects for the first set of functions. As for the second case, the spatial functions are represented through the relationships within the ontology. Spatial functions are performed in the database level but the results are added to the ontology. In this way the spatial analysis is possible within the application for proper spatial knowledge management.

Domain Rules and Analysis: The introduction of horn logic as base in logic programming has allowed the description of knowledge with predicates. Extensional knowledge is expressed as facts, while the intentional knowledge is defined through rules (SPACCAPIETRA, et al. 2004). These rules are defined through different Rule languages to enhance the knowledge in the ontology. The ArchaeoKM platform takes the advantage of such languages and allows the archaeologists to create their own rule in order to validate or discover knowledge. This is particularly helpful for difficult findings which cannot be immediately classified as anything but posses certain features. Later through some analysis a rule could be formulated to classify them. An example would be *an unknown object with red bricks and round structure near the chimney is an oven*. ArchaeoKM facilitates archaeologists to create such rules through its interface and classify the objects. It is shown in figure 4 where a simple rule of *site having oven with brown wall is a gluehaus* is created which could be seen at the bottom of the interface.

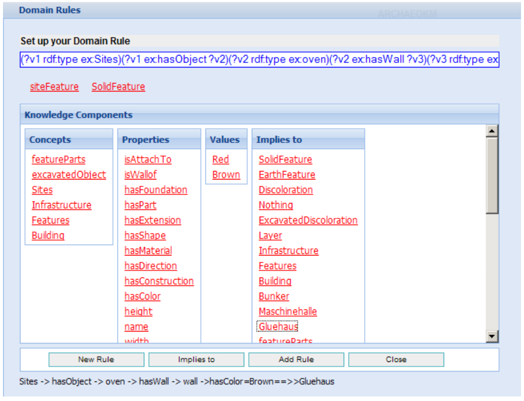

Figure 4: *Interface for domain rule creation.*

Interestingly, the rule can be composed of semantic axioms or spatial axioms or a combination of both. The extension of spatial rules is an added feature to the rule language itself provided by the ArchaeoKM project. Now the above example would be used as *an unknown object with red bricks and round structure and within 50 meters for a chimney is an oven*. In this examples both semantic rule such as *red bricks* and *round structure* and spatial rule *within 50 meters* are applied to classify the unknown object.

The semantic reasoner, reasoning engine, rules engine, or simply the reasoner within the OWL language allows the ArchaeoKM platform to infer new knowledge. This is used within the application to for knowledge analysis. The analysis tool in the application can reason the ontology to extract the new knowledge. Query languages like the SPARQL language are used for the purpose but it is the reasoning capacity that enhances the returned results with new knowledge.

4. Conclusion

It can be noticed that concept behind ArchaeoKM is relatively new. The involvement of archaeologists in each and every step within the system helps the knowledge within the system grow rapidly which will benefit archaeologists.

Currently, ArchaeoKM is under development. The first version of prototype is developed and in the process of testing. This version incorporates complete semantic component and few spatial components. However, the complete integration of spatial component is underway and should be ready soon. Once, the whole package is ready for the testing, we believe the application will be cutting edge application in the field of archaeology.

In addition to assist archaeologists in the data management ArchaeoKM provides a new dimension to spatial manupulation of spatial data. The concept could be integrated within current GIS applications to add the dimension of knowledge in their existing systems

References

KARMACHARYA, A., C. CRUZ, F. BOOCHS, and F. MARZANI., 2009. ArchaeoKM: Toward a Better Archaeological Spatial Data Sets Management. *37th CAA2009 - Computer Application and Quantitative Methods in Archaeology*. (Abstract: Williamsburg: CAA 2009) (Williamsburg, 2009). pp:227.

KOLLIAS, S., 2008. Achieving Semantic Interoperability in Europeana. *Digital Heritage in the New Knowledge Environment: Shared spaces and open paths to cultural content.* Athens: Hellenic Ministry of Culture Directorate of the National Archive of Monuments, (Nov. 2008), (Proc. Intl Conf in Athens), pp:143-145.

LANG M., 2009. *ArcheoInf - Allocation of Archaeological Primary Data.* 2009. http://www.archeoinf.de/archeoinf/archeoinf-english (Last Visited: 18. January 2010).

McGUINNESS, D L., and HARMELEN, F, van., 2004. *OWL Web Ontology Language.* 10. February 2004. http://www.w3.org/TR/owl-features/ (Last Visited: 20. January 2010).

SPACCAPIETRA, S., N. CULLOT, C. PARENT, C. VANGENOT., 2004: On Spatial Ontologies. *GEOINFO.* Campos do Jordao, 2004.