

Integration of Spatial Technologies and Semantic Web Technologies for Industrial Archaeology

Ashish Karmacharya^{1,2}, Christophe Cruz², Frank Boochs¹, Franck Marzani²

¹ *Institut i3mainz, am Fachbereich 1 - Geoinformatik und Vermessung
Fachhochschule Mainz, Holzstrasse 36, 55116 Mainz
{ashish, boochs}@geoinform.fh-mainz.de*

² *Laboratoire Le2i, UMR-5158 CNRS,
UFR Sciences et Techniques, Université de Bourgogne
B.P. 47870, 21078 Dijon Cedex, France
{christophe.cruz, franck.marzani}@u-bourgogne.fr*

Keywords: OWL, SWRL, spatial analysis, knowledge management, ontology, semantic web, industrial archaeology

Abstract: We propose a method that uses the advancement in spatial technologies from current database systems within the Semantic Web Technologies in order to enrich and to populate the knowledge of a domain defined in an OWL-DL ontology. The results of spatial operations and functions are used to populate and to enrich ontologies with new individuals and new relationships. The advantage of spatial analysis within Semantic Web technologies is the diversity of the functionalities provided by the combination of spatial operations and the rule language of the Semantic Web (SWRL). This method is applied in the industrial archaeology domain in order to enhance the knowledge management.

1 INTRODUCTION

Geometry has always been the dominant component in any system related to an archaeological project. The objects extracted on the excavation sites are represented by using their geometries. This fact has led to the assumption that a system related to such projects is either a 3D object modeling system or Geographic Information System (GIS), as they both use object geometries and their relations with their surroundings. However, in the whole process the semantics of the geometric objects and their relationships with the surroundings are neglected. With the advancement of survey technologies, data can be collected more accurately. On the one hand, this has brought a great advantage in the analysis process as we possess more and diverse data to perform the precise analysis. On the other hand, it has created difficulties in managing them with existing database systems due to their size and diversity. This issue is even more visible in an industrial archaeology project. Indeed, the sites of excavations are available for a very limited time only and thus the data have to be collected and

stored in a very short time. In addition, the diversity of the data makes the management of the information with the existing database systems more complex. Hence, a lot of research is done in the field of data indexation and information retrieval in order to reach the level where this vast amount of information can be managed through the knowledge defined by the archaeologists. Actually, the knowledge about the objects excavated from the sites can only be defined by the archaeologists.

Consequently we propose a method which is adjusting the old methods while, at the same time, taking advantage of the emerging cutting edge technology. We propose in our method to retain the storing mechanism with the existing database management systems and to consider geometry as one of the major data types. In addition we suggest the use of a collaborative web platform based on semantic web technologies and knowledge management so that the information can be handled by several archaeologists and technicians. The platform will allow to store data during the excavation and to manage it through the knowledge acquired during the identification process. Furthermore, it facilitates the collaborative process

between the archaeologists concerning the generation of knowledge from the data sets. The main principle of our approach is the use of semantic annotation to provide a semantic view on the data sets. The shared ontology that defines an index on the semantic annotations allows us to build a global schema between the data sources. This global schema allows us to annotate, index, search and retrieve data and documents.

The semantic tool is being used in a wide range of applications ranging from data integrations to knowledge management. Given that, this is a relatively new topic, so great amount of researches have been conducted on the different aspects of this technology. However, most of the research hardly includes spatial information and if they do they are primarily focusing on spatial data integration with semantic technologies (Green, 2008). The ArchaeoKM (Karmacharya, 2008) project aims at the inclusion of the spatial data process within Semantic Web technologies in order to not only establish a comprehensive data integration process between spatial data but to also combine the benefits of spatial operations with the deductive reasoning capabilities of OWL DL ontologies for a comprehensive knowledge management. The benefit of spatial analysis within Semantic Web technologies lies in the diversity of the functionalities provided by the combination of the spatial operation and the rule language of the Semantic Web.

In the following section, we will discuss the technical background of the project. In section 3 we will introduce the Web platform ArchaeoKM. Section 4 focuses on the spatial facilitator. It explains the spatial integration of functions and operations concerning the enrichment of ontologies, as well as the SWRL extension. The last section concludes the paper.

2 BACKGROUND

The sharing of knowledge in archaeology and its dissemination to the general public through wiki has been discussed in (Costa, 2008). Likewise the use of knowledge to build up a common semantic framework has been discussed in (Kansa, 2008). Research works exist in the field of archaeology, but most of the research is carried out in other related fields. The existing research focusses more on the use of a common language for efficient interoperability. The research project in (Kollias, 2008) concerns the achievement of syntactic and

semantic interoperability through ontologies and the RDF framework in order to build a common standard. Data integration through ontologies and their relationships are discussed in (Doerr, 2008). Although the work on semantic web and knowledge management in the field of Information systems in archaeology or related fields has made progress with these research works, it remains a fact that they are in a very preliminary phase today. In addition, these projects concentrate more on how to achieve interoperability with semantic frameworks and ontologies. However, no research focuses on the knowledge generation process and more specifically on rules defined by archaeologists in order to build up the system which will use, evaluate and represent the knowledge of the archaeologists.

Industrial Archaeology is perhaps the best suited field in archaeology on which to carry out our research. Actually, Industrial Archaeological Sites (IAS) are available for a very short time only. The limited time available for the storage of the data is one of the concerns we want to address here. Moreover, the amount of data that has to be collected in this short span is very large and diverse. The ArchaeoKM project focuses their attention on the site of the Krupp factory in Essen, Germany. The 200 hectares area was used for steel production during the early nineteenth century and was destroyed in the 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 as far as the collection of the data is concerned. The first challenge consists in creating a relevant data structure which helps to retrieve those data efficiently. In addition, the amount of data that has to be collected is huge, so the system has to be able to handle a huge data set.

The nature of the data set generated during the project is heterogeneous. As it can be seen, the acquired data ranges from scanned point clouds from terrestrial laser scanners to the floor plans of old archives. The primary source of geometric information is provided through a point cloud. The point clouds have a resolution of 0.036 degrees and are in the Gauss Krüger coordinate system, zone II (GK II). This is the main data set used for the 3D object modeling. Beside the point clouds, a great amount of images are also collected during the excavation. Most of the images are taken with a non calibrated digital camera and, consequently, do not contain any information about the referencing system. Even though they do not contain any

referencing information, they possess vital semantic information and can be used for the formulation of knowledge. However, photogrammetric flights take place to acquire aerial images of the area. The aerial images are processed to generate a digital orthophoto with a resolution of 10 cm. The digital orthophoto is again in the Gauss Krüger referencing system (GK II). Furthermore, a huge amount of archive data are collected. These data contain floor plans, old pictures 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 overlaid with the orthophoto in order to identify the interesting locations easily.

3 THE ARCHAEOKM PLATFORM

ArchaeoKM is a Web platform which takes into consideration an adjustment of the old methods and, at the same time, takes advantage of the emerging cutting edge technology. The system still proposes to retain the storing mechanism with the existing database management systems and to consider geometry as one of the major data types. In addition, we suggest the use of a collaborative Web platform based on semantic web technologies (OWL, RDF, SPARQL, SWRL) and knowledge management in order to handle the information provided by several archaeologists and technicians. ArchaeoKM includes deductive rules defined by archaeologists on data of excavated objects. The knowledge is stored in a machine-readable format. Consequently, the knowledge can be translated into a human-readable format.

The Web based system ArchaeoKM has an architecture divided into three major levels. Each level has its own distinct functionality and is interdependent with the others. The syntactic level stores all the information that is excavated on the site. As discussed earlier, information is either stored in file formats like images or archive data or stored in the Relational Database Management System like archaeological notes or scanned/GIS data.

The semantic level allows the management of generated knowledge. It is achieved through the ontological structure set up by archaeologists. Archaeologists are involved actively in this phase as

they are the ones best suited to provide entities and their relationships needed to build up the domain ontology. This level represents a bridge between interpretative semantics in which users interpret terms and operational semantics in which computers handle symbols (Guarino, 1994).

The knowledge level represents the specification of the knowledge of archaeologists concerning the industrial findings. This level provides the user with a graphical interface represented by Web pages in order to display the generated knowledge. The pages are interrelated and can be navigated according to their relevancy.

Besides these three levels, the system architecture contains a component that facilitates the knowledge generation, update and validation through a spatial perspective. This component called the “spatial facilitator” is in charge of the spatial data analyses and provides thus the result in order to enrich and to populate the ontology. The ontology population process is the activity of adding new instances to an ontology. The ontology enrichment is the activity of extending an ontology by adding new elements (e.g. concepts, relations, properties, axioms) (Castano, 2007). The details of the component are given in the next section.

4 THE SPATIAL FACILITATOR

This section highlights our approach to the management of the spatial operations in order to enrich and to populate our ontology. The ontology schema of the ArchaeoKM platform is responsible for maintaining a relation between the enrichment of the ontology, with the corresponding individuals which are the objects excavated from the site, and their semantic annotations on the data and documents. The ontology schema is also responsible for reflecting the archaeological interpretations of the objects through proper relationships between different entities of the objects.

4.1 The ontology schema

The core of the schema is the concept “siteFeature” which stores all the excavated objects. The basic process behind the “ArchaeoKM” is very straightforward. Archaeologists are responsible for the indexation of the findings on the orthophoto. Those findings are then enriched in the domain ontology through respective objects. The spatial facilitator covers also the adjustments carried out within the ontology schema in order to incorporate

the spatial components. The ontology schema represents the terminological definition. It is defined with the OWL-DL language which is a description logic language (Baader, 2003). Actually, it represents the definition of concepts and roles which are properties and relations between concepts.

The ontology schema in the ArchaeoKM platform has to be adjusted in order to incorporate the spatial functions and operations. In general, the spatial operations and functions provided by the current database system can be broadly categorized into two categories – spatial processing functions and spatial relationship functions. The first category represents unary functions and the second represents binary functions. The unary operations return the new geometry itself whereas the binary operations return the binary value. Figure 1 shows the two categories of the spatial functions. By adding spatial relations between site features (feat:siteFeature), archeologists define a certain kind of knowledge concerning the disposition of findings on the current site. For instance, a knowledge specification about the domain can be made as it exist a finding “oven” and a finding “railway” that overlap a finding “building”. It means that the building is a finding “factory”. So, a concept “feat:factory” is defined as a subclass of “feat:siteFeature” and with the condition described previously. It can be easily computed with the help of the 2D/3D annotations of the indexes.

4.2 Enrichment of the ontology schema by adding spatial operations

The two sets of spatial operations are represented with two different approaches in the ontology schema. The first set of spatial operations needs to be treated as we treat the features excavated in the concept “siteFeature” since they result in geometries. This is achieved by introducing a new concept “spatialAnalysis” with sub-concepts to support such 2D and 3D operations. It is important to define a property that represents the relationship between the spatial concepts with the feature excavated. It is defined through predicate “hasSpatialAnalysis”.

The second set of spatial operations provides the status of the particular relationship between two objects. Such relationships are binary relationship and they show whether there exists a particular relationship between two objects or not. As these relationships do not yield new geometry and they

perform much in similar line to the object relationships, they are represented as a form of object relationship. It is shown by “hasSpatialRelAnalysis” and has both range and domain as “siteFeature”. It is possible to perform binary spatial operations between the objects of “siteFeature” and “spatialAnalysis”. From this point, it can be see that spatial information which defines the knowledge of a domain can be added in the ontology. In addition the properties and relationships can be verified with the help of spatial database by the spatial facilitator.

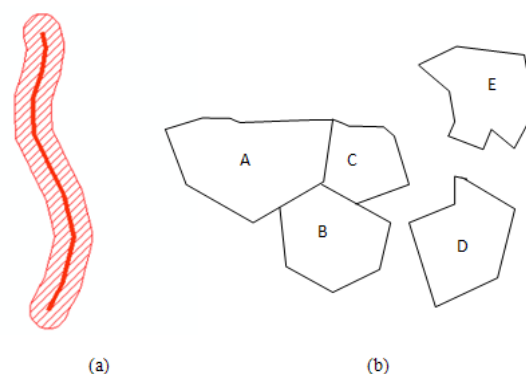


Figure 1. Two types of spatial operations (a) Buffering (spatial processing) a linear feature (red linear feature) generates crossed polygonal features around it (b) Five polygons to demonstrate the touch (Spatial Relationship) options – A touches B → true, A touches D → false

In order to define the new spatial relationships between individuals, any individual from the concept “feat:siteFeature” has relationship to an individual “shape:Feature” which can be 2D or a 3D shape. Almost all of the existing database system supports storage and retrieval of the spatial data with their spatial extensions. They also support spatial operations on these data. However the scales of spatial operations vary from one database system to another. They also vary in the support for the 3D data set. Currently, there are not many 3D spatial operations supported by the existing database systems. Oracle 11g (Oracle, 2007) and PostGIS 1.3.5 (PostgreSQL, 2008) of PostgreSQL 8.3 are the leading database systems supporting the 3D operations. However such operations are mostly limited to unary operations. ArchaeoKM intends to use the advancements in spatial operations in PostGIS to enrich the ontology. All the operations are carried out in accordance to the SQL syntax of the spatial operations of the database systems and will be performed on the data stored in the database. The results that are generated through such

operations are used to enrich the ontology. In this manner, the database is merely used as the tool to store the spatial data and to carry out the required spatial operations. The relationships and the results are managed through the ontology.

4.3 The Extension of SWRL with spatial analysis

This section presents the method used to integrate the spatial operations (unary, binary, 2D and 3D operations) with the help of Horn clauses (SWRL language) in order to define knowledge on the Industrial Archaeological Site (IAS) with ontologies as well as rules.

Example of a SWRL expression. The following example creates a new relationship “cooperatedWith” between authors if they worked on the same publication.

```
Publication(?a) ^
hasAuthor(?a, ?y) ^
hasAuthor(?a, ?z) ^
differentFrom(?y, ?z)
    → cooperatedWith(?y, ?z)
```

In addition it exists “built-in” predicates in the SWRL language that allow the computation of advance information. For instance: `Person(?p) ^ hasAge(?p, ?age) ^ swrlb:greaterThan(?age, 18) → Adult(?p)`. The method presented in this section consists in showing how can be extended the built-in predicates with spatial operation.

Examples of a SWRL rule (Horn clauses) with a spatial operation Buffer. The result is a new relation between site feature x and y that satisfy the “buffer” operation. Actually it consists to enrich the ontology by adding a new relation between two individuals.

```
river(?x) ^
Building(?y) ^
archaeokm:Buffer(?x, ?y, 50)
    → isLiableToFlooding (?y)

oven(?x) ^
Building(?y) ^
archaeokm:Buffer(?x, ?y, 50)
    → hasOven(?y, ?x)
```

The spatial facilitator component in our architecture is composed of an SWRL engine improved with spatial operation predicates that allow the definition of complex rules. In order to realize these operations, each spatial operation is converted to an SQL request. The example buffer given is a combination of spatial operation Buffer

and “within”. The operation consists to define a SQL query (e.g. next example). Thus, an SQL query has to be defined for every SWRL operation in order to process the expected result.

Example of a built-in operation converted into an SQL request

```
SELECT y
FROM Ovenbb_tb
WHERE
    within(the_geom,
        buffer((
            SELECT the_geom
            FROM Ovenbb_tb
            WHERE name = x), 50)
        )
```

4.4 An example

This section presents a scenario to present our case study. We are using the bounding boxes of five distinct objects that are found in the industrial archaeological site. Those findings are specialized concepts of the concept “siteFeature”. These concepts are “Oven”, “Railway”, “Structure”, “Chimney”, “Pipeline” and “Plant”. Those concepts represent the objects excavated.

Once the objects are excavated from the site, they are used to enrich the ontology against their respective concept. The geometries of these objects are stored in the PostgreSQL database as the spatial data type provided by PostGIS – spatial extension of the database system.

To illustrate the spatial operations we discussed in the previous sections, we take one operation from each unary and binary spatial operation and demonstrate how they enrich the ontology. To begin with we take “Buffer operation” which buffers the feature and is a unary operation. We define a buffer of 50 meters around the “Oven_1” and populate the ontology with a new specialized concept “Buffer” of the concept “spatialAnalysis”. Then we populate this concept with the corresponding object “buffOven_1_50m” and store the resulted coordinate.

Example of a spatial operation “buffer”.

```
SELECT AsText(
    buffer((
        SELECT the_geom
        FROM Ovenbb_tb
        WHERE name = 'Oven_1') , 50))
```

It is clear that when we specialize the concept “spatialAnalysis”, a respective specialized object property under “hasSpatialAnalysis” has to be

created too. In this case “hasBuffer” has to be created under “hasSpatialAnalysis” simultaneously. So the new RDF triplet from the operation above would be (“siteFeature”, “hasBuffer”, “Buffer”). The knowledge base is then populated with “Oven_1” “hasBuffer” “buffOven_1_50m”.

The next operation is the binary operation and we take as an example the “within” operation which will show whether or not an object is contained in the next one. This will generate binary results of the operations. But to make this operation more appropriate for our case, we modify the operation so that it will extract all objects within the feature. The spatial operation listed below will list out all the features that are within the feature “Plant_1”.

Example of a spatial operation “within”.

```
SELECT name
FROM Ovenbb_tb
WHERE
  within( the_geom,
    (SELECT the_geom
     FROM Ovenbb_tb
     WHERE name = 'Plant_1')
  )
```

The binary operations are used as the object property “hasSpatialRelAnalysis” in the ontology. A new specialized property “hasWithin” is created with the RDF triplet as (“siteFeature”, “hasWithin”, “siteFeature”). The knowledge base is then enriched with these triplets (“Plant_1”, “hasWithin”, (“Oven_1”, “Railway_1”, “Pipeline_1”, etc.)).

5 CONCLUSION

In this paper, the ArchaeoKM platform has been presented by focusing on spatial analyses and by showing the combination of these analyses with Semantic Web technologies. These benefits are materialized by the population and the enrichment processes of a domain ontology with the help of spatial operations using industrial archaeological site data. An additional benefit is the extension of the SWRL language by providing built-in “spatial operations”. Thus, this extension allows the definition of rules supplying new knowledge on the IAS. These processes are managed by the spatial facilitator component of the ArchaeoKM platform. Although the case study uses industrial archaeology for the description of the approach, it can be used in other areas where the spatial data are the

predominant data type. Future work will be the identification of all spatial operations that can be handled by spatial database systems in order to offer an overview of its capabilities. At the moment only few of them are prototyped as a proof of concept.

REFERENCES

- Baader, F., et al, 2003, *The Description Logic Handbook – Theory, Implementation and Applications*, ISBN: 0521781760, Cambridge University Press, January
- Castano, S., Espinosa, S., Ferrara, A., Karkaletsis, V., Kaya, a., Melzer, S., Moller, R., Montanelli S., Petasis, G., 2007. *Ontology Dynamics with Multimedia Information: The BOEMIE Evolution Methodology*, In Proc. of International Workshop on Ontology Dynamics (IWOD) ESWC 2007 Workshop, Innsbruck, Austria
- Costa, S. and Zanini, E., 2008. *Sharing knowledge in archaeology: looking forward the next decade*. Digital Heritage in the New Knowledge Environment: Shared spaces & open paths to cultural content, 31 October – 02 November in Athens, Greece
- Doerr, M., 2008. *The CIDOC Conceptual Reference Model – A New Standard for Interoperability*. Digital Heritage in the New Knowledge Environment: Shared spaces & open paths to cultural content, in Athens, Greece
- Green, J., Dolbear, C., Goodwin, J., 2008. *Creating a semantic Integration System using Spatial Data*, 7th International Semantic Web Conference (ISWC2008), Karlsruhe, Germany, October 26-30
- Guarino, N., 1994; *The ontological level*, in R. Casati B. S. & White G., eds, *Philosophy and the cognitive sciences*, Hölder-Pichler-Tempsky
- Kansa, E. C., 2008. *Opening Archaeology to Mash-ups: Field Data and an Incremental Approach to Semantic*, Digital Heritage in the New Knowledge Environment: Shared spaces & open paths to cultural content in Athens, Greece
- Karmacharya, A., Cruz, C., Marzani, F., Boochs, F., 2008. *Industrial Archaeology: Case study of Knowledge Management for Spatial Data of Findings*. 2nd International Workshop on Personalized Access to Cultural Heritage, in conjunction with 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Hannover, Germany
- Kollias, S., 2008. *Achieving Semantic Interoperability in Europeana*, Digital Heritage in the New Knowledge Environment: Shared spaces & open paths to cultural content, 31 October – 02 November in Athens, Greece
- Oracle, 2007. *Oracle Spatial Developer Guide 11g Release*, Oracle
- PostgreSQL, 2008. *PostGIS Manual*, PostgreSQL documentation