Industrial Archaeology: Case study of Knowledge Management for Spatial Data of Findings

Ashish Karmacharya[1,2], Christophe Cruz[2], Franck Marzani[2], Frank Boochs[1]

[1] Institut i3mainz, am Fachbereich 1 - Geoinformatik und Vermessung
Fachhochschule Mainz, Holzstrasse 36, 55116 Mainz
{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

Abstract. Shifting from conventional approaches to an unusual approach in industrial archaeology, we suggest the use of a web platform based on semantic web technologies and knowledge management. This platform is used to store data during the excavation process and to manage knowledge acquired during the identification process of the findings. The principle of our approach consists in using semantic annotations in order to have a semantic view on data sets. The shared ontology that defines an index on the semantic annotations allows us to build a global schema between data sources. This global schema allows annotating, indexing, searching and retrieving data and documents.

Keywords: Industrial archaeology, knowledge management, information system, ontology, owl, spatial data

1 Introduction

Oriented data management is widely used in archaeological projects to store and retrieve data generated during the excavation process. Today, with the rapid growth of advance technologies it is possible to generate huge amount of data in short time. Hence, it has become problematic to manage data with the conventional methods. This has prompted huge researches in the field of data indexation and information retrieval in order to reach a better level of data management. The level consists of identifying and managing knowledge from the data collected during excavation. Today, as different technologies are being used during excavation, different pattern of data are generated. Primary source of data in any excavation site is the set of point clouds obtained through the terrestrial laser scanning process. They are generally used for creating 3D object models. Besides, floor plans, images and other data like archaeological notes are collected during the project. They provide great value in analysis of the archaeological findings in any project. The process of identifying and storing data from findings is a process of knowledge capitalization on archeological sites. Industrial archeology generates a huge amount of data in a very short time due to the fact that industrial archeological sites are available only for a very short time.

Thus, the collected data is stored during the process in a repository without any relevant structure. Once data are stored, the process of identification of industrial findings with the help of the data repository is carried out. Two main issues need to be addressed here – first about the data structure for efficient access of data and second – the process allowing archaeologists to have efficient retrieval of the findings from the above repository. Shifting from conventional approaches, we suggest the use of web platform through semantic web technologies and knowledge management. The platform is used in both storing data during excavation process and in managing knowledge acquired during identification process. The collaborative process between archeologists is facilitated by the platform in order to generate knowledge from the data set once the data are stored in relevant data structure. The principle of the approach is to use semantic annotation to have semantic view on the data sets. The shared ontology that defines an index on the semantic annotations allows us to build the global schema between the data source. This global schema allows annotating, index, searching and retrieving data and documents.

2 Data and Knowledge Management

There has been many research works in the field of 3D Object Modeling but most of them focus on some specialized area and do not cover the whole. However, projects like 3D MURALE [1] and DILAS [2] attempts to take other factors into account making them most comprehensive. 3D MURALE system is composed of a recording component, a reconstruction component, a visualization component and database components. The findings are managed through a database management system. Once the findings are stored in the database with a proper data structure, the objects are reconstructed through the reconstruction component. This is done by modeling the objects in 3D space. These 3D models are displayed in the visualization component. DILAS is a generic, fully object oriented model for 3D geo-objects. The 3D geometry model is based on a topologically boundary representation and supports most basic geometry types. It incorporates also the concept of multiple levels of detail (LOD) [3] as well as texture information. As most of the research works are geometry management oriented, they lack semantic information. Actually, semantic information allows the management of knowledge on geometrical objects. An interesting approach on how to represent an object through the semantic information in a 3D scene has been discussed in [8]. The use of spatial and orientation relationships between objects can represent the objects in an adequate manner with respect to its surrounding.

Knowledge about documents has traditionally been managed through the use of metadata. The Web semantic proposes to annotate the document content using semantic information from domain ontologies [4]. The result is a set of Web pages interpretable by machine with the help of mark-ups. The goal is to create annotations (manually or automatically) with well-defined semantics. In the Semantic Web context, the content of a document can be described and annotated using RDF and OWL. Semantic Web annotation brings benefits of two kinds to this platform - enhanced information retrieval and improved interoperability. Information retrieval is

improved by the ability to perform searches, which exploit the ontology in order to make inferences about data from heterogeneous resources [7].
Our platform aims at not only managing the concepts defined to annotate documents (which most of the research projects currently focusing on), but also the instances of concepts with their own property values. In this manner, an object found in a point cloud can be linked, with the help of an instance in the ontology to other documents that contain the same object. The second aim of our platform is to give archaeologists the possibility to manage Wikipedia pages on findings. These Wikipedia pages represent the knowledge formalized by archaeologists and are managed through a 3D scene where 3D objects are linked to Wikipedia pages.

3 Data pattern and format

The case study site is the Krupp factory in Essen, Germany. The 200 hectares area was used for steel production during early 19th century and was destroyed in Second World War. Most of the area has never been rebuilt and thus provides an ideal site for industrial archaeological excavation. The area will be used as a park of the ThyssenKrupp main building in 2010. Actually, we are running out of time to collect data. The first challenge consists in creating a relevant data structure which helps in retrieving those data efficiently. In addition, the data which have to be collected are huge so the system should be able to handle a huge data set. The nature of the dataset generated during the project contain heterogeneous. They range from scanned point cloud from terrestrial laser scanners to the floor plans of old archive. The primary source of geometric information is provided through the point cloud. The point clouds have resolutions of 0.036 degree and are in Gauss Krüger coordinate system (GK II). It is the main data set used for the 3D object modeling. Beside point clouds, huge amount of images are also collected during the excavation. Most of the images are taken with non calibrated digital camera so do not contain any information about the referencing system. Those images posses vital semantic information and could be used for the formulation of knowledge. However, there were photogrammetric flights to acquire aerial images of the area. The aerial images were processed to generate a digital orthophoto with a resolution of 10 cm and is again in Gauss Krüger referencing system (GK II). Additionally, huge archive data have been collected. Those data contains floor plans of the buildings and other semantic information. Likewise, the notes taken by archaeologists are also important to acquire semantic information of the findings. ArcGIS databases are also available depending on the site and its nature. These databases are in the GK II reference system. For our example, this database gives an overview of the site and can be overlayed with the orthophoto in order to identify the interesting locations easily.

4 Principle and method

The scenario of the case study consists of three steps. The first step is to annotate semantic information in the excavation data to identify the findings in the document.

Then during the second step a relationship should be formulated between documents of same industrial finding. The third step consists in managing semantic objects in order to manage the knowledge with the help of Wikipedia pages. First of all it is necessary to consider the storing structure of the repository and the services that will be available to store and search data on the various data sets. Geometric and semantic relationships between various objects should be taken into account for efficient management of the objects. The simplest approach would be to store the objects with respect to a 2D map through the bounding boxes. The images of those objects taken from different view points are then related to the respective objects' bounding box by referencing them against the map. Similarly, the points of view of those images are referenced to their respective points in the map. The theory is similar to the scanned point clouds. The geometries of the objects are stored in the database and linked them through the bounding boxes with the 2D map. A similar process is also applied to other datasets. Every datasets are transformed in a common referencing system with respect to the referencing system of the 2D map. Thus, all the datasets are linked through a common referencing system and becomes easier to extract information. The second step should require archaeologists to annotate the documents indexed in 2D map and identify the common archaeological findings in order to create knowledge. It is very important to involve archaeologists in this step as they are the best person to identify the findings. They are the one who should determine the rules through those annotations to generate the knowledge. These rules between the objects will help to enrich the knowledge base and should be incorporated within the ontology. Thus, the ontology will help to create a relationship between the documents. The ontology and the instances of the ontology classes will be defined by archaeologists. In addition, they will also define industrial objects in relation to the documents indexed in the 2D map. The last step, the findings during the excavation should be managed properly with the knowledge discovered in the archaeological site. Ontology plays a major part in achieving it. All the findings are referenced against the 2D map through the bounding boxes as semantic objects in the ontology. This could be roughly termed as semantic mapping and is of great value to the archaeological process to determine different behaviors of object in different scenarios. Additionally, those semantic annotations can be interpreted by the machine to be shared, published, queried or used in more general way.

The architecture of the proposed system consists of three levels. The first level is the Syntactic level and about the indexation of all data and documents in a 2D map. It is composed with RDBMS (Relational Data Base Management System) that allows us to store geometrical data. Database systems like Oracle 11g or PostgreSQL have their own spatial extension to store the geometric information and can perform spatial operations. The second level represents the semantic index composed of semantic annotations. This level defines the nature of data and documents and defines relationships between semantic objects. This level is called the ontological level and represents a bridge between interpretative semantics in which users interpret terms and operational semantics in which computers handle symbols [5]. The last level is the highest level and the most concrete one which represents the organization of the knowledge on the semantic map. Our platform is close to the semantic extension of Wikipedia [6], but data handling and managing extends beyond textual data. The platform will guide archeologist in order to define Wikipedia pages concerning subjects and objects of the site that represent knowledge added by archeologist. This level is called the knowledge level because it represents the specification of the knowledge of archeologists concerning the industrial findings.

5 Conclusion

We have presented a platform based on knowledge management which is used to handle archaeological data. We are currently prototyping our architecture using JENA on PostgreSQL. The process works on computers in a local network. To implement the framework, we are using JENA (Semantic Web Framework for Java) [9] in order to build and to manage ontologies in JAVA. JENA helps us to handle an OWL database. We use the request language of JENA to retrieve data. What was not presented here is the collaborative process between archaeologists. The next issue to resolve is the collaborative work on the ontology which will enable all archaeologists to work on the same Wikipedia page.

References

1. Cosmas, J., Itagaki, T., Green, D., Grabczewski, E., Waelkens, M., Degeest, R.,: 3D MURALE: A Multimedia System for Archaeology, Proc. ACM Virtual Reality, Archaeology and Cultural Heritage (2001)
2. Wüst T., Nebiker S. Landolt R., Applying the 3D GIS DILAS to Archaeology and Cultural Heritage Projects-Requirements and First Results, Basel University of Applied Sciences, Muttenz, Switzerland (2004)
3. Balletti, C., Guerra, F., Adami, A,. 3D Multiresolution Representations in Archaeological Sites, CIPA 2005 XX International Symposium,Torino, Italy, (2005)
4. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web, Scientific American (2001)
5. Guarino N., The ontological level, in R. Casati B. S. & White G., eds, Philosophy and the cognitive sciences, Hölder-Pichler-Tempsky (1994)
6. Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., and Studer, R.: Semantic Wikipedia. In Proceedings of the 15th international Conference on World Wide Web (Edinburgh, Scotland, 2006). WWW '06. ACM, New York, NY, 585-594 (2006)
7. Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., Ciravegna F.: Semantic annotation for knowledge management: Requirements and a survey of the state of the art, Journal of the Web Semantics: Sciences and Agents on the World Wide Web 4, Elsevier,14-18 (2006)
8. Web Ontology Language (OWL), http://www.w3.org/2004/OWL/
9. McCarthy, P.,: Introduction to Jena, [Online] Available: http://www-128.ibm.com/developerworks/java/library/j-jena/ (2004)