7

Spatialization of the Semantic Web

Ashish Karmacharya[1], Christophe Cruz[2] and Frank Boochs[1]
[1]Institut i3mainz, am Fachbereich 1 - Geoinformatik und Vermessung, Fachhochschule Mainz,
[2]Laboratoire Le2i, UMR-5158 CNRS, UFR Sciences et Techniques, Université de Bourgogne,
[1]Germany
[2]France

1. Introduction

The abstraction of the real world melds the semantics of its objects with the spatial characteristics seamlessly. This is visible in a way the human perceives the real world where it is often difficult to pin point the spatial characteristics of the objects from their semantics. In other words the spatial characteristics are generally hidden with the semantics of the objects. As for example, describing relations of objects the terms near, far or touching are often used which are spatial relations but in general considered as semantic properties which is not true. Hence, it is a trend to consider that the spatial behaviors of objects are parts of its semantics. Similar approaches where the spatial properties are considered as part of semantics have been translated in technical advancements made by the technologies. There is a general trend to mix up spatial components in the semantics or the semantics in the spatial components within technologies. For instance, a classic GIS ignores semantics of objects to focus on the spatial components whereas a non GIS uses spatial components as the semantic parameters of the objects. As the technology is getting matured, it is moving closer to the human perception of the real world. Today, the knowledge management is being researched in real sense to model and to manage knowledge possessed by humans which is basically the perception of the real world.

The emergence of Internet technologies has provided a strong base to share the information in a wider community. As the needs of information have grown it has become necessary to represent them in a proper and meaningful way. It involves attesting semantics to the documents. The major approach to attach semantics to documents involves first to categorize them properly and then to index them with the relevant semantics for efficient retrieval. This categorization and indexing of the Web documents have become important topic for research. These researches focus on the use of knowledge management to structure documents which involves ontologies to conceptualize knowledge of a specific domain. Then, there is knowledge representation which is a vital part of knowledge management. It provides the possibilities to represent knowledge in order to be inferred. Knowledge representations and reasonings have traditionally been a domain within Artificial Intelligence. However, the recent growth in Semantic Web technologies has added fuel to the use of knowledge explicitly in a Web environment. The XML-based knowledge languages could be inferred through different inference mechanisms in order to infer knowledge.

1.1 Knowledge management and the semantic web

The current version of the Web could only be processed through human intelligence. Though the Internet technologies have taken a huge leap forward since it evolved, the fact is the information within the technology still needs to be interpreted by human brain. However, in his paper (Berners-Lee et al., 2001), Tim Berners-Lee and coauthors have envisaged the next generation of the Web which they call "the Semantic Web". In this Web the information is given with well-defined meaning, better enabling computers and people to work in cooperation. Adding on, the Semantic Web aims at machine-processable information enabling intelligent services such as information brokers, search agents and information filters, which offer greater functionality and interoperability (Decker, et al., 2000). Since then the technology has moved forward significantly and has opened the possibility of sharing and combining information in more efficient way.

The association of knowledge with Semantic Web has provided a scope for information management through the knowledge management. Since both the technologies use ontology to conceptualize the scenarios, Semantic Web technology could provide a platform for developments of knowledge management systems (Stojanovi & Handschuh, 2002). We believe the framework has adopted the knowledge technologies as sub technologies within. The ontologies are core underlying knowledge technologies within this Semantic Web framework. These ontologies are defined through XML based languages and the advanced forms of them.

The major context behind this project is the use of knowledge in order to manage huge sets of heterogeneous dataset in a Web based environment. It primarily focuses on the spatial dataset and its management through the available spatial technologies incorporated within the knowledge technology. As the Web technologies get matured through the Semantic Web, the implementation of knowledge in this domain seems even more appropriate. This research paper puts forward the views and results of the research activities within the backdrop of the Semantic Web technologies and the underlying knowledge technologies.

1.2 Knowledge representation and ontologies

Knowledge representation has been described in five distinct roles it plays in (Davis et al., 1993). Those roles are

- A surrogate for the thing itself used to enable an entity to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than acting it.
- A set of ontological commitments, i.e., an answer to the question: In what terms should I think about the world?
- A fragmentary theory of intelligent, reasoning, expressed in terms of three components
  - The representation's fundamental conception of intelligent reasoning
  - The set of inferences the representation sanctions; and
  - The set of inferences it recommends

- A medium for pragmatically efficient computation, i.e., the computational environment in which thinking is accomplished.
- A medium for human to express, i.e., a language human expresses things about the world.

Semantic Web technologies use these roles to represent knowledge. The first and the last roles are primarily theoretical roles through which knowledge could be better understood. The remaining roles are conceptual roles which are being implemented within the technology. If those roles are carefully evaluated, it could be seen that knowledge representation begins with ontological commitments. That is selecting a representation means and making a set of ontological commitments (Brachman et al., 1978). Thus defining ontology is a major activity with the process of the Semantic Web.

The term Ontology is being used for centauries to define an object philosophically. The core theme of the term remains the same in the domain of computer. Within the computer science domain, ontology is a formal representation of the knowledge through the hierarchy of concepts and the relationships between those concepts. In theory ontology is a formal, explicit specification of shared conceptualization (Gruber, 1993) In any case, ontology can be considered as formalization of knowledge representation and Description Logics (DLs) provide logical formalization to the ontologies (Baader et al., 2003).

Description logics (DLs) [(Calvanese et al., 2001); (Baader & Sattler, 2000)] are a family of knowledge representation languages that can be used to represent knowledge of an application domain in a structured and formally well-understood way. The term "Description Logics" can be broken down into the terms description and logic. The former would describe the real world scenario with the real world objects and the relationships between those concepts. More formally these objects are grouped together through unary predicates defined by atomic concepts within description logics and the relationships through binary predicates defined by atomic roles. The term logic adds the fragrance of logical interpretations to the description. Through these logics one could reason the description for generating new knowledge from the existing one.

As the Semantic Web technologies matured, the need of incorporating the concepts behind description logic within the ontology languages was realized. It took few generations for the ontology languages defined within Web environment to implement the description language completely. The Web Ontology Language (OWL) (Bechhofer, et al., 2004); (Patel-Schneider et al., 2004)] is intended to be used when the information contained in documents needs to be processed by applications and not by human (McGuinness & Harmelen, 2004). The OWL language has direct influence from the researches in Description Logics and insights from Description Logics particularly on the formalization of the semantics (Horrocks et al., 2004).

The horn logic more commonly known the Horn clauses is a clause with at most one positive literal. It has been used as the base of logic programming and Prolog languages (Sterling & Shapiro, 1994) for years. These languages allow the description of knowledge with predicates. Extensional knowledge is expressed as facts, while intentional knowledge is defined through rules (Spaccapietra et al., 2004). These rules are used through different Rule Languages to enhance the knowledge possess in ontology. The Horn logic has given a platform to define Horn-like rules through sub-languages of RuleML (Boley, 2009).

Summarizing, it could be said that ontology defines the data structure of a knowledge base and this knowledge base could be inferred through various inference engines. These inference engines can be perform under Horn logic through Horn-like rules languages.

Semantic Web technology is slowly revolutionizing the application of knowledge technologies. Knowledge technologies though existed before the Semantic Web, the implementation in their fullness is just being realized. This research benefits from the existing inference engines through the inference rules and reasoning engines to reason the knowledge. However in current stage, the research works moves beyond semantic reasoning and semantic rule processing and attempts to integrate the spatial reasoning and spatial rule inference integrating spatial components in its structure. This research project introduces the approach on achieving the spatial functionalities within those inference engines.

1.3 Spatial components in semantic web

The Semantic Web technologies is slowly gaining acceptance in the wider community. It is thus paramount to include every type of information within the technology. The core within Semantic Web technologies is the semantics of the resources. These semantics may be the spatial or non-spatial. However, the focus of the technology is mainly on utilizing the non-spatial semantics for managing the information. Thus, the spatial information is widely neglected. Nevertheless, it has been realized inclusion of spatial components within Semantic Web framework is important for way forward. Those researches mainly focus on semantic interoperability of spatial data for efficient exchange of spatial data over heterogeneous platforms or efficient data integration. In cases like (Cruz, et. al 2004; Cruz, 2004), the ontologies are used to map their concepts to a global concept within a global ontology and thus providing a common platform for data integration. This is a common trend of practice for managing heterogeneous data source through Semantic Web technologies. The same practice is applied for geospatial data sources. In other cases like (Tanasescu, et al., 2006), ontologies are used to manage the semantics within different data sources to maintain the semantic interoperability of spatial data within different platforms.

In the realm of geospatial and temporal concepts and relationships, the work has not yet reached a level of either consensus or actionability which would allow it to be basis of knowledge interoperability (Lieberman, 2007). The Open Geospatial Consortium (OGC) is playing a major role to develop a consensus among different stakeholder on various aspects of geospatial technologies. The 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 reported different spatial ontologies that exist in the Web (Lieberman et al., 2007).

It is evident that the geospatial ontologies are developed to solve individual spatial problem and are not being used to be effective for knowledge formulation within the Semantic Web framework. Existing ontologies or the ontologies in the process of creation are mostly targeting the usage of vocabularies for the proper data management and not the knowledge management. One implication of such approach is that there is no possibility of geospatial reasoning to enhance the knowledge base. It is widely noticed there is the lack of a known, robust geospatial reasoners. Furthermore, it has been argued that while geospatial reasoning is an ever-evolving field of research, spatial data constructs are not yet accommodated within most current Semantic Web languages as the OWL language (Reitsma & Hiramatsu, 2006).

The seamless integration of spatial components within Semantic Web technologies is the major topic of this research project. Hence, the approach in which this component is integrated within the global framework of the Semantic Web technology is covered extensively within this research project. Additionally, it discusses different components involved in spatial activities within the framework.

1.4 Spatial components on database systems

It has been seen that in the previous section that the ontology engineering has not gained enough momentum to assist spatial activities through ontology. Hence, this project work utilizes the existing potentiality of spatial extensions within the current database system to carry out the spatial activities within the ontology.

Most of the database systems support spatial operations and functions through their spatial extensions. Over the past decade, as Relational DataBase Management System (RDBMS) has seen a huge growth in the database technology. Likewise, the spatial components within them also seen a tremendous improvement in their functionalities. In early days, spatial data were organized in dual architectures which consist of separate administrative data for data management in a RDBMS and spatial data for a GIS system. This could easily result in data inconsistency hence all the database systems today maintain the spatial component in a single RDBMS.

In order to have a common standard among different database systems, they implement their spatial performance accordance to the Open Geospatial Consortium (OGC 1998) Simple Features Specifications for SQL (OGC 1999). Since OGC Simple Feature Specifications are built within simple spatial features in 2D space, most of the spatial operations are restricted to 2D spatial data. It is also possible to store, retrieve and visualize 3D data but it does not follow OGC simple feature specifications. Some RDBMS system today also supports certain 3D spatial queries as well.

According to OGC specification any object is represented spatially following two structures – geometrical and topological. The geometrical structure is the feature providing the direct access to the coordinates of the objects. The topological structure provides the information about the spatial relationships of the objects. The database systems store the geometrical information of the objects and not their topology. They then use their spatial operations to retrieve topological relationships between these geometries (Hellerstein et al., 1995).

1.5 Aims and the motivation of the project

It is a general fact that technologies always shift for the betterment and the components of the previous technologies must be upgraded to the shifting technology. The world is experiencing a shift in technology from the database oriented Information technology to ontology oriented knowledge technology and thus each individual technology that have matured under previous technology requires to be shifted to this emerging technology. The tasks of shifting these components have always presented challenges as the principle foundations between the two technologies are entirely different in most of the cases.

One of the major technical components in the database oriented technologies is the spatial technology. The immense strength of spatial technology was realized long before the emergence of database or even the computers. Maps were used to analyze the problems and derive solutions spatially (Berry, 1999). With the evolution of computers, a new discipline emerged to analyze the problems spatially, which is termed as Geographic Information System (GIS). GIS technology was one of the first to use the spatial technology for the analysis of the geographic locations. Spatial analysis is used in other domains too besides GIS. Before the emergence of sophisticated database systems, GIS technologies used files to store the spatial data. Each vendors of the technology had their own algorithms for spatial operations and functions. This in turn provided lots of inconsistency in the analysis process. As the database technology matured, it started to include those spatial components into it. In this manner, the spatial technology got immersed within the database technology. As previously mentioned they followed the specifications provided by OGC to maintain a common standard and hence most of these inconsistencies were revolved. With the advancement in database systems the spatial technology also got matured and today it is not necessary to depend on a GIS to perform spatial analysis. This has clear advantages for the other domains which use spatial analysis as part of their analysis process.

When viewed from the Semantic Web point of view, the integration of spatial component will trigger the integration process of other data component adding an open layer for data type which could be argued as non-typically semantic within its framework. This data could be spatial or temporal data or even process data. Such level within the technical framework of Semantic Web will give clear advantages for the technology to grow.

The main aim of this research project is to initiate the process of setting up a layer in the Semantic Web framework for the non-typical semantic information that is not covered through the semantics. In order to illustrate its applicability, this research centers on integration of spatial component within the Semantic Web technologies. This work focuses beyond data interoperability and addresses the spatial processing through knowledge querying and inferring. In addition, the work attempts to change and to improve the ongoing data management process of archiving documents in the industrial archaeology domain through knowledge management process. This work also aims to initiate the usage knowledge for performing spatial analyses in the existing GIS tools. It tries to draw attention towards the benefit of introducing a knowledge level in the universal GIS model. This in fact supports the relevancy theory of the need to transfer the technical component in the wake of technology change.

1.6 Industrial archaeology: the case study

The research project is drawn around the case study of industrial archaeology. The discipline of industrial archaeology fits perfectly to demonstrate the effectiveness of the implementation of the research activities. In general the industrial archaeological sites are available for very short duration of time and the amount of information collected is huge and diverse making it impossible for the conventional technologies to manage them. This research takes on the Semantic Web and its underlying knowledge technology to manage them. The knowledge possess by archaeologists is used to identify the objects and map the data and documents to the respective objects. In this process the knowledge about the objects is acquired through first identifying the objects and defining their behavior at the ground. This knowledge can then be used during the management of these objects. In fact the research project is based on 4Ks processing steps: Knowledge Acquisition, Knowledge Management, Knowledge Visualization and Knowledge Analysis. In each of these 4Ks, the knowledge of archaeologists is used.

The research site lies in Krupp belt Essen. This 200 hectares site was used for steel production in early nineteenth century but was later destroyed. The majority of the area was never rebuilt. The site was excavated in 2007 in order to document the findings. The area is being converted to a park of the main building of ThyssenKrupp so there was not much time available to document the findings properly due to ever changing structure of data and documents and their volume. It is hence not possible to use the traditional technology for their rigid nature and huge dependency on human manipulation of the data and documents. Possibility to engage machine to understand the information and processed them through the collaboration of the knowledge possess by archaeologist was realized through an application tool – The Web platform ArchaeoKM.

The research highlights the importance of non-typical semantic information within the Semantic Web framework. The research discusses the possibility of including spatial technology within the framework. A layer is proposed for spatial data pattern that utilizes the Semantic Web component to process spatial knowledge. This layer could host other data patterns as well and follow the same trend of spatial integration.

The integration of spatial technology within the Semantic Web technologies adds up benefit to the geospatial community. Instead of depending on the information based on the data, the analysis process should be more efficient and less demanding through the application of knowledge. The approach of using knowledge that is supported by underlying spatial data to execute the analysis process was embraced by the research.

The paper is divided into four major sections with chapters discussing them. The first chapter introduces the domain of the case study and how the existing technologies are contributing on it. It mainly discusses these issues with backdrop of ArchaeoKM – an application tool developed during the research work. The next section covers the state of art and basically highlights the state of the art in underlying knowledge technologies within the Semantic Web technology and their relations to this research work. The fourth chapter points out the possibilities of spatial extension in the knowledge technologies thus proposing a separate level dedicated to geospatial integration within the Semantic Web framework. The paper concludes with the conclusion where an effort has been made to argue that there is vast implementation of spatial extension and the benefit could be realized in third party domain as shown within the case study domain.

2. The ArchaeoKM project

This chapter begins with a discussion about a general overview of the industrial archaeology. It presents the case study of the research site by discussing the nature of the data collected during the excavation process. It then reviews the current Information Systems that are either being implemented or researched in this domain. It includes the usages of Geographic Information Systems (GIS) in this field. Then after, the chapter continues with the introduction of the ArchaeoKM project through discussion on the principle and how it is different from the existing systems. It concludes with a discussion on the future prospective of the work.

2.1 The domain of the industrial archaeology: a case study

The domain of the Industrial archaeology is the recording, study, interpretation and preservation of the physical remains of the industrially related artifacts, sites and systems within their social and historical contexts (Clouse, 1995). During the period of 18th and 19th century the industrial revolution started from the United Kingdom and spread across the world marking a major turn of the human civilization. In the course of time the industries established during the period were abandoned and replaced by new installments. These abandoned sites however hide many important histories of modern developments which need to be preserved as historical facts. Today, the domain of the industrial archaeology has occupied its position in the archaeological community as a mainstream branch of archaeology which deals with the history of constructions, the development architecture, the history of technologies, socio-economic and cultural history (Boochs, 2009). The domain of the industrial archaeology has its own challenges. It does not involve the excavation process and just documents the standing artefacts in contrast to the conventional archaeology, so the discipline was initially considered as hobby archaeology and not a mainstream archaeology. Though the branch has now been taken more seriously by its contemporary branches, it still needs acceptance by the wider community as the awareness about the importance of this field in archaeology is still minimal. The lack of acceptance has its own impact here as there is no reliable tool to document the artifact as the classical archaeology and hence large scale of existing relicts get lost forever. Usually the industrial archaeological sites are available for limited amount of time as they are not mostly conserved for continuous excavation and they are most often the sites for new constructions. Adding on, the advancement of current data capturing technologies made it possible to capture huge and heterogeneous datasets in this limited duration. It is absolutely not possible to manage this nature of datasets in such a limited amount of time without the intervention of machine to assist human. It thus requires human machine collaboration to manage them which is not possible through the conventional technologies.

The project points out these limitations and provides a prospective solution to handle the dataset through the knowledge possessed by the archaeologists and facilitated by knowledge management tools within Semantic Web technology. This section presents the case study site used within this research work discussing the diversity and amount of data acquired through the modern technologies.

2.2 The main excavation area

The main excavation area lies in Krupp area in Essen belt, Germany. The 200 hectares area was used for steel production during early 19th century. The work on steel production has a critical impact on the settlement development of Essen. In this way the history of Essen is closely related to the activities of steel production in Krupp. The site grew over the decades and formed a so-called Krupp Belt. The site was destroyed during the Second World War. Most of the area is never rebuilt. In between 1945 to 2007, the area was basically a wasteland making it an ideal site for an industrial archaeological excavation. However, the ThyssenKrupp is returning to build its new headquarters in the site by then 2010. This has raised the problem of limitation of time period for a proper management of the recovered objects. The objects are recorded as soon as they are recovered and these records are stored in a repository in their respective data formats. Hence, there is a clear lack of well-defined structure for data management. Moreover, in contrast to the conventional archaeology where the data collection and data analysis goes side by side so in that case the data structure could be designed at the beginning, the data analysis is carried out at the end in industrial archaeology so it is not possible to perceive the structure of the data at the beginning. The first challenge consists of creating a proper data structure which helps in retrieving those data efficiently. As there was not enough time to filter the collected data concurrently, the amount of data that are collected is huge. Hence, the system that has to handle the collection of data should be able to handle this huge set.

Archaeologists with assistance of photogrammetric specialists were involved in data acquisition process. They were responsible to decide the methods of measurements. The findings were scanned through terrestrial laser scanning instruments. Two scanners were used to acquire the scanned data. They were the Zöller and Fröhlich scanner (ZF) and the Riegl scanner. Those two scanners were used according to their requirement. Large objects scanning were carried out with the help of the Riegl scanner whereas the ZF scanner is used whenever some important findings are recovered. The Riegl scanner was installed on the roof of the Kreuzhaus (the building marked at the bottom of the site in figure 1) so that the scanner gets a good overview of the area. The findings were scanned with a resolution of 0.036 degrees (6 mm on 10 m) hence the point cloud is very dense. All the data were stored in the Gauß Krüger zone II (GK II) coordinate system.

An orthophoto was orthorectified from the aerial images (that were taken during the course of research work). The orthophoto has 10 cm resolution and is in GK II coordinate system. Huge numbers of digital pictures were taken during the research activities and they were stored in their original formats. These photos were taken with non-calibrated digital cameras. However, certain knowledge can be extracted from them by the archaeologists. Besides, photographs documents like the site plan of the area and some documents with relevant information of the site or the objects recovered were collected during data acquisition process. These data and documents were digitized and stored for proper mapping with the relevant objects. Archaeological notes taken by archaeologists during these excavation processes are of high importance. Hence, these notes are digitized and stored in the repository. Similarly, the site plan of the area was digitized and stored as .shp format in ArcGIS.

The nature of the dataset that was collected during the research work is varied. There are four distinct kinds of data which ranges from textual documents as the archaeological notes to multimedia documents as images. The heterogeneity of dataset is evident through the nature of each type of dataset varying completely from others in terms of their storages, presentations and implementations.

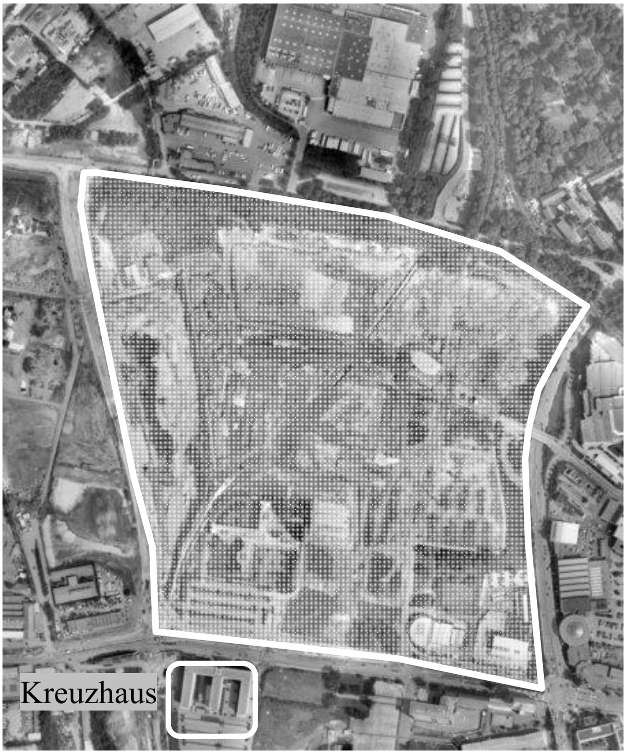

Fig. 1. The main excavation area Site.

2.3 GIS for archaeology

What does a GIS do? Basically providing a definition of GIS and referring to its abilities to capture and manipulate spatial data doesn't provide much insight into its functionality. The basic tasks of a GIS system can be broken down into five groups, data acquisition, spatial data management, database management, data visualization and spatial data analysis (Jones, 1997). Most archaeological data such as artifacts, features, buildings, sites or landscapes, have spatial and aspatial attributes that can be explored by GIS. These attributes include the spatial location that informs about the local or global context concerning the pieces of information, and the morphology that defines the shape and the size of an object.

The acquisition of spatial data is undertaken with the help of existing digitizing functionalities within the application software providing them. They are responsible for the acquisition of data and integrating it to the existing spatial sets. Spatial data include, but are not limited to, topographic maps, site locations and morphology, archaeological plans, artifacts distribution, aerial photography, geophysical data and satellite imagery.

The spatial data management process uses sophisticated database management systems in order to store and retrieve spatial data and their attributes. Data collected from different sources have to be transformed in the same coordinate system in order to integrate them.

The database management system, involving conceptual and logical data modeling is an important part of GIS because it ensures that the construction and the maintenance of database is done and that the spatial and aspatial datasets and components are correctly linked.

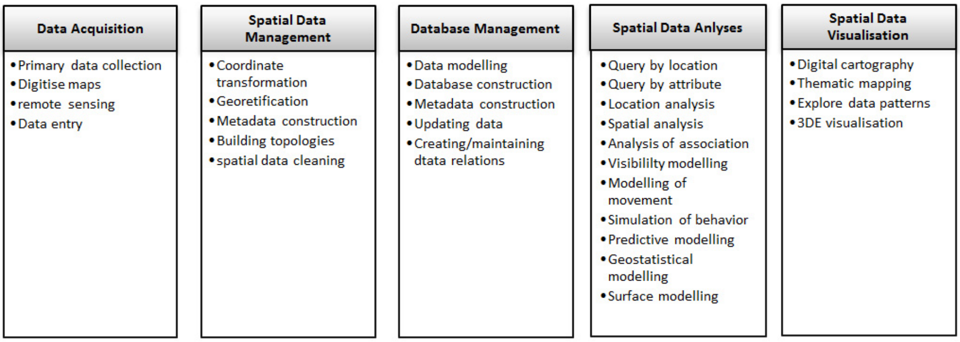

Fig. 2. The five main groups of tasks performed by GIS.

Some limitations appear visible in currents GIS system in the context of the Industrial Archaeology. The lack of GIS platforms that uses data like point cloud is one of such visible limitations. It however is a fact that conventionally an Information System for archaeologists is a Geographic Information System or 3D object modelling system. The statement has been supported by the current commercial applications for the archaeologists. Applications like ArchaeoCAD from ArcTron and PointCloud from Kubit rely heavily on the geometry of the objects excavated. The applications are thus used primarily to represent objects excavated in a 3D space. Similarly, GIS vendors like ESRI uses the spatial information of the objects to analyze them spatially. Meanwhile, the data collection process has seen a tremendous change in the last few years. Today, it is not only the amount of data that needs consideration, the diversity of data should also be taken into account. It is becoming increasingly difficult to manage them solely with the current database system due to the size and diversity of the data. In addition, information systems in archaeological projects or cultural heritage projects lacks from a complete package. There have been lots of researches going on but they are on the independent components. However, research projects like 3D MURALE (Cosmas et al., 2001) and GIS DILAS (Wüst et al., 2004) contains most of the elements needed for a complete package and hence could be considered as comprehensive Information System. The 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 the 3D space. These 3D models are displayed in the visualization component. The DILAS is generic software, fully object oriented model for 3D geo-objects. The 3D geometry model is based on a topological boundary representation and supports most basic geometry types. It also incorporates the concept of multiple levels of detail (LOD) (Balletti et al., 2005) as well as texture information. It is thus clear that the existing systems rely heavily on the geometries of excavated objects for their representations, but the interoperability of these systems and the knowledge sharing remains a gap.

In addition, the sharing of knowledge in archaeology and disseminate it to the general public through wiki has been discussed in (Costa & Zanini, 2008). Likewise the use of knowledge to build up a common semantic framework has been discussed in (Kansa, 2008). Research works in data interoperability exist in the field of archaeology, but most of the research is carried out in other related fields. However, it could be applied in archaeology as well. The existing researches focus more on using the common language for efficient interoperability. The research project (Kollias, 2008) concerns the achieving syntactic and semantic interoperability through ontologies and the RDF framework to build a common standard. Data integration through ontologies and their relationships is discussed in (Doerr, 2008). Although the work on the Semantic Web and knowledge management in the field of Information System in Archaeology or related fields is stepping up with these research works, the fact is they are in very preliminary phases. Additionally, these projects concentrate more on how to achieve interoperability with semantic frameworks and ontologies. However, none of them focuses on the knowledge generation process and more specifically on rules defined by archaeologists in order to build up the system which should use, evaluate and represent the knowledge of the archaeologists.

Knowledge contained in documents has been traditionally managed through the use of metadata. Before going on details about knowledge management, let us first understand the perspective about the whole idea. Every activity begins with data. However data is meaningless until they are put in context of space or an event. Additionally, unless the relationship between different pieces of data is defined, simply data do not have any significance. Once the data are defined in terms of space or events and are defined through relationships, they become Information. Information understands the nature of the data but they do not provide the reasons behind the existence of data and are relatively static and linear by nature. Information is a relationship between data and, quite simply, is what it is, with great dependence on context for its meaning and with little implication for the future (Bellinger, 2004). Beyond every relationship, arises a pattern which has capacity to embody completeness and consistency of the relations to an extent of creating its own context (Bateson, 1979). Such patterns represent knowledge on the information and consequently on data. The term Knowledge Management has wide implications. However, very precisely Knowledge Management is about the capture and reuse of knowledge at different knowledge level. In order to access the knowledge, data are annotated and indexed in the knowledge base. This is in line to the concept proposed by Web Semantic where it proposes to annotate the document content using semantic information from domain ontologies (Berners-Lee et al., 2001). The goal is to create annotations with well-defined semantics so they can be interpreted efficiently. Today, in the context of Semantic Web, the contents of a document can be described and annotated using RDF and OWL. The result is a set of Web documents interpretable by machine with the help of mark-ups. With such Semantic Web annotation, the efficiency of information retrieval is enhanced and the interoperability is improved. The information retrieval is improved by the ability to perform searches, which exploit the ontology in order to make inferences about data from heterogeneous resources (Welty & Ide, 1999).

2.4 Towards knowledge processing

The project ArchaeoKM plans to complement the principle of knowledge management by implementing it in the application through formulating the knowledge rules that can be used by archaeologists on the excavated data. The knowledge stored in machine readable format is inferred to bring result which could be well understood by human. Moreover, it moves beyond managing the concepts defined to annotate documents, which most of the research projects currently focusing on, to 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 (a part in an image or a section of archive document) that contains the same object.

One of the main focuses on ArchaeoKM project is to determine an approach of integrating the spatial data within its overall framework of data integration. The integration process did not only serve for the data integration but also has taken a step forward in data analysis and management through the knowledge management techniques.

2.4.1 The web platform ArchaeoKM

The challenges possessed to document the artifacts in such a site could be handled through utilizing the knowledge of responsible archaeologists. The platform ArchaeoKM focuses on the use of the knowledge of archaeologists to document the objects with respect to the surrounding. In the process a tool based on the Semantic Web technology and its underlying knowledge technology was develop to provide the archaeologists to share their knowledge and document the information collected during the excavation process. One of the challenges is to bring all the datasets previously presented in one common platform. As a knowledge representation format, the top level ontology acts as the global schema for data integration in the platform. The application tool provides a common platform for archaeologists to share their experience and knowledge.

2.4.2 The ArchaeoKM architecture

The GIS technology performs a group of five tasks to execute the result. These tasks as already been mentioned are acquisition of spatial data, spatial data management, database management, spatial data analysis and the spatial data visualization. The ArchaeoKM project attempts to complement the five major processing steps of a GIS through its four processing activities which it calls the processing steps of 4Ks: knowledge acquisition, knowledge management, knowledge visualization, knowledge analysis.

The knowledge acquisition task consists in general term defines metadata on data acquired during the survey process. The spatial data acquisition process is still involved during the process, but in addition metadata on these data are defined using a knowledge representation language. Actually, an ontology, which defines the semantic of the recovered features, is defined to capture and capitalized the knowledge of archeologists on the archaeological site. Hence the schema of the ontology is defined at this level. This is done by the help of a specialist on ontologies. The relationships and their semantics are stored into the ontology. This semantic could be provided through an example of the relation of "insideOf" which is transitive relationship. In mathematics, a binary relation R over a set X is transitive if whenever an element a is related to an element b, and b is in turn related to an element c, then a is also related to c by the same kind relation. The ArchaeoKM platform deals with this issue.

The acquisition process constitutes of generation of knowledge base through enriching the ontology. The knowledge of archaeologists is used again to identify the recovered objects and enrich them in the ontology schema formulated. In short the process consists of populating the ontology with "individual" which represent objects recovered from the archaeological site. This creates a knowledge base from the ontology schema.

The knowledge management task consists of storing and the retrieving data along with its semantics. Knowledge is defined through the relationships and it is the relationships between individuals that create the real knowledge in the knowledge base. These relationships not only imply the relations between objects but also relation to their spatial signatures in spatial database. A specialized tool has to be developed in order to retrieve data from the ontology and from its spatial representation stored in a GIS. The ArchaeoKM platform deals with this issue.

The knowledge analysis task is the ability of the system to perform inferences on datasets. This cannot be undertaken without the help of the semantic definition on the archaeological objects. Usually inference or deduction is conducted on attributive data which are defined in the ontology. Today, no tool is defined to compute inference on the individuals of ontology and its spatial definition store in a spatial database. The ArchaeoKM platform deals with this issue.

The knowledge visualization task provides powerful visualization capabilities used for viewing spatial datasets and its semantics counterparts. Tools for the visualization of ontologies are of benefit to visualize the results of knowledge analysis. The ArchaeoKM platform deals with this issue.

As illustrated in figure 3 the system architecture of the ArchaeoKM platform is a three layered architecture with a structure for spatial component standing parallel against them.

The bottom level is the Syntactic level. This level contains all the information recovered from the site. Most of the data and documents collected during the excavation process are stored in their original formats. Certain data which needs to be stored in database system such as GIS data are stored in the RDBMS. This level basically performs as the repository of the dataset. One of the main tasks of the syntactic level is to explain the data. For a proper identification, the data needs to be analyzed with reference to the objects illustrated in the index. One of the first features within the application is the identification process.

A proper identification mechanism allows defining the identified objects. The ArchaeoKM platform utilizes the knowledge of archaeologists to identify the object. The identification is carried out by tagging the objects in the orthophoto of the site provided in the application. Attaching the semantic characteristics through semantic analysis on these objects generates knowledge. Different methods are used for the associating the semantic information according to the data pattern. Three distinct methods are applied to associate the semantic information which depends on the nature of the datasets with which it is associating with: Minimum Bounding Rectangles (MBRs) for the spatial data set, Uniform Resource Identifier (URI) for images and archive data and mapping to the data tables for datasets stored within

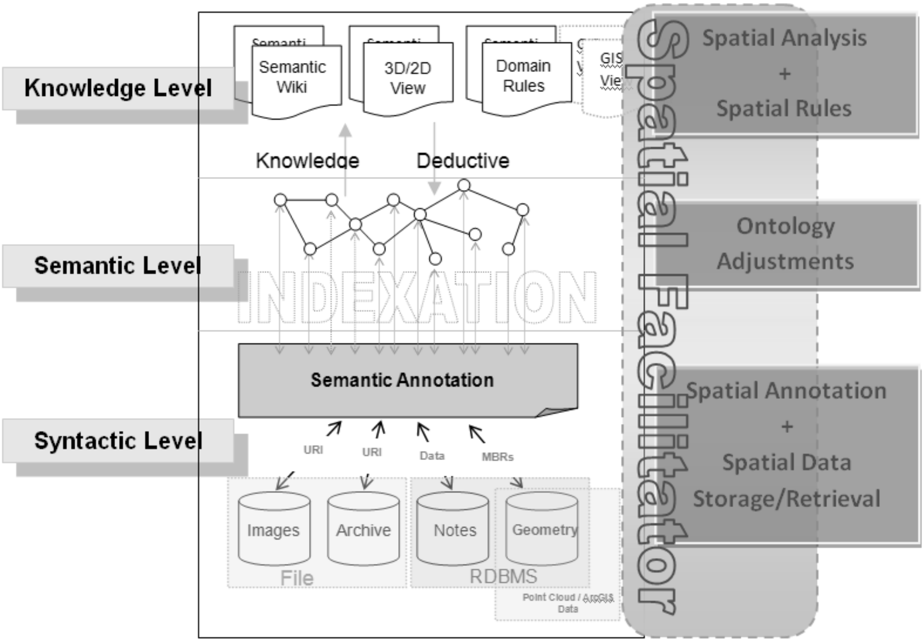

Fig. 3. The system architecture of the ArchaeoKM.

RDBMS. The method is reflected by the feature Semantic Annotation within the platform. These annotations are carried out through creating individual Resource Description Framework (RDF) triplets for each annotation process technology. RDF triplets also map the identified objects to the relevant classes in the domain ontology.

The next level is the semantic level, which manages the extracted knowledge. As stated, it is achieved through the ontological structure established through the descriptions, observations and rules defined by the archaeologists. These descriptions and rules are represented through different axioms in the domain ontology. Archaeologists are involved actively in this phase as they are the one best suited to provide entities and their relationships needed to build up the domain ontology. The semantic annotations from the Syntactic level will be indexed semantically to the entities of the domain ontology in this level. This semantic index through the identification process is the building block of the domain ontology and through semantic annotations provides a semantic view of the data. It also provides a global schema between various data sources making the data integration possible at certain level. 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 is also managed through assigning semantic properties to the objects through proper relationships with other objects.

The top most level is the most concrete one as this level represents the organization of the knowledge on the semantic map through different visualizing tools. This level provides the user interfaces and they are visualized in form of Web pages as illustrated in figure 3. These Web pages represent knowledge which are generated through the knowledge management process discussed above. The pages are interrelated and can be used according to their relevance. The main representation of the knowledge is, however, demonstrated through Detail View pages. These pages are not only designed to illustrate the knowledge that has been generated and to manage it through the bottom two levels, but to also perform semantic research in order to gain new knowledge. Various techniques of the Semantic Web technology are being integrated within ArchaeoKM structure for acquiring new knowledge. Domain rules through inference engine provide one of those features in ArchaeoKM structure. In archaeology it is sometimes not possible to analyze the finding immediately and needs some properties or relationships to support them later. These inference rules provide the archaeologists such functionalities within the application.

In addition to the three levels, the system architecture contains components that facilitate the acquisition, validation, upgrade, management and analysis of the spatial knowledge. These components are packaged into the Spatial Facilitator as illustrated in figure 3. This component is responsible for analyzing the spatial data and providing results; either to update the current ontological structure in the semantic level or to populate the knowledge base. Through the inference capabilities in Semantic Web technology, new theories could be explored.

2.5 Discussion

This chapter has presented the case study of industrial archaeology for implementation of the arguments the research proposes. Industrial archaeology is the best suited for the research for the nature of the domain. The discipline of industrial archaeology generates huge and diverse data in very short duration of time and amount of time is short making it not possible to manage information through the conventional technologies. It is thus apprehending that this huge and diverse information could only be managed through active involvement of the archaeologists and the knowledge possess by them.

The ArchaeoKM project uses the knowledge possessed by the archaeologists to manage the information they gathered during the excavation. It is handled through a platform based on Semantic Web technologies and knowledge management and is termed as ArchaeoKM itself abbreviating Archaeological Knowledge Management. It uses the processing steps of 4Ks representing knowledge acquisition, knowledge management, knowledge analysis and knowledge visualization complementing the five steps of a GIS process. These 4Ks processing steps use the knowledge of the archaeologists in manipulating the data and to manage them.

This chapter establishes a relation between the case study of industrial archaeology and the spatial knowledge modeling paving the direction of the research. Primarily based on knowledge management of Semantic Web framework, it uses the spatial nature of case study to implement the spatial tools provided by the current spatial technology within the Semantic Web framework. The capabilities in existing tools to use the current database systems and their spatial extension are evident of the ability of database systems to manage spatial data. It however lacks the flexibility to adapt itself into new scenarios that might arise through generation of new information or changes in the contexts. This research carries these capabilities forward by using the spatial knowledge processing through knowledge tools which provides the proper data management in archaeology that addresses the limitation in adaptation of the conventional technologies.

This chapter has presented the concept of the inclusion of spatial knowledge in handling the spatial nature of data recovered. This is new domain of research and probably one of its kinds. Hence it is important to understand the current state of art in both spatial and Semantic Web technologies. The next chapter thus discusses the state of art in the Semantic Web.

3. The Semantic Web

The World Wide Web (WWW or the Web) is the single largest repository of information. The growth of Web has been tremendous since its evolvement both in terms of the content and the technology. The first generation Webs were mainly presentation based. They provided information through the Web pages but did not allow users to interact with them. In short they contained read only information. Moreover, the early pages were text only pages and do not contain multimedia data. These Web sites have higher dependency on the presentation languages as Hypertext Markup Languages (HTML) (Horrocks et al., 2004). With the introduction of eXtensible Markup Language (XML), the information within the pages became more structured. Those XML based pages could hold up the contents in more structured method but still lack the proper definition of semantics within the contents (Berners-Lee T., 1998). Needs of intelligent systems which could exploit the wide range of information available within the Web are widely felt. Semantic Web is envisaged to address the need. The term "Semantic Web" is coined by Tim Berners-Lee in his work (Berners-Lee et al., 2001) to propose the inclusion of semantic for better enabling machine-people cooperation for handling the huge information that exists in the Web.

The term "Semantic Web" has been defined numerous time. Though there is no formal definition of Semantic Web, some of its most used definitions are.

The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. It is a source to retrieve information from the Web (using the Web spiders from RDF files) and access the data through Semantic Web Agents or Semantic Web Services. Simply Semantic Web is data about data or metadata (Berners-Lee et al., 2001).

- A Semantic Web is a Web where the focus is placed on the meaning of words, rather than on the words themselves: information becomes knowledge after semantic analysis is performed. For this reason, a Semantic Web is a network of knowledge compared with what we have today that can be defined as a network of information (Mazzocchi, 2000).
- The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise and community boundaries (Herman et al., 2010).

Any information systems which have to interoperate with various other information systems have to face the problem of interoperability. The archaeological community has seen the tremendous change in the manner the data are collected and manipulated. In one hand the technology growth provides the added functionalities to handle information which archaeologists cherish but at the same time they provide heterogeneity in the information pattern. The differences in manners and methods of individual community with the archaeological domain have led to development of independent systems and this has contributed in data incompatibility. A platform providing interoperability between different systems and in particular different sets of information has been widely felt within the community. Actually, the data heterogeneity is the main issue when the time comes to exchange and to manage information that describe the real world.

The issue of interoperability has always been there in the field of Information Technology ever since the computer systems started to communicate with each other through various modes. Factors like data authority, system autonomy and data heterogeneity are involved in the concerns of achieving efficient interoperability among different information systems. During the initial stages of the technology when a system was restricted to a department or at most a company, the issue of interoperability was limited within departments of a company. Hence the concern of data authority was not a big issue. However, the involvement of different departments and with them different players raised the issue of data heterogeneity. The evolvement of database management system (DBMS) fuelled up the necessity for data interoperability. Different underlying issues needed to be considered for achieving data interoperability in database systems like the structural differences, constraints differences or the difference in query languages. These information systems are based on DBMSs and hence the efficiency of system interoperability depended on tackling the question of heterogeneity of underlying data models of these DBMSs. As data models are represented through their schemas, the most common approach was to compare the schemas of the DBMSs and convert a schema of a DBMS to the next DBMS. Other approaches like building up a common model which acts as a broker to interchange the data between different DBMSs were also preferred to achieve the interoperability. In short, the first generation problem of data interoperability was mainly due to the fact of the differences in technical issues such as structures, constraints and different techniques. These problems are short term problems as they could be sorted out with a broker technology mediating between different technical approaches. The main problem of interoperability arises when there is a difference in understanding. The semantic differences between information fuel up the interoperability issues as the information gets more accessible and easy to use.

The next generation of systems saw gradual acceleration in the data types which are not necessarily structured. Those kinds of data could be semi-structured data or digital data like multimedia data. During this period data like geospatial data or temporal data got more acceptances within structured data community expanding the horizon of structured data. The influx of tailored made software applications for these kinds of data has raised the arguments of interoperability in much stronger manner. To add this there is the rapid growth of Internet technology and rapid growth in tendency to depend on internet for information. The information is thus distributed through various systems with their independent methods of developments and presentations. The issue of interoperability revolved around factors like technology for dealing heterogeneous systems with different data structures and patterns or handling the semantic interoperability through handling the difference in terminologies (Sheth, 1999). The necessity to have a common understanding of the information led to the concept of inclusion of some form of semantics to represent them. Metadata provided the semantic representations to the information. Metadata is data of data which provides information about the data in terms of their creation, storage, management, authority and in certain term their intended purpose. Metadata became essential part of any reliable information source and a medium to maintain interoperability. Likewise the trend to have standardization or adoption of ad hoc standards made significant progress towards achieving system, syntactic and structural interoperability (Sheth, 1999).

The current generation has followed the previous trend of heterogeneity in the data source and has carried it even further. The users have become more sophisticated in using this information. They expect the system to help them not at the data level but at the information and increasingly knowledge level (Sheth, 1999), thus expecting to have interoperability at the semantic level. Though metadata provides certain level of semantics for the data, they are generally not enough for managing the ever exploding information. The contexts of information needs to be taken into account to understand the information and these contexts are managed through the ontologies as traditionally they are built for specifying the vocabularies and their relationships. The underlying semantics in ontology provides foundation to interpret the knowledge within. This has provided a huge boosting in achieving interoperability between systems. The use of knowledge to understand information between systems and find a common linkage between them provides a framework for the interoperation. The issue of interoperability which started with technical differences has come to difference in understanding. The technical differences in dealing with interoperability is long been exercised but the semantic differences has come in a big way. It became even bigger issue with the amount of information that is available today. The problem could be tackled with resolving the differences in understanding of information. So a form of semantic mapping can address such issues of understanding.

Web 3.0 aims to make computers understand semantics behind information. This would make them intelligent to process information and deliver the required knowledge. It could be argued that the information when encapsulated by semantics would provide knowledge. The relationship between Web 3.0 and Semantic Web is a topic of argument. There are suggestions that they are the same whereas some argue that Semantic Web is a sub-set of Web 3. Sir Tim Berners-Lee has described Semantic Web as a component of Web 3.

> "People keep asking what Web 3.0 is. I think maybe when you've got an overlay of scalable vector graphics - everything rippling and folding and looking misty - on Web 2.0 and access to a Semantic Web integrated across a huge space of data, you'll have access to an unbelievable data resource"
>
> Tim Berners-Lee, 2006, (Shannon, 2006)

This chapter covers different features of the Semantic Web.

3.1 The knowledge base

Description Logics supports serialization through the human readable forms of the real world scenario with the classification of concepts and individuals. Moreover, they support the hierarchical structure of concepts in forms of subconcepts/superconcepts relationships of a concept between the concepts of a given terminology. This hierarchical structure provides efficient inference through the proper relations between different concepts. The individual-concept relationship could be compared to instantiation of an object to its class in object-oriented concept. In this manner, the approach DL takes can be related to classification of objects in a real world scenario.

Description logics provide formalization to knowledge representation of real world situations. This means, it should provide the logical replies to the queries of real world situations. This is currently most researched topic in this domain. The results are highly sophisticated reasoning engines which utilize the capabilities of expressiveness of DLs to manipulate the knowledge. A Knowledge Representation system is a formal representation of knowledge described through different technologies. When it is describe through DLs, they set up a Knowledge Base (KB), the contents of which could be reasoned or infer to manipulate them. A knowledge base could be considered as a complete package of knowledge content. It is however only a subset of a KR system that contains additional components.

Figure 4 (Baader & Nutt, 2002) sketches the architecture of any KR system based on DLs. It could be seen the central theme of such a system is a Knowledge Base (KB). The KB constitutes of two components: the TBox and the ABox.

TBox statements are the terms or the terminologies that are used within the system domain. In general they are statements describing the domain through the controlled vocabularies. For example in terms of a social domain the TBox statements are the set of concepts as People, Male,Female, Father, Daughter etc. or the set of roles as marriedTo, siblingOf, sonOf, hasDaughter etc. ABox in other hand contains assertions to the TBox statements. For example Ashish is a Male is an ABox statement. In object oriented concept ABox statements compliant TBox statements through instantiating what is equivalent to classes in TBox and relating the roles (equivalent to methods or properties in OO concept) to those instances. The DLs are expressed through the concepts and roles of a particular domain. This complements well with the fact how knowledge is expressed in the general term. Concepts are sets of classes of individual objects. Classes provide an abstraction mechanism for grouping resources with similar characteristics (Bechhofer, et al., 2004).

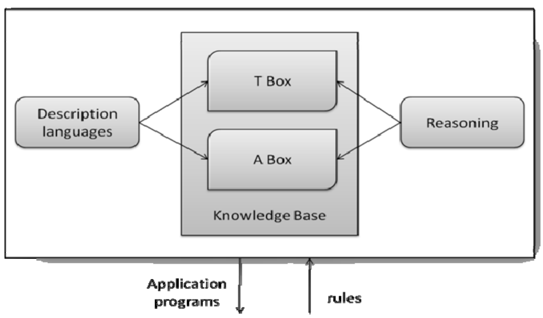

Fig. 4. The Architecture of a knowledge representation system based on DLs.

The concepts can be organized into superclass-subclass hierarchy which is also known as taxonomy. It shares the object-oriented concepts in managing the hierarchy of superconcept-subconcept. The subconcepts are specialized concepts of their superconcepts and the superconcepts are generalized concepts of their subconcepts. The subsumption algorithm determines the superclass-subclass relationships. For an example all individuals of a class must be individuals of its superclass. In general all concepts are subsumed by their superclass. In any graphical representation of knowledge concepts are represented through the nodes. Similarly the roles are binary relationship between concepts and eventually the relationships of the individuals of those concepts. They are represented by links in the graphical representation of knowledge. The description language has a model-theoretic semantics as the language for building the descriptions is independent to each DL system. Thus, statements in the TBox and in the ABox can be identified as first-order logic or, in some cases, a slight extension of it (Baader & Nutt, 2002).

3.2 The Semantic Web stack

The Semantic Web stack also called the Semantic Web cake is basically a hierarchy of the technologies composed of different layers. Each layer takes advantages of the capabilities concerning all the sub-layers. The following figure 5. illustrates the Semantic Web cake.

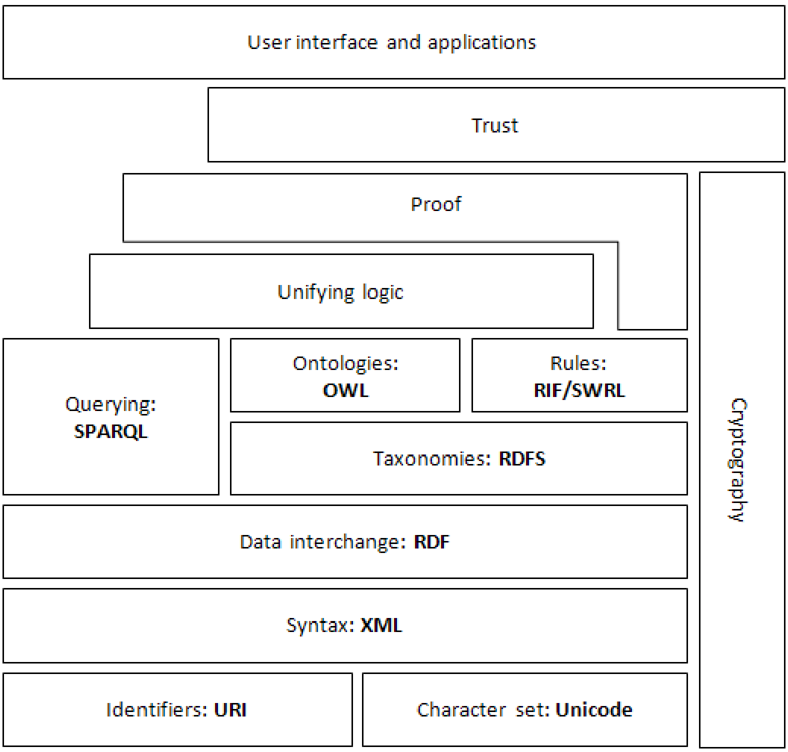

Fig. 5. The Semantic Web Stack.

There is a degrees of uncertainty in which the Semantic Web cake is defined. There are four versions of this cake till date, and none of them have been published in the literatures. All the four versions are presented by Berners-Lee in his presentations (Gerber et al., 2008). The components and their relationships are hence been defined profoundly. It is thus necessary to isolate each component and discuss their role in terms of the Web. The definitions of some layers within the semantic cake are illusive and could be interpreted in many ways. However, some layers especially the lower layers have clear definitions. Here, these hierarchical layers are discussed in terms of the knowledge representation approach. The layers in the Semantic Web stack can generally be divided into three categories: syntactic layers, knowledge layers and certifying layers composing of different technologies to support the technology.

The bottom layers are information holding layers and are either presented in uniform language or the through XML based information. The components within this layer hold the technologies that are direct descendent technologies from the hypertext Web. Though they are carryovers from basic technologies, they provide strong base to the Semantic Web. The technologies within these layers present syntactical representation of the information and thus be grouped into one common category of syntactic layers. They are capable to hold huge amount of information in each of the individual technologies within the level. These technologies include basic technologies as URI or content based technologies as XML and RDF. Despite rich with contents they lack interpretations as they do not possess semantics within.

The middle section contains layers which represents knowledge. These layers generally represent the technologies standardize by W3C for processing knowledge and can be grouped together to knowledge layer. The technologies here utilize the syntactically rich technologies in layers beneath. The knowledge is generated through attaching semantics to the information. RDFS provides vocabulary to RDF thus providing semantics to the structured statements representing the information as triplets. Through RDFS technology it is possible to derive hierarchical representations of objects and relate the objects to each other. The technology bridges the gaps between syntactically rich contents and tools to interpret knowledge from these contents. RDFS can define ontologies. Ontologies play important roles in order to provide semantics to the information or to the contents by providing suitable vocabulary to the contents and uplift contents to resources which could be related to real world objects. As a result of the work of the W3C Web Ontology Working Group, the "Ontology" layer has now been instantiated with the Web Ontology Language (OWL (Smith et al., 2004)) (Horrocks et al., 2005) due to its extended constructs to describe the semantics of the RDF statement. The semantic within the ontologies and expressed through OWL can be used within the ontologies and the knowledge bases themselves for the inferences. However, in order to express the rules independent to the languages two standards are emerging in the form of RIF (Boley & Kifer, 2010) and Semantic Web Rule Language (SWRL) (Horrocks et al., 2004). The rules are supported through inference engines. Simple Protocol and RDF Query Language (SPARQL) (Prud'hommeaux & Seaborne, 2008) is SQL equivalent language for querying data stored as RDF resources. As OWL is basically written in RDF pattern so the query could be applied to it as well. The topmost layer within knowledge layer is the unifying logic layer. This layer provides the logic behind knowledge manipulation through the reasoning capabilities of reasoning engines. This layer has not been formally defined so subjected to certain degrees of manipulation.

The top two layers in the stack are not yet fully conceptualized in terms of their applicability. These layers contain technologies which are not standardized yet but still they point toward maintaining the authenticity in the knowledge generated. The layer describing proof is therefore presumed responsible for providing evidence for the accuracy. At present no technology recommended to support this layer exists but there is an attempt for developing a proof language called Proof Mark-up Language (PML (da Silva et al., 2004; Al-Feel et al., 2009)) by knowledge systems laboratory at Stanford University. The top most layer Trust is to certify the knowledge reliability and there is a degrees of confidence in the knowledge generated within the layers under it. Again, at present there is no technology to support the layer.

The figure 5 can hence be updated with the three categories defined in this section and is illustrated in figure 6.

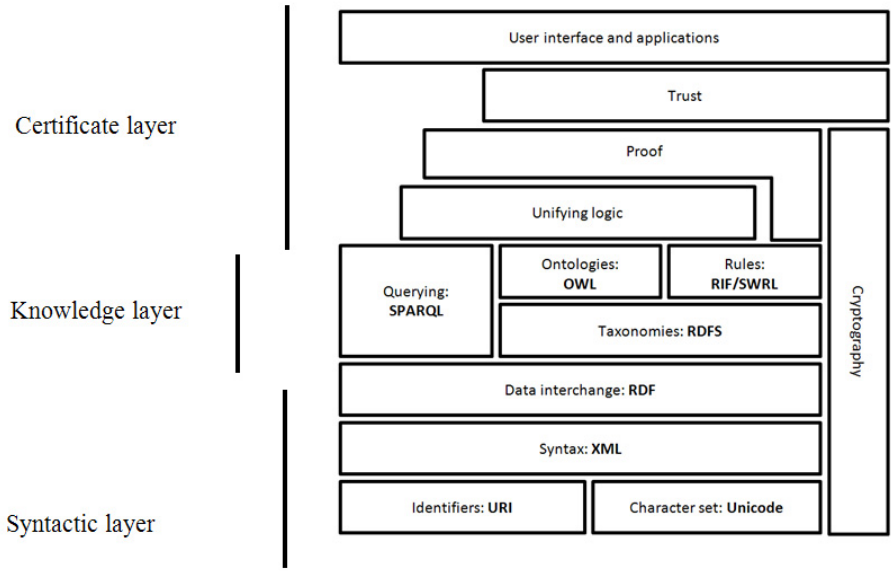

Fig. 6. The layers of the Semantic Web stack.

3.2.1 The syntactic layer

Semantic Web technologies are built up through the Web technologies that could hold up contents. The emergence of the eXtensible Markup Language (XML) marked the beginning of content based information in the Web environment. The language can encode information in machine readable format. The XML syntax is recommended in various data models and this syntactical approach laid a foundation for data models for defining metadata as Resource Description Framework (RDF). Resources are conventionally described through their metadata. The W3C recommended RDF as a standard to define the resources on Web.

W3C has defined five major reasons for developing the standard (Klyne et al., 2004). They focus on automatization of the information processing through serialization. That means the contents inside the documents are machine processable. In order for the documents to be machine processable they need to be machine readable and since the syntax of RDF is based on XML, it provides a mechanism to represent the information in machine readable manner.

The RDF (Resource Description Framework) is a graph data model. It is basically a framework to represent information on the Web. It has also been assigned as the standard model for data interchange on the Web by W3C because it can merge different sets of data irrespective to the underlying schema. RDF is conceptualized through graph data model which demonstrates the underlying structure of its expression. The nodes in the graph model are resources which can represent Uniform Resource Identifiers (URI reference or simply URIRef) or literals or even blank. The link in the graph representing properties are generally URI references. The literals within RDF expressions are generally assigned values of certain data types. RDF syntax is primarily based on its predecessor XML and is defined by RDF abstract syntax. This abstract syntax is the syntax over which the formal semantic are defined. It is a set of triples known as RDF graph (Klyne et al., 2004). It consists three parts which are normally called RDF triplet and represent a statement of relationship between the objects.

3.2.2 The Knowledge Layer

Knowledge representation has been described in five distinct roles it plays in (Davis et al., 1993). Those roles are

- A surrogate for the thing itself used to enable an entity to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than acting it.
- A set of ontological commitments, i.e., an answer to the question: In what terms should I think about the world?
- A fragmentary theory of intelligent, reasoning, expressed in terms of three components
    - The representation's fundamental conception of intelligent reasoning
    - The set of inferences the representation sanctions; and
    - The set of inferences it recommends
- A medium for pragmatically efficient computation, i.e., the computational environment in which thinking is accomplished.
- A medium for human to express, i.e., a language human expresses things about the world.

With these roles in view, different languages that represent the knowledge have been conceived over the time. They vary in terms of their characteristics, expressive power and computational complexity. The effectiveness of any representation language can be measured in:

- The expressiveness of the language is measured in terms of the range through which the language can use its constructs to describe the components in knowledge model.
- The strictness in the language is measured through the consistency and satisfiability within the knowledge model. The consistency and satisfiabilty issue is important in any

knowledge model because they decide the reliability of the model. If any model contains statements which contradict with each other, the model cannot be considered reliable. For an example A cannot be a father and son of B at the same time. Such statements should be rigorously audited for the model to be reliable enough.
- The semantic within the model should not be ambiguous. The meaning of each statement within the model should be clear and unambiguous.

RDFS

RDFS or the RDF Schema is the semantic extension of RDF. The applications using RDF uses it to describe its resources and those descriptions can be modeled as relationships among Web resources. These models constitute of interrelationships among the resources. They are carried out through the named properties and values. It however lacks the mechanism of defining the relationships between properties and other resources. Furthermore RDF data models do not declare these properties. They are hence information without any semantics. RDFS is designed to address these shortcomings. RDFS provides mechanisms for describing groups of related resources and the relationships between these resources.

The Web Ontology Language (OWL)

OWL or the Web Ontology Language is a family of knowledge representation language to create and manage ontologies. It is in general term an extension of RDFS with addition to richer expressiveness that RDFS lacks through its missing features (Antoniou & Harmelen, 2003). The OWL Working Group has approved two versions of OWL: OWL 1 and OWL 2. This research work uses OWL 1 for the applications of ontology as this version was the most used version at the time of research. The later version of OWL 1 was just evolving during the period. This research work discusses its activities in terms of OWL 1.

The expressiveness of OWL depends upon the level of serialization. The expressiveness of OWL comes at the cost of computational efficiency and reasoning effectiveness. This tradeoff between expressiveness and reasoning support was addressed through classifying OWL into three sub languages by the W3C Web Ontology working group.

OWL Full contains the maximum expressiveness but may lack in computational processing capability. It may also have restricted reasoning efficiency. OWL Full is completely compatible with RDF/RDFS both syntactically and semantically. OWLDL is compatible to the components of description logics and provides the functionalities of DLs. It provides the complete computational efficiency and reasoning capabilities. It is sub language of OWL Full with all OWL language constructs which could be used only through certain restrictions (McGuinness & Harmelen, 2004). This restriction is even more in OWL Lite – the third sublanguage of OWL. The advantage of this language is its easiness to understand and implement but the drawback is it is just a simple and fast migration from thesauri and other taxonomies.

The SPARQL language

It has been stated before that RDF statements store data in the form of informative contents. In this manner, it could be easily argued RDF documents are datasets complimenting the data storage capability of its conventional counterparts as database systems. As database systems provide efficient retrieval of the data through its query language in form of

Structured Query Language (SQL), the dataset within a RDF document can be retrieve through the query language called SPARQL. As with its counterpart SPARQL is also used to manage the RDF document. It is a key component of Semantic Web technology. As a query language, SPARQL is "data-oriented" in that it only queries the information held in the models; there is no inference in the query language itself. SPARQL does not do anything other than taking the description of what the application wants, in the form of a query, and returns that information in the form of a set of bindings or an RDF graph. In addition, the SPARQL is able to query OWL ontologies which use RDF graphs to structure them. However, no inferences are possible on that structure. SWRL is used for that purpose.

The query language has been standardized by W3C and has been recommended as official query language to retrieve RDF data (Prud'hommeaux & Seaborne, 2008).

SPARQL queries the RDF data in four distinct forms.

- SELECT returns the resulted dataset from this form. The results could be used accessed by the APIs as well could be serialized into XML or RDF graph.
- CONSTRUCT form constructs a RDF graph through running the query to derive the solution in solution sequence and then combines these triplets.
- ASK form is used to ask the authenticity of the query pattern. That means whether certain query pattern returns a solution or not.
- DESCRIBE forms describe the RDF data about its resources.

The SWRL language

An inference process consists of applying logic in order to derive a conclusion based on the observations and hypothesis. In computer science Inferences are applied through inference engines. These inference engines are basically computer applications which derive answers from a knowledge base. These engines depend on the logics through logic programming.

The horn logic more commonly known Horn clause is a clause with at most one positive literal. It has been used as the base of logic programming and Prolog languages (Sterling & Shapiro, 1994) for years. These languages allow the description of knowledge with predicates. Extensional knowledge is expressed as facts, while intentional knowledge is defined through rules (Spaccapietra et al., 2004). These rules are used through different Rule Languages to enhance the knowledge possess in ontology. The Horn logic has given a platform to define Horn-like rules through sub languages of RuleML (Boley, 2009). There have been different rule languages that have emerged in last few years. Some of these languages that have been evolving rapidly are Semantic Web Rule Language (SWRL) and JenaRule. Both have their own built-ins to support the rules. This research work uses SWRL to demonstrate the concepts but it could be applied to others rule language based on Horn clauses.

Semantic Web Rule Language (SWRL (Horrocks et al., 2004)) is a rule language based on the combination of the OWL-DL (SHOIN(D)) with Unary/Binary Datalog RuleML which is a sublanguage of the Rule Markup Language. One restriction on SWRL called DL-safe rules was designed in order to keep the decidability of deduction algorithms. This restriction is not about the component of the language but on its interaction. SWRL includes a high-level abstract syntax for Horn-like rules. The SWRL as the form, antecedent→consequent, where both antecedent and consequent are conjunctions of atoms written a1^ ... ^ an. Atoms in rules can be of the form C(x), P(x,y), Q(x,z), sameAs(x,y), differentFrom(x,y), or builtIn(pred, z1, …, zn), where C is an OWL description, P is an OWL individual-valued property, Q is an OWL data-valued property, pred is a datatype predicate URIref, x and y are either individual-valued variables or OWL individuals, and z, z1, … zn are either data-valued variables or OWL data literals. An OWL data literal is either a typed literal or a plain literal. Variables are indicated by using the standard convention of prefixing them with a question mark (e.g., ?x). URI references (URIrefs) are used to identify ontology elements such as classes, individual-valued properties and data-valued properties. For instance, the following rule asserts that one's parents' brothers are one's uncles where parent, brother and uncle are all individual-valued properties.

$$\text{parent}(?x, ?p) \wedge \text{brother}(?p, ?u) \rightarrow \text{uncle}(?x, ?u) \quad (1)$$

The set of built-ins for SWRL is motivated by a modular approach that will allow further extensions in future releases within a (hierarchical) taxonomy. SWRL's built-ins approach is also based on the reuse of existing built-ins in XQuery and XPath, which are themselves based on XML Schema by using the Datatypes. This system of built-ins should also help in the interoperation of SWRL with other Web formalisms by providing an extensible, modular built-ins infrastructure for Semantic Web Languages, Web Services, and Web applications. Many built-ins are defined and some of most common built-ins can be found in (Horrocks et al., 2004). These built-ins are keys for any external integration. The research work develops spatial built-in for the integration of spatial data structure.

3.3 Discussion

The Semantic Web, a set of technologies complementing the conventional Web tools proposed by Sir Tim Berners-Lee is seen as the most probabilistic approach to reach the goal of semantic interoperability. The Semantic Web is envisaged as an extension to the existing Web from a linked document repository into the platform where information is provided with the semantic allowing better cooperation between people and their machines. This is to be achieved by augmenting the existing layout information with semantic annotations that add descriptive terms to Web content, with meaning of such terms being defined in ontologies (Horrocks et al., 2004). Ontologies play crucial role in conceptualizing a domain and thus play an important role in enabling Web-based knowledge processing, sharing and reuse between applications.

This research takes advantages of the tools of Semantic Web technology to make a case of information management through knowledge. The case study of Industrial Archaeology fits perfectly to put forward the concept of information handling through knowledge as the domain generates huge and heterogeneous dataset. In addition the sites are not preserved for continuing excavation as in case of the conventional archaeology, making it ideal for utilizing knowledge techniques to manage the information because of the flexibility in knowledge techniques to handle information long after they are collected. The definition of a domain ontology representing the site is sketched out by the archaeologists. It is again their task to fill in knowledge in the domain ontology to make it a knowledge base where one can reason to derive new knowledge. Archaeologists use collaborative Web platform based on Semantic Web technology to identify the objects and define them in the ontology. These objects once defined, performs as common schemas between data sources to achieve a sense of data interoperability. The definitions of objects add semantics to the objects and thus adding knowledge about the objects. Knowledge techniques based on Description Logics (DLs) exploit these semantics to manipulate implicit knowledge within the knowledge base. Inference engines utilize the definition of DLs to infer the knowledge base through Horn based rules. The knowledge base stored in OWL syntactic structure is inferred through SWRL to infer the rules. This inference is complimented through querying with SPARQL.

Carrying the discussion from last chapter, this research attempts to use the Semantic Web techniques to perform spatial analysis in form of spatial SPARQL and spatial SWRL. The spatial analysis through Semantic Web can only be possible through providing spatial signatures to the defined objects in the ontology. This will allow the knowledge techniques to process spatial solutions. The spatial integration is carried out through OWL/RDF again and the spatial management is carried out again through tools as SWRL and SPARQL. This simplistic yet but effective approach of spatial integration into Semantic Web technologies provides the possibility to include different modes of data into its framework.

The Semantic Web stack shown in figure 5 and 6 can adjust a layer of spatial information into it. The research proposes such an arrangement in the stack. A layer of spatial data mixing seamlessly with the semantic proposition in the layer Ontology through its OWL/RDF based syntax can be envisaged. This layer since uses the standard syntax of OWL/RDF can perform spatial queries through SPARQL or infer rules through standards as SWRL. The next chapter discusses this integration process of spatial technology and Semantic Web technology which is undertaken by defining spatial FILTERs for SPARQL queries and spatial Built-ins for SWRL rules. Ideally the layer should be the top most layer of knowledge level but spatial layer does not yet possess any standards that are standardized by W3C so could not be placed there. It is hence placed as the bottom layer in the certificate level. The next chapter discusses this adjustment in stack in detail and how to apply spatial queries and rules on any existing ontology.

4. The spatial layer of the WS stack

This chapter presents the integration process of spatial technologies and the Semantic Web technologies at the backdrop of Industrial Archaeology, and its associated tool called the spatial facilitator which is a query and rule engine. The technologies discussed in previous chapters are used and adjusted for processing the spatial knowledge through knowledge technologies within the Semantic Web framework in the research works. This chapter attempts to outline the methods and the processes of these adjustments and how they return the results through knowledge tools as SWRL and SPARQL.

The discussions of the last two chapters aim at laying a background on the concepts of integration process. The discussions on Semantic Web and its underlying technologies and the spatial technology in GIS in the last two chapters have clearly pointed out that the technical advancements toward semantic technologies are integrating every data structures so it will integrate spatial data structure in future. However, for now it is still a topic of research. It could be conceived from earlier discussions that the integration process requires adjustments of the spatial components within the ontological framework. This chapter is dedicated to discuss the steps and process of this adjustment. The spatial signature of objects plays an important role in determining them. The identification of objects is the process of signing these spatial signatures on them. These signatures should be integrated within the semantics of the objects seamlessly in order to process the spatial knowledge through the knowledge technology. It should be noted however that the Semantic Web technologies are in the maturation process and hence there exists certain processing problems within especially for the non-conventional data type as that of spatial data. Thus, it needs to be sorted out through the existing tested techniques. The research in GIS systems uses the capabilities of existing RDBMS to process the spatial data through spatial operations and functions and use the results of these processes.

The Semantic Web stack discussed in the previous chapter can be updated to address the inclusion of a spatial component. Every tangible object has its spatial signature and thus it becomes indispensable to address the spatial component within its semantic framework. The Semantic Web technologies and its architecture are mostly influenced by the nature of information available on the Internet. Hence, these levels deals mostly with managing the semantic based information through knowledge technologies. However, in recent years there has been huge surge of other forms of information on Web platform and they need to be managed as well. With the advancement in spatial technologies, the trend of disseminating spatial information through Web based environment is rapidly growing. This has raised the issue of the integration of spatial component into the Semantic Web framework.

A layer representing geospatial data in the Semantic Web stack can be placed just above the knowledge layers as could be seen in figure 7. As the technologies within knowledge level are standardized by W3C, the geospatial layer needs to be above the level. However, the technologies within knowledge level needs to blend spatial components seamlessly both syntactically and semantically to maintain the satisfiability required for the consistency of the ontology. This integration procedure should be adjusted within the knowledge tool within the knowledge level of the stack. This approach thus uses the knowledge techniques through adding the spatial structures within them and implementing the spatial knowledge processing along with semantic knowledge processing. The first Semantic Web tool that comes direct in contact with the integration procedure is the structural schema of the knowledge base which is termed as top level ontology in general sense. The top level ontology is the structural schema that represents the nature of knowledge the ontology possesses. It should include the components to adjust the behavior of the knowledge base. Hence the initial task that needs to be adjusted within any top level ontology to perform spatial knowledge processing is to include spatial components within it.

The top level ontology is the structural schema that represents the nature of knowledge the ontology possesses. It should be noted that the top level ontology is syntactically presented through OWL/RDF and contains the top level concepts of the domain. Among these top level concepts, the concepts presenting the spatial components for storage, retrieval and processing of the spatial knowledge should be present. Moving down to the enrichment process, the spatial signatures are mapped to the objects within the knowledge base is again

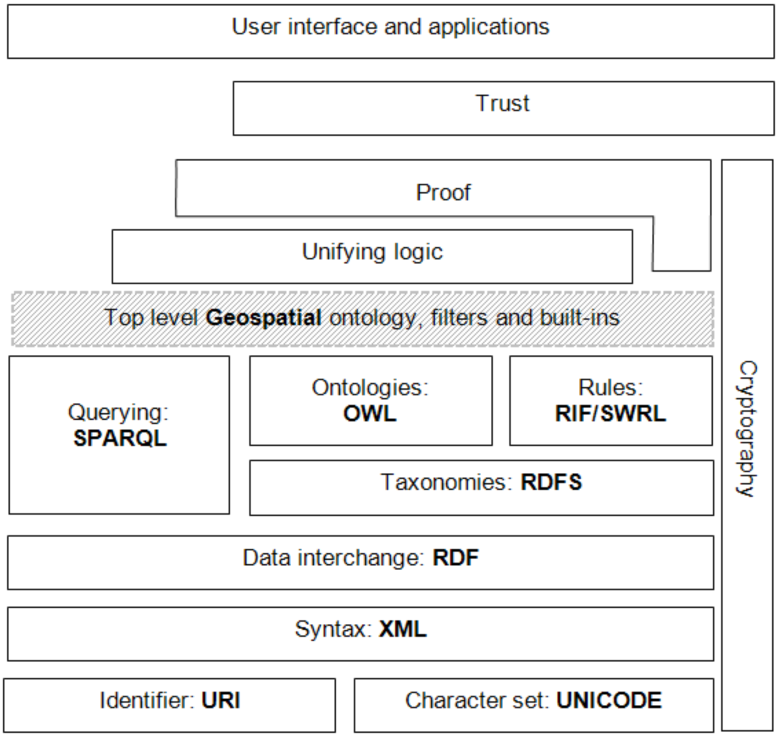

Fig. 7. The inclusion of a Geospatial layer in the Semantic Web Stack.

encoded with OWL/RDF syntax. The methodology of this integration is discussed in later sections within this chapter.

Similarly, the spatial filters and spatial built-ins defined in this layer facilitate the spatial querying and the spatial rule definition. The layers of Rules: SWRL/RIF and Querying: SPARQL provide a base to knowledge management through processing the spatial information semantically within the knowledge base. The only adjustment that is needed is to execute the built-ins and filters in conjunction to the processing capabilities of spatial extensions within current database systems.

Putting forward the arguments on the authenticity of the layer with respect to other layers, the geospatial layer exploits the capabilities of the layers below maintaining the trend of the stack. At the time of integration, the spatial components are included within the top level ontology which stores, retrieves and processes spatial knowledge and utilizes the capabilities of the other technologies in the stack. The spatial components on the top level ontology and the mapped spatial signatures are encoded through the OWL/RDF syntactical structure thus justifying the involvement of ontologies in the stack. Then after, the capability of the SWRL language is exploited through spatial built-ins for spatial SWRL rules. Similarly, the querying capability of the SPARQL language is exploited through spatial filters for the query language. These filters and built-ins can be used with conjunction to already standardized filters and built-ins of both the technologies thus forwarding the arguments of the process in standardizing these built-ins too.

4.1 The top level ontology

The top level ontology or more popularly upper ontology describes the general concept behind the knowledge domain. This ontology varies with the domain it addresses. There are efforts to come out with a universal upper ontology which addresses the requirements of every knowledge domains but they still are in the phase of researches. Every domain uses its own standard upper ontology for its purpose. This research work attempts to propose an upper level ontology for the domain of industrial archaeology. This top level ontology is the main driving force behind the ArchaeoKM framework. It represents the knowledge possessed by archaeologists in form of descriptions, observations and rules represented through different axioms within the ontology. This ontology serves as a foundational ontology to which objects can be instantiated during identification process. The axioms are the building blocks of the ontology. The integration of spatial components within the framework holds major importance and is required to be adjusted within the top level ontology of ArchaeoKM. The spatial extension of the top level ontology is discussed in the next section.

4.2 The spatial top level ontology

The realization of spatial signatures of the identified objects in the knowledge base has been discussed earlier. The attachments of these spatial signatures provide a framework that could exploit the developments in spatial technology to provide the objects their spatial identity in respect to their surrounding objects. However, it is important to adjust the components of the spatial technology in the top level ontology. This section covers the spatial top level ontology of the ArchaeoKM framework.

Although the impact of spatial integration is realized in the semantic level when the spatial components are integrated in the ontology, the usage of spatial features begins earlier than that. The spatial functionalities provided by database system form foundations to how they should be adjusted. A parallel structure facilitating the spatial components in different levels of the system architecture has already been presented in chapter 2 through figure 3. At the syntactic level where most of knowledge generation activities are carried out, spatial components are handled through spatially annotating the identified objects. This spatial annotation process draws a Minimum Bounding Rectangles (MBRs) around the objects and stores them as spatial data type in PostgreSQL database system. These MBRs would be used to carry out spatial rules while managing knowledge. It should be noted that the MBRs are not the optimal way of representing the objects and would constitutes some degrees of error during the analysis process. The ideal approach would be to use the boundaries of the objects for representation and analysis purpose. The algorithm to extract point cloud from the boundary is still in the domain of research and not completely matured and hence this research uses MBRs to put forward the ideas.

It is the semantic level where the most of the integration work is carried out. The domain ontology is modified to represent the spatial functions and operations within it. The research work revolves around two categories of spatial operations and the integration process takes the functions and operations within these two categories which are the georelationship functions and the geoprocessing functions. These functions are defined by the OGC consortium. The Open Geospatial Consortium, Inc. (OGC) is an international industry consortium of 404 companies, government agencies and universities participating in a consensus process to develop publicly available interface standards. OpenGIS® Standards support interoperable solutions that "geo-enable" the Web, wireless and location-based services, and mainstream IT. The standards empower technology developers to make complex spatial information and services accessible and useful with all kinds of applications.

The top level ontology should model spatial technology in terms of its spatial functions and operations. This modeling process should accommodate the spatial functions and operations and maintain their true identity.

4.3 Translation engine

The translation engine is a part of the spatial facilitator that allows the computation of spatial SPARQL queries and spatial SWRL rules. In both cases, the translation engine interprets the statements in order to parse the spatial components. Once the spatial components are parsed, they are computed through relevant spatial functions and operations by the translation engine through the operations provided at the database level. The results are populated in the knowledge base thus making it spatially rich. After that, the spatial statements are translated to standard statements for the executions through their respective engines. With the inference engine, the enrichment and the population of the ontology through the results of the inference process is stored in the ontology.

The next sections present in details the translation engine and more specifically the translation process of spatial SPARQL queries to regular queries. The following one presents the translation process of spatial SWRL rules to regular SWRL rules. These two processes have in common the use of SQL statements to query to the spatial database.

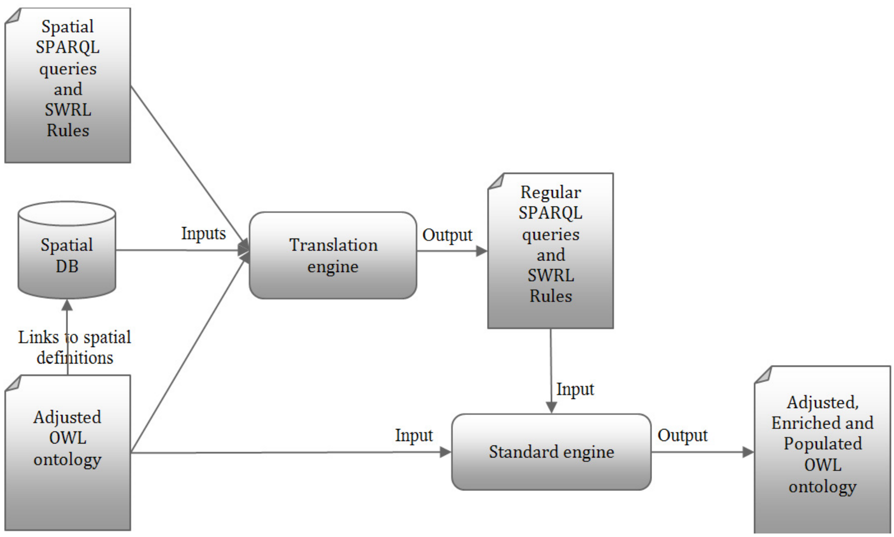

Fig. 8. The spatial processing of the translation Engine.

4.3.1 Spatial SPARQL queries

FILTERs can be used to compare strings and derive results. The functions like regular expression which matches plain literal with no language tag can be used to match the lexical forms of other literals by using string comparison function. In addition, SPARQL FILTER uses the relational operators as = or > or < for the comparison and restrict to the results that they return. The FILTER principle can hence be extended in order to process the geospatial functions.

Geoprocessing FILTER

Geoprocessing functions need to be addressed through enriching the knowledge base with the spatial operations which is related to them during the execution of the query. The enrichment process should be rolled back after the results are returned into its original form iff the SELECT statement is used under the filter. The optimization of the SPATIAL_FILTER is discussed later which highlights the management of the knowledge base during the execution of the SPARQL queries.

The following example demonstrates the syntax of geoprocessing filters in SPARQL. It could be seen that a new spatial filter through the keyword SPATIAL_FILTER is introduced which helps the translation engine during the parsing process. The SPARQL statement with spatial filters in the example returns names of all the buildings in class feat:Building which are intersecting with the buffer of 2000 meters of the rivers in class feat:River with their respective rivers names.

```
SELECT  ?name1 ?name2
WHERE
{
        ?feat1      feat:name      ?name1
        ?feat2      feat:name      ?name2
        ?feat1      rdfs:type            feat:River
        ?feat2      rdfs:type            feat:Building

        SPATIAL_FILTER [buffer (?x, 2000,?feat1)]
        SPATIAL_FILTER [intersection (?y,?x,?feat2)]
}
```

Georelationship FILTER

In case of georelationship filter, it is straightforward as the enrichment process requires enriching the object properties imitating spatial relationship between objects through the results of the spatial operations at the database level. As with the previous case, the georelationship filter uses the keyword SPATIAL_FILTER. This keyword parses the spatial components from the SPARQL statements. The following example illustrates the execution of SPARQL with these filters. The first feature is a feat:River which is of kind of feat:Feature, and the second feature is a feat:Building which is also of kind of feat:Feature. The SPATIAL_FILTER selects the rivers and buildings which are touching spatially.

```
SELECT  ?name1 ?name2
WHERE
{
            ?feat1            feat:name        ?name1
            ?feat2            feat:name        ?name2
            ?feat1            rdfs:type              feat:River
            ?feat2            rdfs:type
                 feat:Building

            SPATIAL_FILTER [touches (?feat1, ?feat2)]
}
```

4.3.2 Inference rules through SWRL

In an attempt to define the built-ins for SWRL, a list of eight built-ins was proposed during the research work. These eight built-ins reflect four geoprocessing functions and four georelationship functions that are discussed previously. The built-ins reflecting geoprocessing functions are built up in combinations with the spatial classes adjusted in the ontology and their relevant object properties. The built-ins for georelationship functions are object properties and corresponding spatial functions in database system.

The domain of archaeology benefits from this work and could surely be of benefit for lot of other domains. To show this we present a simple example to determine the location of possible flooding zone when the river bank bursts with excessive water during rainy season. This is a very common exercise for a flood management system in hydrology and it gives interesting clues for archaeology. In general with a common GIS, a set of activities are carried out which are mentioned in the following sequences:

- Buffer the river by certain distance (e.g. 100 meters)
- Determine the elevation of land parcel inside the buffer zone
- Check whether the land parcel elevation is above the threshold (e.g. 25 meters)
- Select areas below the threshold area and determine them as flood liable zone.

It should be understood that this example is provided just as a proof of the concept. Hence details on other hydrological factors are ignored on purpose. For a simple location analysis as such requires at least four steps of spatial analyses. This paper provides an alternative through the spatial extension of SWRL in one step. We combine the existing built-ins in existing SWRL and the spatial built-in mentioned in this paper to execute this analysis.

River(?x) ^ LandParcel(?y) ^ hasElevation(?y, ?Elv) ^ swrlb:lessThan(?Elv, 25) ^
spatialswrlb:Buffer(?x, 50, ?z) ^ spatialswrlb:Intersection(?z, ?y, ?res)→
FloodingLandParcel(?y) (2)

The result of this rule is that the individuals which respect the rule and belong to LandParcel, belong also to the concept FloodingLandParcel.

5. Conclusion

This research has made an attempt to contribute through including the functionalities of spatial analysis within the Semantic Web framework. Moving beyond the semantic information, it has opened the chapter of inclusion of other form of information. It is important for the development of the technology itself. The world is witnessing a shift in technology and the Semantic Web is the direction the shift is moving towards. This would mean that the technology including that of GIS is moving towards the flexible solutions through knowledge based systems from static solution through current database systems. Hence, it is important to raise issues of integrating non-typical semantic data into it. This research work at least provides certain vision towards the direction the technology is taking to integrate these forms of data. It discusses the direction in terms of spatial integration. There are other data patterns like temporal data which need to be addressed too.

This concluding chapter begins with summarizing the work contribution that has been presented in previous chapters. It then discusses the contribution made to different related discipline. Lastly, the chapter concludes the future prospect and the direction of the research work in this field.

5.1 Contribution

This research attempts to highlight the possibilities to integrate spatial technology in Semantic Web framework. It moves beyond the scope of data interoperability while presenting the concept and makes efforts to utilize the potentiality in other areas of the Semantic Web technologies. The underlying technologies of knowledge processing provide the Semantic Web capabilities to process the semantics of the information through close collaboration with the machine. It makes not only the understanding of data easier for achieving interoperability among different data sources, but it also provides valuable knowledge which could enrich the knowledge base in order to equip it with new knowledge through the knowledge management techniques. This helps the users understand the data better.

5.1.1 In the industrial archaeology domain

This research benefits from the advancement in Semantic Web technologies and its knowledge representation formalization tools and techniques. The primary principle of 4Ks processing is based on the knowledge formalization techniques. The research uses the case study of the industrial archaeology to demonstrate the possibility of implementation of application based on Semantic Web and utilizes the knowledge possessed by the archaeologists to manage the information recovered. This turns out to be an ideal case for the experimentation as the site for industrial archaeology is available for short duration of time. With the conventional technology it is difficult to manage the information due to share volume of data and the limitation of available time. It is however seen that with 4Ks implemented within the application prototype of the ArchaeoKM framework, the information could be managed. There has always been active involvement of archaeologists in every phase of design and development. The domain ontology and its axioms and theorems are based on their experiences. The enrichments of domain ontology through the identification of objects are carried out by them. It is the first K, Knowledge Acquisition. The knowledge acquired through the identification process is managed through defining relationships. It is again the archaeologists with the ArchaeoKM platform to manage knowledge through adjuring proper relationships (which reflects archaeologists view of the world) to the objects and semantically annotating them to the data and documents collected. The process is second K: Knowledge Management. The third K is Knowledge Visualization which generally means that knowledge identified and managed could be visualized through the interfaces of the ArchaeoKM platform. The knowledge base enriched and managed through the collaborative approach of archaeologists could be analyzed through inferring the knowledge base with rules formulated by archaeologists. These rules are inferred through SWRL – a rule language for Semantic Web standardized by W3C (Horrocks et al., 2004). It is the last K, Knowledge Analysis.

5.1.2 In the geospatial domain

The 4Ks processing principle is implemented during the integration of spatial technology. The domain ontology is modified to adjust the spatial components into it. The research work considers the advancement in spatial technology in modern database systems. It implements the notations standardized by OGC simple feature specification (Herring, 2010) during the inclusion of the spatial components as axioms into the ontology. The spatial technologies provide spatial functions and operations to perform spatial analysis. These functions and operations are categorized into four major categories as documented in PostGIS documentation. However, the research implements functions under geoprocessing and georelationship functions as these two categories consist of mostly all the spatial functions. Geoprocessing functions are implemented as class axioms which relate to the classes containing features through the respective object properties. Likewise the georelationship functions are treated as object properties relating the classes containing features spatially to each other.

The knowledge acquisition process comprises of acquiring spatial signatures of the object. In general they are acquired during the identification process. However, the spatial signatures are formalized during spatial annotations of the objects which are then stored in database as spatial data type. The spatial operations and functions which are encoded as classes and object properties within the ontology provide the management of spatial knowledge. The ontology was spatially enriched through the spatial operations and functions at the database level. This enriched knowledge base can be inferred spatially through the spatial built-ins for SWRL proposed in the research. The research also proposes the spatial filters for query language of the Semantic Web (SPARQL) (Harris & Seaborne, 2010).

The benefits to geospatial community are prominent. The shift from data oriented to knowledge oriented GIS gives the GIS an edge. The flexibility of knowledge based systems should add the flexibility to GIS in terms of data acquisition, data management and data analysis. The data acquisition process though still remains to the conventional digitization techniques; the possibility of linking it up to its semantics adds knowledge to the whole process. This added knowledge then could be utilized for different purposes including semantic interoperation between other data from other sources. However, this paper discusses in terms of knowledge management and analysis. The knowledge query through

SPARQL or knowledge inference through SWRL to the spatially rich knowledge base generates new knowledge which is more authentic in a sense that this new result is the manipulation of knowledge base through the existing one. It is not just data any more. The semantic behind the results provides support to their authenticity.

This research has provided GIS community an alternative to conventional spatial data analysis through spatial rules. It can be opined that the proposed approach of knowledge analysis is apparent and less complicated to the conventional one. As the spatial rules could be combined with general rules they have wider implications. Additionally, the rules are based on formal logics which relate to day-to-day human interpretations; they should be easy to understand and implement. Consequently, the research proposes a rule based approach for spatial analysis and provides an evidence of possibilities through the experimentation performed.

5.1.3 In the Semantic Web domain

A spatial layer in the Semantic Web stack presented through this paper is not enough to address the overall problems of non-semantic data within the framework but at least there is something to start with. The full potential of underlying knowledge techniques through the reasoning or inferring capabilities within Semantic Web has not been identified in Geospatial community. The primary focus on these technologies is to achieve data interoperability within different data sources (Cruz, 2004; Cruz et al., 2004) Even W3C concentrated its priority in proposing comprehensive geospatial ontology acceptable to all through its Geospatial Incubator Group (Lieberman et al., 2007). All these research works show that the emphasis on using geospatial ontology lie in achieving data interoperability and thus ignores the capabilities of underlying knowledge techniques for carrying out complex spatial analysis. This research presented a concept to carry out spatial analysis through inferring knowledge base spatially.

The realization of spatial integration into Semantic Web framework is demonstrated through a demonstration application. The application demonstrates that through a suitable translation engine, it is possible to infer the spatially enriched knowledge base in order to deduce spatial knowledge. The translation engine developed within the demonstration application translates the spatial built-ins and enriches the knowledge base through results of spatial operations of these built-ins making the knowledge base ready to be inferred.

5.2 Way forward

This research work has highlighted the benefits of tools and techniques of the Semantic Web and especially underlying knowledge technologies and their usages with the spatial technologies for the efficient management of spatial information. It has also been discussed that the approach presented here benefits both the Semantic Web and spatial technology. The research activities has just initiated the integration of spatial technology into the Semantic Web framework and still has long way to go. This section presents few areas where the research work could be continued in this area.

Researches in the field of spatial technology within the Semantic Web framework have not moved beyond geospatial ontology and the possibility of semantic interoperability between different sources. This research attempts to break that trend and use knowledge to manage the spatial data through knowledge management techniques. In the process, it provided the mechanisms to infer spatial rules through spatial built-ins for SWRL. This was done first through populating domain ontology with the spatial components so that spatial knowledge could be enriched into it and this spatially rich knowledge base is inferred through SWRL. It could also be queried through SPARQL. However there are number of issues that need to be addressed in future work. The first one is about the dependability on the database systems to conduct the spatial operations and functions. This research uses the spatial operations and functions provided by PostGIS, the spatial extension of PostgreSQL to enrich the knowledge base through their result. Future works should make an attempt to free them with such dependency through providing such functionalities within spatial built-ins themselves.

Another area where the research could concentrate is the area of using current reasoning engines to reason the spatial knowledge base and deduce the implicit spatial knowledge. In other words addition to the the inference engine to infer the rules through SWRL, the constraint axioms should be introduced within the ontology which automatize the enrichment of knowledge base through reasoning mechanism. The constraint axioms in particular should be able to include the spatial built-ins and run through the respective spatial operations and functions to automatize the enrichment process while reasoning the knowledge base. It can be clarified with one of the typical examples in industrial archaeology: “chimney should be 5 meters around an oven and should be round”. Currently it is possible to execute this only through SWRL rule.

feat:Object(?x) ^ feat:Oven(?y) ^ spatialswrlb:Buffer(?y, 5, ?x) ^ att:hasShape(?x, round) → feat:Chimney(?x) (3)

This infers the spatial knowledge base to annotate the result to the class feat:Chimney. However an alternative could be a theorem

$$\text{feat:Chimney} \sqsubseteq \text{Within(Buffer(feat:Oven,5))} \sqcap \text{hasShape.\{round\}} \quad (4)$$

can be thought upon. The existing reasoning engine then reasons every object with round shape around 5 meters of every oven and terms them as individuals of chimney.

Lastly, it is important to have standard terms for every built-in that will be developed to process spatial knowledge. With other built-ins in the tools standardized by W3C, the spatial built-ins should also get standardized by the consortium. In addition to W3C, OGC should also get involved in standardizing the built-ins. An effort in this direction should be carried out.

6. References

Al-Feel, H., Koutb, M. A., & Suoror, H. (2009). Toward An Agreement on Semantic Web Architecture, *World Academy of Science, Engineering and Technology* (49), 806-810.

Antoniou, G., & Harmelen, F. v. (2003). Web Ontology Language: OWL, *Handbook on Ontologies in Information Systems* (pp. 67 - 92). Springer Series.

Baader, F., & Nutt, W. (2002). Basic Description Logics, *Description Logic Handbook* (pp. 47 - 100). Cambridge: Cambridge University Press.

Baader, F., & Sattler, U. (2001). An overview of tableau algorithms for description logics. *Studia Logica* , (pp. 5 – 40), Springer.

Baader, F., Horrocks, I., & Sattler, U. (2003). Description Logics as Ontology Languages for the Semantic Web, *Festschrift in honor of Jörg Siekmann, Lecture Notes in Artificial Intelligence,* (pp. 228 – 248) Springer

Balletti, C., Guerra , F., & Adami, A. (2005). 3D Multiresolution Representations in Archaeological Sites, *ISPRS* (pp. 654-659), ISPRS XXXVI-5/C34.

Bateson, G. (1979). Mind and Nature: A Necessary Unity (Advances in Systems Theory, Complexity, and the Human Sciences). Hampton Press.

Bechhofer, S., Harmelen, F. v., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., et al. (2004, February 10). OWL Web Ontology Language, In: *W3C Recommendation*, Date of Access: November 27, 2009, Available from: http://www.w3.org/TR/owl-ref/

Bellinger, G. (2004). Knowledge Management—Emerging Perspectives, In: *Systems thinking*, Date of Access: December 13, 2010, Available from http://www.systems-thinking.org/kmgmt/kmgmt.htm

Berners-Lee, T. (1998, October 14). Why RDF model is different from the XML model, In: *W3C: A roadmap to the Semantic Web*, Date of Access: October 2010, 22, Available from: http://www.w3.org/DesignIssues/Overview.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001, May). The Semantic Web, *Scientific American*, pp. 34-43.

Berry, J. K. (1999). GIS Technology in Environmental Management: a Brief History, Trends and Probable Future. In D. L. Soden, & B. S. Steel (Eds.), *Handbook of Global Environmental Policy and Administration* (pp. 49 - 81). Decker, Marcel Inc.

Bobillo, F., & Straccia, U. (2009). Supporting Fuzzy Rough Sets in Fuzzy Description Logics, *Proceeding of 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. 5590/2009, pp. 676-687. Verona: LNCS.

Boley, H. (2009, August 11). RuleML: The Rule Markup Initiative, In: *RuleML*, Date of Access: September 27, 2009, Available from: http://ruleml.org/

Boley, H., & Kifer, M. (2010, June 22). W3C Recommendation, In: *RIF Basic Logic Dialect*, Date of Access: August 24, 2010, Available from: http://www.w3.org/TR/rif-bld/

Boochs, F. (2009). Spatial information system for acquisition, documentation and analysis of industrial archaeological objects, *Proceeding of The International Committee for the Conservation of the Industrial Heritage Conference*. Freiberg: TICCIH 2009.

Brachman, R., Eugene, C., Norton, G., & Martin, Y. (1978). KLONE Reference Manual, Cambridge: Bolt Beranek And Newman Inc.

Calvanese, D., Giacomo, G. D., Lenzerini, M., & Nardi, D. (2001). Reasoning in expressive description logics, *In A. Robinson, & A. Voronkov (Eds.), Handbook of Automated Reasoning* (Vol. II, pp. 1581-1634). Elsevier Science Publishers.

Clouse, R. A. (1995). Industrial Archaeology, In: *Minnesota Historical Society of Archaeology*, Date of Access: September 23, 2010, Available from: http://www1.umn.edu/marp/indus/indus.html

Cosmas, J., Itagaki, T., Green, D., Grabczewski, E., Waelkens, M., & Degeest, R. (2001). 3D MURALE: A Multimedia System for Archaeology, VAST *(p. Virtual Reality: Archaeology and Cultural Heritage)*. Glyfada, Nr Athens: ACM.

Costa, S., & Zanini , E. (2008). Sharing knowledge in archaeology: looking forward the next decade(s), *Proceeding of Digital Heritage in the New Knowledge Environment: Shared spaces & open paths to cultural content*, Athens: Hellenic Ministry of Culture

Cruz, I. F. (2004). Geospatial Data Integration, Chicago: ADVIS Lab, Dept. of Computer Science, University of Illinois.

Cruz, I. F., Sunna, W., & Chaudhry, A. (2004). Semi-Automatic Ontology Alignment for Geospatial Data Integration, *GIScience 2004*, (pp. 51 - 66). MaryLand: LNCS.

da Silva, P. P., McGuinness, D. L., & Fikes, R. E. (2004). A Proof Markup Language for Semantic Web Services, Stanford: Stanford University.

Davis, R., Shrobe, H., & Szolovits, P. (1993). What is Knowledge Representation, *AI Magazine*, 14(1), (pp. 17 – 33)

Decker, S., Melnik, S., Harmelen, F. V., Fensel, D., Klein, M., Erdmann, M., et al. (2000). The Semantic Web: the roles of XML and RDF, *IEEE Internet Computing*, (pp. 63-74)

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,* Athens: Hellenic Ministry of Culture.

Gerber, A., Merwe, A., & Barnard, A. (2008). A functional Semantic Web architecture, *Proceedings of the 5th European Semantic Web conference on The Semantic Web: research and applications* (pp. 273 -287). Tenerife: Springer-Verlag.

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications, *Knowledge Acqusition*, 5(2), (pp. 199 – 220)

Guarino, N. (1994). The Ontological Level, R. Casati, & G. White, Eds., *Philosophy and the congitive sciences*

Harris, S., & Seaborne, A. (2010, January 26). SPARQL Query Language 1.1, In: *Minnesota Historical Society of Archaeology*, Date of Access: September 2, 2010, Available from: http://www.w3.org/TR/2010/WD-sparql11-query-20100126/

Hellerstein, J. M., Naughton, J. F., & Preffer, A. (1995). Generalized Search Trees for Database Systems, *Proceedings of the 21th International Conference on Very Large Data Bases* (pp. 562 - 573). San Francisco: Morgan Kaufmann Publishers Inc.

Herman, I., Halpin, H., Hawke, S., Prud'hommeaux, E., Raggett, D., & Swick, R. (2010, February 06). W3C Semantic Web, In: *W3C Semantic Web Activity*, Date of Access: August 15, 2010, Available from: http://www.w3.org/2001/sw/

Herring, J. R. (2010). OpenGIS Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture, Open Geospatial Consortium Inc.

Horrocks, I., Parsia, B., Patel-Schneider, P., & Hendler, J. (2005). Semantic Web Architecture: Stack or Two Towers. *Proceeding of 3rd Internation Workshop on Principles and Practice of Semantic Web Reasoning* (pp. 37 - 41), Dagstuhl: LNCS .

Horrocks, I., Pater-Schneider, P. F., McGuinness, D. L., & Welty, C. A. (n.d.). OWL: a Description Logic Based Ontology Language for the Semantic Web.

Horrocks, I., Schneider, P. f., Boley, H., Tabelt, S., Grosof, B., & Dean, M. (2004, December 21). SWRL - A Semantic Web Rule Language - Combining OWL and RuleML, In: *SWRL: A Semantic Web Rule Language - Combining OWL and RuleML*, Date of Access: May 22, 2009, Available from: http://www.w3.org/Submission/SWRL/

Jones, C. (1997). Geographical Information Systems and Computer Cartography, Addison Wesley Longman Ltd.

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,* Athens: Hellenic Ministry of Culture.

Klyne, G., Caroll, J. J., & McBride, B. (2004, February 10). Resource Description Framework (RDF): Concepts and Abstract Syntax, In: *W3C Recommendation*, Date of Access: August 26, 2010, Available from: http://www.w3.org/TR/rdf-concepts/

Kollias, S. (2008). Achieving Semantic Interoperability in Euroeana, *Digital Heritage in the New Knowledge Environment: Shared spaces & open paths to cultural content*, Athens: Hellenic Ministry of Culture.

Lieberman, J. (2007, June 23). Geospatial Incubator Group, In: *W3C Incubator Activity*, Date of Access: July 2010, 14, Available from:
http://www.w3.org/2005/Incubator/geo/charter

Lieberman, J., Singh, R., & Goad, C. (2007). W3C Geospatial Ontologies - W3C Incubator Report, W3C Incubator Group, In: *W3C Incubator Activity*, Date of Access: September 15, 2009, Available from:
http://www.w3.org/2005/Incubator/geo/XGR-geo-ont/

Mazzocchi, S. (2000). Toward the Semantic Web: a view of XML from outer space, *Proceeding: ApacheCON 2000,* London: Apache.

McGuinness, D. L., & Harmelen, F. v. (2004, February 10). OWL Web Ontology Language: Overview, In: *W3C Recommendation*, Date of Access: July 2010, 12, Available from: http://www.w3.org/TR/2004/REC-owl-features-20040210/

Patel-Schneider, P. F., Hayes, P., & Horrocks, I. (2004, February 10). OWL Web Ontology Language, In: *W3C Recommendation*, Date of Access: July 2010, 12, Available from: http://www.w3.org/TR/owl-semantics/

Prud'hommeaux, E., & Seaborne, A. (2008, January 2008). SPARQL Query Language for RDF, In: *W3C Recommendation*, Date of Access: May 22, 2010, Available from http://www.w3.org/TR/rdf-sparql-query/

Reitsma, F., & Hiramatsu, K. (2006). Exploring GeoMarkup on the Semantic Web, *Proceedings of 9th AGILE Conference on Geographic Information Science* (pp. 110 - 118). Visegrád: 9th AGILE Conference on Geographic Information Science.

Shannon, V. (2006, May 23). The New York Times – Technology, In: *A 'more revolutionary' Web*, Date of Access: August 16, 2010, Available from:
http://www.nytimes.com/2006/05/23/technology/23iht-Web.html

Sheth, A. P. (1999). Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics, *Interoperating Geographic Information Systems* (pp. 5 - 30), MA: Kulewar Academic Publishers.

Smith, M. K., Welty, C., & McGuinness, D. L. (2004, February 10). OWL Web Ontology Language, In: *W3C Recommendation*, Date of Access: November 17, 2009, Available from: http://www.w3.org/TR/owl-guide/

Spaccapietra, S., Cullot, N., Parent, C., & Vangenot, C. (2004). On Spatial Ontologies, *GEOINFO*, Campos do Jordao.

Sterling, L., & Shapiro, S. (1994). The Art of Prolog, MIT Press.

Tanasescu, V., Gugliotta, A., Domingue, J., Gutiérrez-villarías, L., Davis, R., Rowlatt, M., et al. (2006). A Semantic Web Services GIS based Emergency Management Application, *Proceeding of The 5th International Semantic Web Conference (ISWC06)*, Athens, Georgia: ISWC06.

W3C OWL Working Group. (2009, October 27). OWL 2 Web Ontology Language: Document Overview, In: *W3C Recommendation*, Date of Access: October 28, 2010, Available from: http://www.w3.org/TR/2009/REC-owl2-overview-20091027/

Welty, C., & Ide, N. (1999). Using the right tools: enhancing retrieval from marked-up documents, J. Computers and the Humanities, 33 (10), pp. 59 - 84.

Wüst, T., Nebiker, S., & Landolt, R. (2004). Applying the 3D GIS DILAS to Archaeology and Cultural Heritage Projects Requirements and First Results, *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences* (Vol. 35 Part 5, pp. 407 - 412). Natural Resources Canada.