Formal Ontology for Knowledge Management in Archaeology

Ashish Karmacharya, Christophe Cruz, Frank Boochs

C6

Formal Ontology for Knowledge Management in Archaeology

Ashish Karmacharya
Christophe Cruz
Frank Boochs

INTRODUCTION

With the advancement of survey technologies, data can be collected more accurately. On the one hand, this has provided a great advantage in the analysis process as we have more and diverse data to carry out the precise analysis. On the other hand, it has created difficulties in managing data with the existing database systems. This is a problem due to their size and diversity. This issue is even more visible in an industrial archaeological project. Indeed, the excavation sites are available for very limited time and thus the data need to be collected and stored rapidly. In addition, the diversity of the data makes the management of information with the existing database systems very complex. Hence, a lot of research is undertaken in the field of data indexation and information retrieval in order to reach the level where this vast amount of information can be managed through the knowledge defined by the archaeologists. Actually, the knowledge about the objects excavated from the site can only be defined by the archaeologists. Semantics is used to focus on data integration among heterogeneous datasets as well as to build up a common language in order to develop a common framework. Consequently, data is self-describing and allows generic and automatic processes. The lack of semantic data that describes objects spatially is an issue that we address. Actually, our proposition is based on a web platform which uses semantic web technologies and knowledge management processes. Our proposition focuses on the identification process which consists in managing data generated during the excavation process. The spatial data are linked to knowledge bases acquired during the identification process. By annotating data with semantic definitions, our web platform provides a semantic view on spatial data sets.

This project is motivated by the concern to preserve knowledge on archaeological sites which will soon disappear. Consequently, it is necessary to structure data as efficiently as possible for future reuse and to facilitate the interoperability. In addition, the process of capitalization of the knowledge will identify good practices and the web platform aims at sharing these good practices. The fact that a great amount of data is collected in a very short time means that this amount of data has to be managed rapidly. The structuring and the managing of a huge amount of data define a process that should be realized by an e-Archive. For all these reasons, we have focused our research on the use of knowledge management processes including semantic technologies in order to make it possible to realize a platform that constitutes an e-Archive.

The following section focuses on the formal ontology that is used as a semantic layer for our e-Archive. Section 3 presents the knowledge management domain. Section 4 explains how archaeology can benefit from formal ontology and KM. The last section concludes the paper.

FORMAL ONTOLOGY

The Semantic Web initiative of the World-Wide Web Consortium (W3C) was inspired by the vision of its founder, Tim Berners-Lee, of a more flexible, integrated, automatic and self-adapting web, providing a richer and more interactive experience for users. The W3C has developed a set of standards and tools to support this vision, and after several years of research and development, these are now usable and could make a real impact. These technologies benefit from more than 30 years of research in artificial intelligence and more specifically the domain of ontology. Ontology aims at representing knowledge about a specific domain that is understandable by both developers and computers. For this, ontology enumerates concepts and relations between concepts[1] and defines properties, functions, constraints and axioms.[2] The major issues in ontology development include ontology representation, ontology acquisition, evaluation and ontology maintenance.[3] Ontology representation is the main issue in ontology development because its representation has to be understandable by computers and humans. Consequently, an ontology representation language should provide representation adequacy for humans and inference efficiency for computers. Ontology dialects based on description logic (DL) provide a frame-based knowledge representation and profit from the expressiveness of DL reasoning systems. Ontology acquisition refers to the process of the ontology creation such as concepts, relations, individuals and axioms.

KNOWLEDGE MANAGEMENT

Knowledge management (KM)[4] is a process that helps organizations to identify, select, organize, disseminate, and transfer important information and expertise that is part of the organization's memory. It aims at exploiting an organization's intellectual assets for greater productivity, new value, and increased competitiveness. Organizations require knowledge management for users in order to help them understand the evolution of the knowledge on the organization.

When it comes time taking into account the huge quantity of documents, existing systems show some drawbacks concerning the following four points. First of all, **searching information** retrieves irrelevant and incomplete information due to the use of inappropriate words in the context of use during the keyword-based search. Secondly, **extracting information** is a human based process that is required because artificial agents fail to extract and to integrate distributed data sources. Thirdly, **maintaining unstructured information** such as texts, images, videos, scans, etc. is a tedious and a time-consuming process. It is even more tedious and time-consuming to keep these sources up to date and consistent. Fourthly, **generating automatic documents** according to the user profile through adaptive web systems is still very difficult to realize. Actually, the profile has to be defined regarding the context of use. The adaptive system has to be defined as well, and finally the semantic definition of unstructured data such as texts or images needs to be defined.

The KMS challenge is to identify and integrate the following four technologies in order to meet the KM needs of an organization. Firstly, **communication technologies** allow users to access required knowledge and to communicate with each other. This is usually materialized by a web-based application that allows all users to access all data of the platform regarding access right. Secondly, **collaboration technologies** provide the means to perform group work. The capitalization of knowledge concerns all the manners in which an organization works, and how all processes are divided and executed sequentially or at the same time. Thirdly, **multimedia storage technologies** use a database management system to capture, store and manage knowledge according to a large number of data source formats. This underlines the issue of the interoperability of the data source access which has to be dealt with by the KMS. Fourthly, semantic web technologies allow the definition of formal ontology. In this manner, a **semantic layer** is introduced in the KMS in order to create the meaning of data in order to be used by artificial agents. This is not the panacea but it improves greatly the capability of the system to search, to extract and to maintain unstructured information and to generate automatic documents regarding the requirements.

HOW ARCHAEOLOGY CAN BENEFIT FROM FORMAL ONTOLOGY AND KM

The semantics required by KM system is introduced by the semantic web technologies. Actually, all the semantic web technologies allow the definition of formal ontology that represent knowledge useable by machines. These semantics are useful during all the cycle of knowledge management. The following figure shows this cycle and the implication of the semantics during all this cycle. The first process is the creation of knowledge. By analyzing available data, archaeologists are able to identify piece of new information that defines in its context new knowledge. The role of the experts is important because they are the only ones who are able to create knowledge at this stage. Actually, after several cycles, machines are able to extract new knowledge by using intelligent processes.[5] In order to permit experts to capture and to refine knowledge a framework is required. The knowledge capture is undertaken with the help of a framework that defines formal ontology. The experts intervene at several levels. Firstly, they get involved at a schematic level which consists in defining categories of things and semantic relations between these categories. For instance, the semantic relation "isComposedOf" is a meronymy relationship which is transitive. If a house is composed of floors, and a floor is composed of rooms, then a house is composed of rooms, this assertion is possible because the relationship between categories of objects is transitive. Secondly, experts intervene at an individual level. This intervention consists in defining objects that are part of a category and to define their attributes and relationships. For instance, experts will define house n°1 and its associated floors and rooms. Consequently, the archaeologists are able to capture all information of an archaeological site and to define formally the knowledge about it. Refinement process consists in improving previously defined knowledge by correcting mistakes concerning the ontology consistency and satisfiability. A consistent ontology means that no axioms of the ontology can be true and false at the same time. For instance, a dog cannot be a cat at the same time. A satisfiable concept is a concept that accepts individuals. The knowledge storing process consists in storing knowledge in order to make it persistent. Usually, this is realized using a knowledge database system which permits to store but also to retrieve information. The subsequent processes are the management processing and the dissemination processing of knowledge. They are related to a common process which is the collaboration between experts. In order to make this collaboration possible, information and communication technologies are used. This collaboration is possible by using a collaborative web platform which allows the user to access, to retrieve and to manage information. Multimedia data can be also managed and can be indexed by the knowledge base with the help of the previously defined ontology. At the same time, each modification that is made on data is accessible to all the experts that take part in the project. Consequently, they participate in the criticism of available knowledge in order to create or to update knowledge. From this point on, another cycle starts.

CONCLUSION

In this paper we have discussed briefly the use of formal ontology in knowledge management systems in order to help archaeologists to capture and manage knowledge. This is made possible by the introduction of the web semantic technologies. The principle of this process was used in the e-Archive project called ArchaeoKM in order to manage a great amount of archaeological data from an industrial site. Different sources are indexed with help of ontology and from this ontology some deductions are possible in order to find out new knowledge based on spatial data for instance.

[1] Guarino, N., Formal ontology, conceptual analysis and knowledge representation, International Journal of Human-Computer Studies 43, 625–640 (1995)
[2] Studer, R. Benjamins, R. and Fensel, D., Knowledge engineering: Principles and methods, Data and Knowledge Engineering 25, 161–197 (1998)
[3] Zhou, L., Ontology learning: state of the art and open issues, Information Technology and Management archive Volume 8, Issue 3, 241–252 (2007)
[4] D Fensel, Ontologies: a silver bullet for knowledge management and electronic commerce, Springer-Verlag New York, Inc, 2004
[5] Karmacharya, A., Cruz, C., Boochs, F., Marzani, F., ArchaeoKM: toward a better archaeological spatial datasets management, Computer Applications and Quantitative Methods in Archaeology (CAA), Williamsburg, Virginia, USA, 2009.

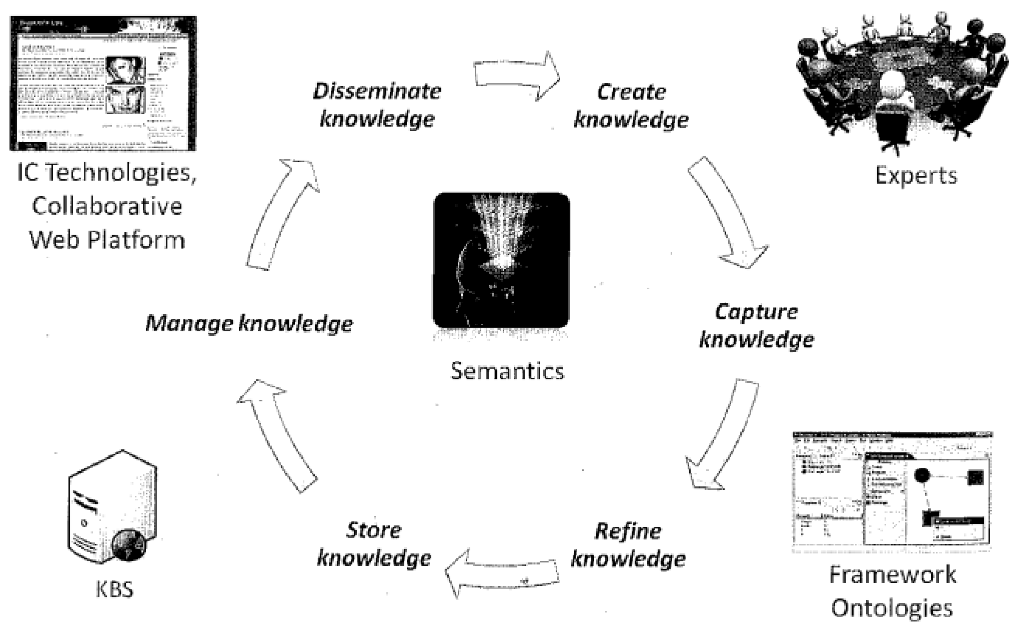

Figure 1. The Knowledge Management Cycle (© author)