Support of Spatial Analysis through a Knowledgebase– A new concept to exploit spatial information shown for Industrial Archaeology

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

[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

Designing and developing spatial ontology is an emerging research topic today and there has been lots of research works going on in the field. However, those researches mostly focus on data interoperability through spatial ontology and rarely provide any assistance to spatial analysis. We propose a unique concept as an extension to our Web based system for data management in the field of Industrial Archaeology, *ArchaeoKM*, a spatial tool which uses the spatial operations and functions provided by the current database system to enrich and populate the knowledge schema of the application so that results of the spatial analysis could be managed through the knowledge base. Domain Rules are the driving force behind *ArchaeoKM* as they are the foundations for the domain ontology within the application. The domain ontology is the main core of the application system. Spatial operations and functions complement these domain rules by providing supports through enriching the ontology with the new entities reflecting the analysis process and finally populating them their results.

Keywords: Industrial Archaeology, Spatial Analysis, Knowledge Management, Ontology,

1 Introduction

The semantic web technology relies heavily in the ontology that helps the information processed through the machine understandable language. Therefore for any semantic web application to be efficient and successful, it is very important to have a strong ontology behind it. In recent years there has been a huge upsurge in researches in semantic web and most of them primarily focus on the ontology engineering. The topic is so vast that there has been research on the use of ontology in almost every field. However, we will be discussing the applications and research works on the spatial ontology and our ideas of its usage for better management of the spatial analysis. We have used the field of Industrial Archaeology as our background to prove the case because it provides sufficient materials both spatial and semantic to illustrate the concept. The demonstrate site is Krupp in Essen belt, 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 is never rebuilt making it an ideal site for industrial archaeological excavation.

During the excavation, findings are scanned through the terrestrial laser scanner thus making point cloud as the major data set. The point cloud represents the geometric information of the findings making it ideal dataset for any GIS application. However, we intend to take it further through our system (which will be termed as *ArchaeoKM* from this point onward). Besides, point cloud there are other pattern of data set which provides semantic information of the findings and it won't be intelligent to ignore those information. *ArchaeoKM* intends to use the advancement in Semantic Web technology and Spatial Database Management System (SDBMs) to integrate the spatial analysis within the knowledge base of Semantic Web. This will provide an add-on to the existing GIS System where only the geometry is considered for the analysis purpose.

The paper discusses various aspects of *ArchaeoKM* and how it could provide the support to the existing development in Geospatial technology. Section 2 provides an overview on existing research works on Semantic Web in Spatial technology. The concept behind the *ArchaeoKM* and its architecture is discussed under Section 3. Section 4 discusses on the integration of Spatial Analysis within the *ArchaeoKM* and how it contributes in the development of Geospatial technology. Finally section 5 concludes by summarizing the paper.

2 Existing Research Works

The existing GIS system does not use semantic explicitly. They primarily focus on geometry and it is quite well known that most GIS systems store the geometry in their native formats. They provide the spatial queries and functions through their custom made interfaces which are strictly geometry based. With the advancement Spatial Database Management Systems (SDBMS), it is now possible to store the geometry in those database systems and do not have to rely on the GIS tools to store or retrieve the geometries. It has even become possible to perform spatial operations within those database systems. Research projects like GIS DILAS [1] or 3D MURALE [2] take advantages of those features of current database systems to carry out spatial operations within their systems.

The inclusion of semantic into any information system adds the efficient on the system as a whole [3]. This applies for the information system that involves geometry greatly like in Geographic Information Systems (GIS). There has been few research works to include the semantic layers within GIS but they are not as many as in some other cases. In addition the current research works mostly focus on the use of semantic for semantic interoperability of the GIS data so that the GIS data could be exchanged over broader and heterogeneous platforms [4]. The ontology is also being used data mapping in order to have comprehensive data integration. This has been discussed in the research works [5,6,7].

Though there are several research works, common consensus foundation ontology has not yet been agreed upon. Open Geospatial Consortium (OGC) is playing a major role to develop a consensus among different stakeholder on various aspect of geospatial technology. Data interoperability is a major area in which OGC is concerned upon and it has developed different standards for this. Groups like Geospatial Incubator have taken the works of OGC to formulate steps in updating the w3c geo vocabulary and preparing the groundwork to develop comprehensive geospatial ontology. In the process it has been reviewing different spatial ontologies that exist in the web [8].

3 ArchaeoKM and the Principle Behind

It has been mentioned in the previous section about the lack of foundational ontology in geospatial technology. However, the existence of such foundational ontology would have minimal influence in our studies since we are focusing more on spatial integration through spatial analysis rather than the semantic interoperability of the data. In this section we will be discussing on the system *ArchaeoKM* and the principle behind the application.

ArchaeoKM is a shift from conventional approaches. It is a web platform based on the semantic web technologies and knowledge management to store data during the excavation process and to manage knowledge acquired during the finding and identification process. The collaborative process between archeologists is facilitated by the platform in order to generate knowledge from the dataset once the data are stored in relevant data structure. The principle of our approach consists in using semantic annotation in order to have a semantic view on datasets. The domain ontology allows us to build a global schema between data sources. This global schema allows annotating, index, searching and retrieving data and documents. The domain ontology represents the descriptions from the excavation in more structured approach. It can be considered as the core of the system as it contains the concepts and the relationship between concepts which actually represent the excavation. The basic knowledge schema of *ArchaeoKM* is shown in figure 1.

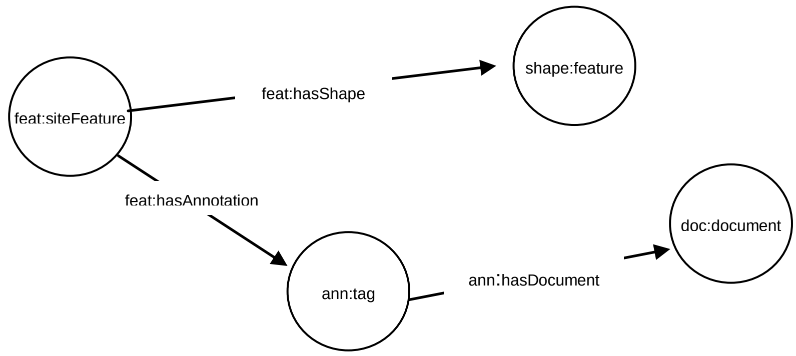

Fig 1: Basic Knowledge Schema of ArchaeoKM

The domain ontology also includes the spatial concepts which are used for spatial analysis within *ArchaeoKM*. They are included as concepts and properties within the knowledge base. Details on the spatial integration within the platform are discussed later in this section.

3.1 The Architecture

The system architecture of *ArchaeoKM* has three distinct but interrelated levels. Each level has its own distinct features and functionalities but they are responsible to provide inputs to the higher level. As could be seen in the figure 2, those layers are Syntactic, Semantic and Knowledge levels. Besides these levels there is a facilitator as Spatial Facilitator which is responsible to provide spatial dimension to the platform.

The bottom most level is the Syntactic level. This is the main repository of the system where the information gathered from the excavation site is stored. Basically data are stored in two formats – file system and Relational Database Management System (RDBMS). The semantic data excavated from site are stored as the file system. This includes the archive data, images and archaeological notes taken during the excavation process. The geometries representing the findings are stored in RDBMS. It has already been mentioned that the findings are scanned through the terrestrial laser scanners to achieve the point cloud representing geometries of the objects. It is however not possible to store each and every point in the point clouds, the geometries representing the boundaries are stored as spatial data type in Spatial Database Management System (SDBMS), the extensions of RDBMS. We are using PostgreSQL and its spatial extension PostGIS to illustrate our idea.

The next level is the Semantic level. This can be considered as the heart of the system as all the major tasks are done within this level. The level contains domain ontology (DO) which can be considered as a structured representation of the observations and analysis of the findings from the excavation site. The DO represents the findings and their relationships with each other and the surrounding as concepts and relations between concepts. Web Ontology Language (OWL) is used to specify ontology or more generally some ontological and terminological resources by defining concepts used to represent a domain of knowledge. As could be shown in the figure 2, this well netted web of concepts is a powerhouse of knowledge which will be exploited through the Knowledge level. The concepts are annotated to their relevant data and documents in the Syntactic level. In the Semantic Web context, the content of a document can be described and annotated using knowledge such as RDF, and OWL. Besides holding the knowledge of archaeological observations of the site the DO also consists of concepts and properties which are the representations of spatial operations that could applied to the geometries of the findings. In collaboration with the semantic knowledge, these spatial concepts could provide a breakthrough in how the knowledge is manipulated.

One of the highlight of Semantic Web technology is the possibility of applying rules to come out with new solutions. This is termed as deductive reasoning in Semantic Web and has several languages supporting it. Semantic Web Rule Language (SWRL)[9] is one such language which could apply in the DO to come out with a solution. However, in current form they could be only applied for semantic information. With our approach, it will be possible to use spatial aspects (which is integrated within the DO) and semantic aspect to formulize a solution through Rule Languages like SWRL. It will be discussed in detail in section 3.2 later.

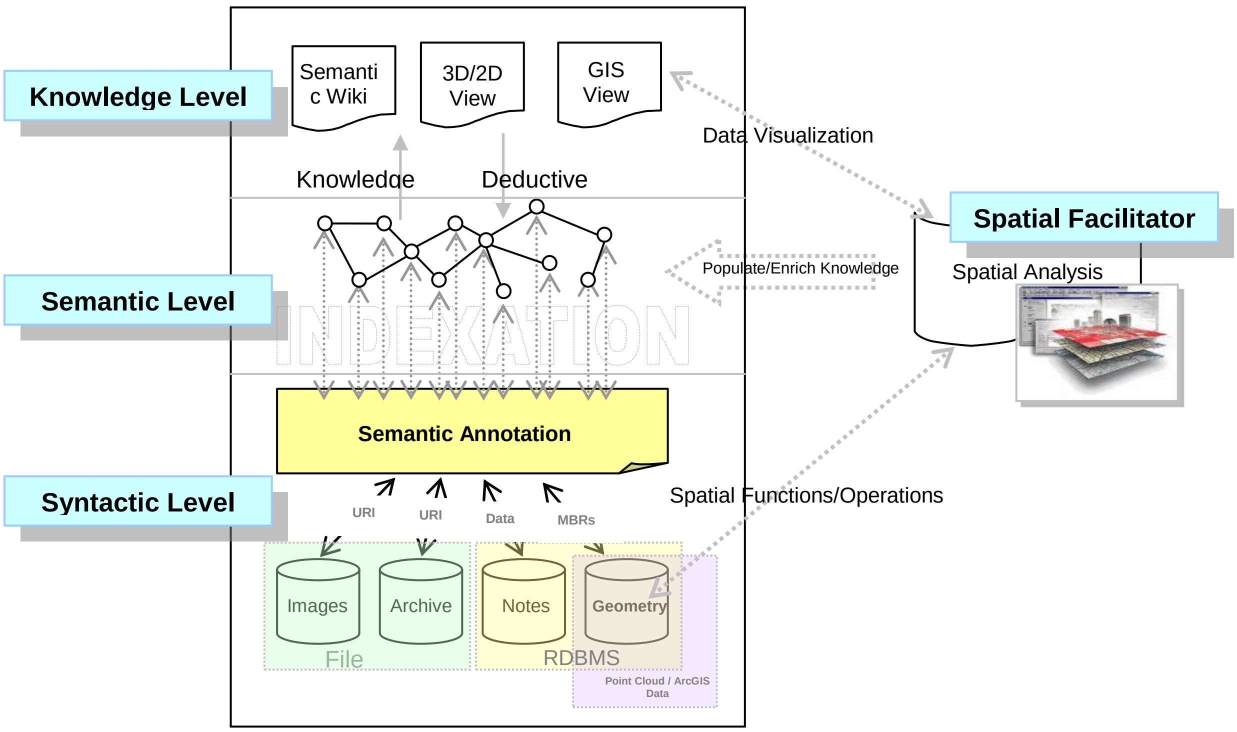

Fig 2: System Architecture of *ArchaeoKM*

The uppermost level is the Knowledge level. It can be considered as the face of the system as it is the user interface where users interact. This level represents knowledge generated in Semantic level in different forms and formats. The main format is which the knowledge is represented is through Semantic Wiki. These wiki pages are not only designed to show the knowledge that are generated and managed through the bottom two levels, they are designed to perform semantic queries to derive new knowledge. This will be possible through the interface within the semantic wiki – the semantic wiki will provide a platform through which user can launch their queries and the results will be displayed through the query languages of RDF like SPARQL [10] or rule language as SWRL. In this way they will be different from the existing wiki pages. Thus, *ArchaeoKM* is close to the semantic extension of Wikipedia [11], but data handling and managing extends beyond textual data. It also handles 3D or 2D object models of the findings besides the textual and image data.

The interfaces in this level are different than that of existing GIS applications as they are made to address the request on knowledge than on the information. However the visualizations of the results might have similar outlooks. The interface in *ArchaeoKM* will be able to process the request having combined semantic and spatial query in a single step which was not possible in existing GIS tool. In this manner, we hope *ArchaeoKM* will be able to process complicated *Location Based Analysis* request in relatively fewer steps. We will be discussing this in next section.

3.2 Spatial Analysis in *ArchaeoKM*

We have already discussed the integration of spatial operations and unction in bit and pieces in the previous sections. This section we will be summarizing the approach of spatial integration within *ArchaeoKM* and illustrate the process with example.

Almost every database management system today includes the spatial extension within it to store the geometry as spatial data type. This demonstrates the importance of geometrical information in any Information Systems today and points out that the spatial analysis of any spatial data can be performed outside the GIS application software. We will be using PostgreSQL with spatial extension as PostGIS to store and retrieve our spatial data.

PostGIS 1.3.2 is the spatial extension of PostgreSQL 8.3 object relational database system that allows the spatial objects to be stored in the database [12]. A strong tendency has been seen in the last few years that big GIS vendors like GRASS and ESRI have shifted their support towards PostGIS. PostGIS supports the storage of *point*, *line*, *polygon*, *multipoint*, *multiline*, *multipolygon*, and *geometrycollections*. It follows the specification provided by *OGC* for the simple features to store these objects. Those are specified in the *Open GIS Well Know Text* (*WKT*) or *Well Known Binary* (*WKB*) Formats[*]. It supports all the objects and functions specified by OGC "Simple Features for SQL" specification. However PostGIS extends by supporting 3D and 4D objects. PostGIS has given those extensions names as EWKB or EWKT (Extended Well Known Binary and Extended Well Known Text). In contrast to the Simple Feature Specification by OGC, those extensions support the embedded Spatial Reference Identifier (SRID) information.

IO of these formats is available with the interfaces

```
a. 2D objects
        bytea WKB = asBinary(geometry); text WKT =
                asText(geometry); geometry = GeomFromWKB(bytea WKB, SRID);
        geometry =
                GeometryFromText(text WKT, SRID);
b. 3D objects
        bytea EWKB = asEWKB(geometry); text EWKT =
                asEWKT(geometry); geometry = GeomFromEWKB(bytea EWKB);
        geometry =
                GeomFromEWKT(text EWKT);
```

Point Clouds collected during scanning of the findings are the main source of geometric information. However, it is not possible to store each and every point from the point cloud so the boundary of the point cloud will be stored as spatial data within the database. The boundary stored in the database will be mapped to the file where the point cloud is stored so that the point clouds could be extracted for 3D object modeling.

PostGIS also supports a vast range of spatial operations that will be utilized by *ArchaeoKM* to perform its spatial analysis. The knowledge base in the Semantic level is modified to fit in spatial operations and functions provided by the database. We have identified two major sets of spatial functions and operations provided by a database system (though there are more than two sets, we categorized them in two general sets for our purpose) – One which returns the geometry and the next which return the Boolean value. The first is similar to Geometry Processing Functions and the second to Geometry Relationship Functions in PostGIS. These two sets of operations are integrated within the DO according to their features. The basic knowledge base of figure 1 is modified to accommodate these spatial functions and could be seen in figure 3.

The first set which returns geometry is treated as a concept within the DO. This makes sense as the results of such operations yield geometries which need to be stored. A new specialized concept *SpatialAnalysis* it representing the operations from the set are introduced. An example of such operation would be *buffer* operation which will create *buffer* geometry of certain distance around the feature. Similarly, an object property is introduced to against each specialized concept to map the respective result with the corresponding feature. In other words, a generalized object property *hasSpatialAnalysis* is introduced which is the spatial relation between the concepts in the generalize concept of *siteFeature* and the concepts in the generalized concept of *SpatialAnalysis* at the individual levels. In short a generalized statement can be represented by the triplet *siteFeature hasSpatialAnalysis spatialAnalysis.* This generalized statement is later specialized through specialized statements with specialized triplet like *Wall hasBuffer Buffer*. So, when a *buffer* is created on an individual of concept *Wall W1*of 50 meters, an individual *BuffW150m* is created within the specialized concept *Buffer*. The *hasSpatialAnalysis* will have a specialized property *hasBuffer* which provide relation between the wall *W1* with the *BuffW150m.*

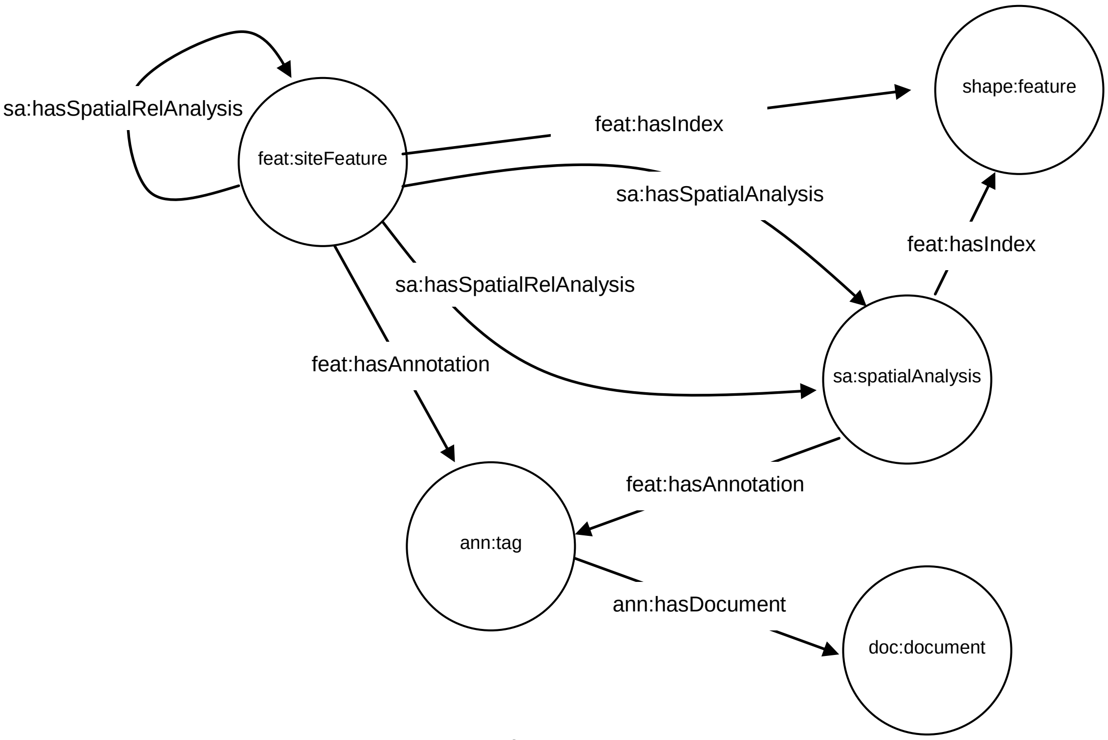

Fig 3: Spatial Adjustment of the knowledge schema in *ArchaeoKM*

The next set of operations which provide binary results are integrated as object properties under generalized property *hasSpatialRelAnalysis.* Since these operations are performed with the features excavated from the sites, the relationships are formulated among the concepts within the generalized concept *siteFeature.* Additionally, these types of relationships exist between the concepts of *siteFeature* and *spatialAnalysis.* An example of such operations can be the spatial operation *within.* This operation determines whether one geometry is within the next. The geometry could be the geometry of the concepts or that of results of the spatial analysis like *buffer*. In our example of *within,* we create an object property *hasWithin* within the *hasSpatialRelAnalysis.* Though, this operation yields only binary results, queries could be performed in a manner that they provide results as the objects which are within the particular object. We will look at it with a small demonstration which implies the *within* operation and generates the results. As already mentioned, the site plan of the area has been stored in the database which includes the geometry information of the different sections of the site. We take a portion of the site plan for this demonstration. The section of the site plan is shown in the figure 4. As could be seen, it is drawn in *ArcGIS* and has object which we named as *R1* as for the room and objects like *O1, O2* and *C1*. We would like to extract the objects inside the block *R1* to enrich it in the knowledge base. This could be done through two PostGIS spatial operations – *ST_Contain* and *ST_Within*. We are taking the operation *ST_Within* with the query mentioned below for this case. However, it should be understood that the use of spatial operations to enrich the knowledge base is directly related to the needs of them for the spatial analysis and generate the domain rules.

```
SELECT betriebs_2 As Objects, AsText(the_geom) As Geometry
FROM triald
WHERE Within(the_geom,
       (SELECT the_geom FROM
       triald WHERE betriebs_2='R1'))
```

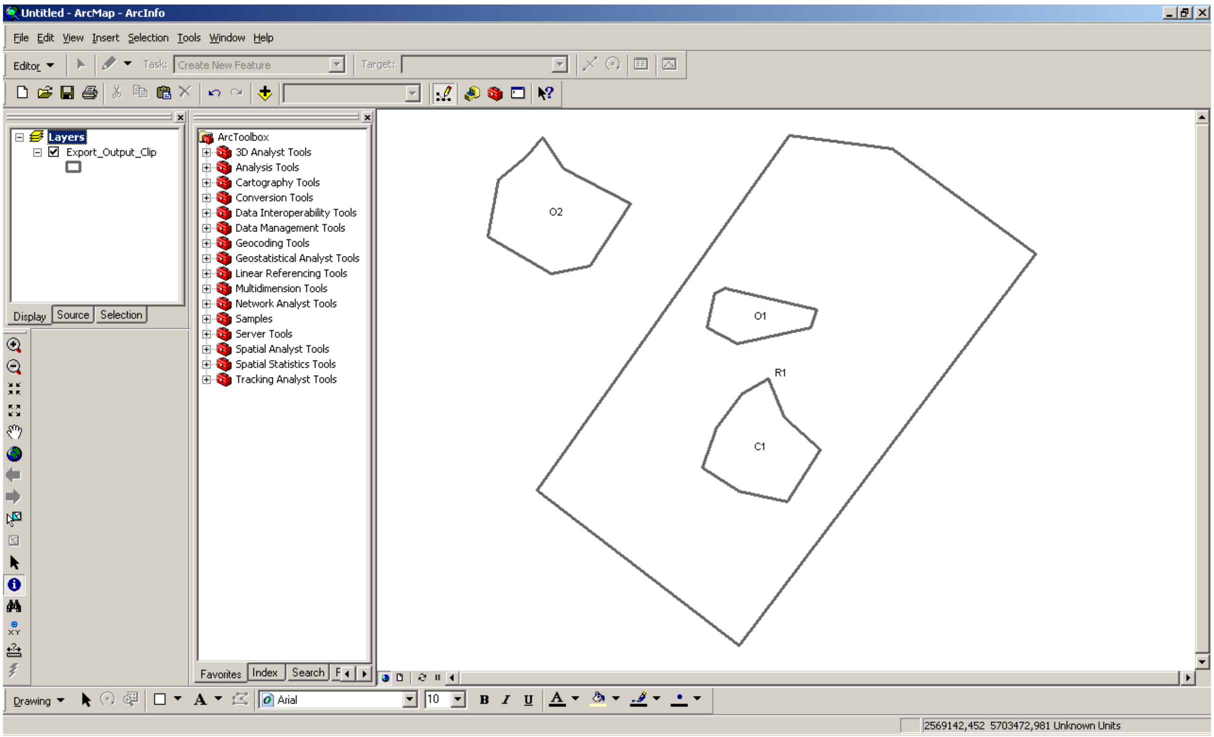

Fig 4: The section of site plan for the demonstration

The query yields result

```
"O1";"MULTIPOLYGON(((2569162.34430857 5703467.77497359,2569167.5503294
5703466.56936876,2569167.22152808 5703465.5281646,2569162.94711098
5703464.59656087,2569161.19350397 5703465.5281646,2569161.63190572
5703467.50097249,2569162.28950835 5703467.82977381,2569162.34430857
5703467.77497359)))"

"C1";"MULTIPOLYGON(((2569165.68712194 5703460.32214377,2569167.76953027
5703458.45893632,2569165.8515226 5703455.44492426,2569163.05671142
5703456.04772667,2569160.91950287 5703457.41773215,2569161.74150616
5703459.71934136,2569163.22111208 5703461.69214925,2569164.75551822
5703462.56895276,2569165.63232172 5703460.43174421,2569165.68712194
5703460.32214377)))"
```

It could be seen that two objects *O1* and *C1* lie within the block *R1*. Now it is turn to enrich the knowledge base with this result. It has already been mentioned that the specialized property *hasWithin* is created within *hasSpatialRelAnalysis* with the domain as the concept *siteFeature* and range as the concepts *siteFeature* and *spatialAnalysis*. The individual *R1* which is an individual of one of the concepts in *siteFeature hasWithin* individuals *O1* and *C1* again individuals within the concepts in *siteFeature*. Thus the knowledge base in enriched with the result of the spatial function.

4 What Next?

The GIS technology has come a long way since its early days. The primary reason behind its growth is its adaptability to integrate the new technologies within it. Since the early days when the technology was using file system to manage the spatial data and operations, it has been using geometry as the main data for analysis. However, there were lots of problem due to the short coming of the file system. The biggest short coming was that each system has its own file system and there were problems in data interoperability. Additionally, the sizes of files were often too big making the processing of them very slow. There were other problems as well. When it migrated to the Relational Database Management System, most of such problems were ironed out. However, the problems regarding data interoperability persisted. The other limitation of using the database system in GIS is that it limits the freedom of analysis and restricts us to the functionalities provided by the application for the analysis.

Today, most of research works in semantic web technology are focusing on solving the problems of interoperability. The ontologies that are developed to address the problem are mostly set of controlled vocabularies or metadata which provides semantic meaning of the terms used in the field. Though we agree on their importance, it is only one area where the ontology can be used for. The next limitation we talked before has got very little attention and this is the area we are focusing on. *ArchaeoKM* uses the ontology and description language provided by ontology language to define the domain rules and enrich the ontology. In the process we integrate the spatial operations and functions within the ontological descriptive language so they can be processed as rules in rule languages of semantic web like Semantic Web Rule Language. By doing this, *ArchaeoKM* is demonstrating how the current GIS application tools can take advantage of those rule languages to make their applications dynamic.

Semantic Web Rule Language (SWRL) is a combination of Web Ontology Language (OWL) and Rule Markup Language to define a rule in a knowledge base. It is expressed through Horn logic. A simple example of SWRL – `hasParent(?x1,?x2) ^ Woman(?x2) → hasMother(?x1,?x2).` Here *Woman* is a concept and *hasParent* and *hasMother* are the object properties of the descriptive language of OWL and *x1*, *x2* are the instances. The rule is self expressive. It says if *x1 hasParent x2* and *x2 is a woman* then *x1 hasMother x2*.

The integration of spatial functions through different logic within the knowledge base has made possible to apply rules in to the spatial data which are annotated through these logics. Continuing with the example of spatial function *within* and its representation *hasWithin* in the knowledge base we can formulate a rule as

```
hasWithin(?x1,?x2) → hasPart(?x1,?x2).
```

This is a very simple example stating that if individual (*x2*) are spatially within another one (*x1*), then the later individual (*x1*) has part earlier object (*x2*). Though this very simple and straight forward case, it answers many questions – first of which is inadequacy of current GIS application to provide semantic relationship of the object. In this case existing software can say that an object is within another spatially but fails to that the object is a part of another (which is semantic relationship between the objects). The knowledge base could be then enriched with the rules as such rules yield new descriptive logics (*hasPart* in this case) and populate them with the results. By enriching the knowledge base dynamically we are providing users the flexibility of interpretations of their view. Spatial operations could be combined with other semantic operations to form more complex rules and to carry out complex analysis.

5 Conclusion

The rapid advancement in Semantic Web technology has taken the world by surprise and today the research works in semantic web and related ontological engineering are one of the most researched. The flexibility and dynamism that it provides has provided lots of possibilities which were not there few years back. Today, the semantic web technology is integrated or at least in the process in every field of Information Science. However, in the field of GIS, it is one of the least researched topics and whatever researched are in the area of data interoperability through ontology mapping.

The concept presented as *ArchaeoKM* could contribute the development in GIS technology. The paper discusses the possibilities of integrating semantic web technology with spatial operations to enrich and populate the knowledge base. The combination of rule languages with the spatial analysis will add a new dimension in which users interpret their views. A layer in between the data layer and the visualization could be added in the existing GIS system which performs the ontological operations. This layer will act as the knowledge base in the current system. The inclusion of knowledge base in the existing GIS system will provide a firm base by providing much needed dynamism to the system.

References

1. 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
2. J. Cosmas,, T. Itagaki,, D Green, E. Grabczewski, M. Waelkens, R. Degeest, et al. "*3D MURALE:A Multimedia System for Archaeology*". Proc. ACM Virtual Reality, Archaeology and Cultural Heritage (VAST 2001). Nov 2001
3. Semantic Interoperability Community of Practice (SICoP), "*Introducing Semantic Technologies and the Vision of the Semantic Web*", 2005
4. Roman, D., Klien, E., Skogan D., "*SWING – A Semantic Web Services Framework for the Geospatial Domain.*" Position Paper at the Terra Cognita 2006 - Directions to the Geospatial Semantic Web Workshop, Athens, USA, 2006
5. Cruz I. F., "*Geospatial Data Integration.*", ADVIS Lab, Department of Computer Science, University of Illinois, Chicago, 2004
6. Chaudhary A., Sunna W., Cruz I. F., "*Semi-automatic Ontology Alignment for Geospatial Data Integration*", 3rd Intl. Conference on Geographic Information System (GIScience), Adelphi, Meryland, 2004
7. Tanasescu, V., Gugliotta, A., Domingue, J., Gutiérrez Villarías, L., Davies, R., Rowlatt, M., Richardson, M., Stinčić, S., "*A Semantic Web Services GIS based Emergency Management Application*." Workshop: Semantic Web Challenge, The 5th International Semantic Web Conference (ISWC06), Athens, Georgia, USA, 2006
8. Lieberman J., Singh R., Goad C., „"*W3C Geospatial Ontologies – W3C Incubator Group Report*", W3C, http://www.w3.org/2005/Incubator/geo/XGR-geo-ont-20071023/, [Last visited on 23 June 2009]
9. W3C, *"SWRL: A Semantic Web Rule Language Combining OWL and RuleML"*, http://www.w3.org/Submission/SWRL/, Last Visited: 25 November 2008
10. W3C, *"SPARQL Query Language for RDF"*, http://www.w3.org/TR/rdf-sparql-query/, Last Visited: 25 November 2008
11. Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., Studer, R., "*Semantic Wikipedia*". In Proceedings of the 15th international Conference on World Wide Web (Edinburgh, Scotland, May 23 - 26, 2006). WWW '06. ACM, New York, NY, 585-594. DOI= http://doi.acm.org/10.1145/1135777.1135863
12. PostgreSQL, "*PostGIS Manual*", PostgreSQL documentation.