Bases for the creation of ontology in the context of Archaeology

JUAN MARÍA FERNÁNDEZ GONZÁLEZ, ANTONIO POLO MÁRQUEZ, ENRIQUE CERRILLO CUENCA

ABSTRACT

The Web has become the main source of knowledge by the great amount of information that is available. Despite this amount of information makes the recovery of the information not an easy work. In order to solve this problem the scientific community is developing the Semantic Web that it includes, in addition to the information of the present Web, semantic descriptions of what it represents and it means part of that information.
We tried to structure the knowledge of a logical way, in such a way that from computer science applications it is possible the retrieval of information from the Web. In our case the knowledge domain is the one of Archaeology. Since in the case of Archaeology a unique conceptualization of the knowledge can not be applied, the possibility that fits so many ontologies exist as projects or specific necessities. At the present time, we are developing a system that allows to store and to query different ontologies defined under different contexts to gather part of the knowledge of Archaeology under the same system.

1. INTRODUCTION

During the last years the Web has become the biggest information system; its knowledge is continuously growing and incorporating new sources of information. It is a fact that the great amount of information makes even more complicated to access the information in which we are interested. In order to solve this problem the scientific community is developing the Semantic Web. "Semantic Web" is the expression used to describe an information system that not only contains data, but the semantic descriptions of what are and what means the data.
In the case of Archaeology, the Web is not just a source for divulgation of research data, some catalogues of artefacts or sites are available on line, so the possibilities of using the Web as a source for archaeological research are increasing continuously. Thus, it would be possible to retrieve information with semantic content. The Semantic Web improves the possibilities of search, by discriminating non-relevant data. For instance, if we search a term such as *sigillata* on the Web, the search engine would retrieve all the web pages in which this word appears. However, the same search on the Semantic Web would give us a list of web sites that include the concept of "roman pottery" and other wider or even more specific related concepts.
In order to do so, the Semantic Web is based on "ontologies". In this paper we present our first results on applying ontological and semantic techniques to Archaeology. We have designed an application, ArqueOnto. This work is being performed within a wider project oriented to the exploit of spatial and temporal archaeological knowledge on the Semantic Web. Three research groups are participating in this project, from which two of them are developing their own ontologies within the domain of Archaeology. A first group from Jaume I University has developed an ontology applied to the study of Spanish Mediterranean shores along Early and Late Prehistory, and from the EXIS group (University of Extremadura) we have designed another ontology oriented to the Early Neolithic in inner basin of Tagus river. We have used the field data of Los Barruecos site, an Early Neolithic site located on the interior of Iberian Peninsula. Both ontologies represent different theoretical positions, but also reflect the particularities of material culture in both times and territories.
Initially we verified that it was not possible to align both ontologies to access data, so instead we designed a mechanism to provide information to the final user of the application depending on its particular interests. ArqueOnto prototype application tries to go further and solve the problem of knowledge integration, using an Ontology of the Context (OC) that will provide information depending on the interest of the user.

2. PRINCIPLES OF THE SEMANTIC WEB: THE ONTOLOGIES

During the last years the Web has become a great source of information where users can get data of any type. Nevertheless, the amount of available data and its heterogeneity makes not easy to retrieve the right information that the user is looking for, or even the method to present it. The Web must be a way to share and process information among applications in order to optimize resources (duplicity, inconsistences) rather than a user-oriented interchange. The Semantic Web has been created to achieve a source of understandable information for the user and which can be managed by computers. Thus, it can be set an intelligent access to information.
In order to make comprehensible the information and other elements of the Web (sounds, videos, images or data) for users and computers, these elements must be identified by some mechanism which should classify them by adding some kind of semantic labels. In order to do so, ontologies are defined from different domains to represent the semantic information of web objects. Ontology" is a term lent from Philosophy, which refers to the general being and its transcendental proprieties.

In Informatics, this word has adopted a different meaning and can be understood as "a formal and explicit specification of a shared conceptualization". That definition was enunciated by Gruber (1993) in a very first work about the informatics use of ontologies. From the previous definition we can extract the following conclusions:

- Explicit specifications. Concepts, relations, instances and axioms are explicitly defined.
  - Concepts are basic ideas which are intended to be formalized. Concepts can be classes of objects, methods, planes, strategies, processes of reasoning, etc.
  - Relations represent the interaction and the link between concepts from the same domain.
  - Instances are employed to represent objects from a concept.
  - Axioms are theorems declared on relations which must comply with the elements of the ontology.
- Formal. The ontologies are understandable by any software.
- Conceptualization. Abstract model from a certain domain of the world.
- Shared. Knowledge is shared in a community.

The main inconvenience of the enunciated definition is that is eventually impossible to achieve an agreement among the archaeologists to get a "shared conceptualization" within the domain of Archaeology. Thus, different theoretical and methodological frameworks do not allow archaeologists to share knowledge by the means of the same conceptual scheme, as is possible in other sciences. This is a fact in a moment in which post-processual Archaeology has definitively broken down and criticised the "general laws" and the systemic perspective of culture. Nowadays, it seems to be as many conceptualizations as archaeologists, so there is an initial trouble to share knowledge and retrieve it from different data sources, since ontologies can be built from different theoretical postulates. A great effort to conceptualize about the documentation of cultural heritage is offered by CRM conceptual model (Crofts *et al.*, 2004), but nevertheless, it does not reflects the particular needs of archaeological research, and especially during the fieldwork.
Ontologies are defined using different standard informatics languages (URI, NS, XML, RDF/S, OWL, OWL-S, etc), defined by the World Wide Web Consortium (http://www.w3.org). As it happens with the Web, in the Semantic Web we can find a great amount of ontologies which can be applied to different domains. The trouble is that the diversity of ontologies can be confuse and would produce the opposite effect from they were created. So, we must arrange the ontologies depending on the domains they were crated to avoid ambiguity, since the knowledge represented must be shared within a community.

3. ONTOLOGIES IN ARCHAEOLOGY

3.1. CRM-CIDOC MODEL

As we stated before, we have not found an ontology focused on the archaeological knowledge valid for our research. Nevertheless the International Committee for Documentation of the International Council of Museums (ICOM-CIDOC) has developed a conceptual model of reference (CRM). This model is known as *CRM-CIDOC*, and contains concepts of general utility in the domain of Heritage and Museums ("definitions and a formal structure to describe concepts and relations used in the documentation of cultural heritage").
Initially, we tried to base our ontologies in CRM-CIDOC as a high-level ontology, deriving specific classes of our application from wider concepts defined in reference model. However, during the developing of the application was obvious that it was a problematic decision, since CRM-CIDOC conceptual model was very abstract for our needs and an eventual adaptation would use too many resources. So we decided to develop our application without this conceptual model, considering a possible alignment in the future by the use of versioning techniques.

3.2. UEX ONTOLOGY

As an initial level, he have defined that we call the "Knowledge Base" (KB), the set of ontologies that tries to represent a part or the whole domain of Archaeology. In our case, the KB is composed by three ontologies: the first of them has been developed at Jaume I University (UJI ontology) and represents the domain of Late Prehistory in the Spanish Mediterranean shores. The two remaining ontologies were thought initially as an unique ontology (the UEX Ontology) as we will explain later.
The UEX ontology has been arranged from five main concepts that try to represent any cultural behaviour at any time or space. Before making the prototype we have discussed about the epistemological organization of Archaeology.
As a starting point, we can state that archaeological research can be understood as a multidimensional level, composed by different levels of conceptualization. In one hand, we use abstract concepts and figures, on the other hand, our research has a strong component of material culture and is often dealing with material elements from the past. Moreover, we can quote even more evident opposites: the plane of research made from the present time and the plane of the social groups that lived in the past

That is, in fact, the duality of agency proposed by some theorist. A wider definition about the role of "agency" in Archaeology can be found in a work about this subject written by Barrett (2004, 141). Thus, following this theoretical perception, we have considered that there are two agents implied in the analysis of the archaeological record: the researcher and the object of the research. Both agents would produce different actions. The actions produced by the researcher would be directed to increase knowledge about the object of research. Meanwhile, the actions made by the object of research are only recognized by the means of the process of archaeological research.

In that way, we can assert that the relationship established between the researcher and the object of research is the base of archaeological analysis. We have tried to make a very first conceptualization of Archaeology in an ontology.

This ontology has been divided in two parallel planes. Plane A represents the point of view of the reasearcher and his personal conditions. On the contrary, "Plane B" has been described to define the archaeological evidence and human behaviour from the social groups been analyzed by the researcher. These ontologies that were thought to represent general archaeological concepts, with a special reference to Late Prehistory concepts in lower levels of conceptualization. In order to represent knowledge we have chosen five main concepts and a dual plane structure.

We have chosen five main concepts with their respective sub-classes and instances: time, space, matter, action and agent. We can represent with them any phrase from natural language and refer any human behaviour form the past and any process of the archaeological research.

In the "Plane A" ontology, *the agent* is any person implied in the process of research in Archaeology (i.e. archaeologists, biologists, etc). With *the action* we can describe any act produced by *the agent* to document a site, a set of items or even any act made at the laboratory. *Time* concept is related to the concrete time in which the act is produced, so we can express that any fieldwork is done during a concrete lapse of time. Meanwhile *matter* and *space* are concepts also included in the "Plane B". In the case of the matter, often related in our ontology to archaeological items, the *matter* which is analyzed by the *agent* of "Plane A" was produced by "Plane B" *agent*. The same circumstance is observed with the *spatial* concepts, since the archaeologist tries to analyze the landscapes produced by past social groups.

On the opposite, in the "Plane B" is defined by the past groups being analyzed by the researcher, *the agent* is not a concrete person, but an abstract collective entity which represents the groups implied in the construction of the archaeological record. *The actions* are recognizable by the means of the archaeological interpretation of sites, artefact or any cultural element.

Other subclasses have been defined depending on these concepts. For example, the concept "Stratigraphical Unit" depends on the "site" and it is as well a *child* of "space" concept. By defining classes and sub-classes on both planes we have achieved a vast conceptualization of Archaeology to be used as an ontology.

4. THE "CONTEXT"

The concept of "context" is frequently used in Archaeology to make reference to the social, cultural or even the personal conditions implied in the formation of the archaeological record, and, what is more important, its individual interpretation (Hodder, 1986). However, from the point of view of Informatics, "context" must be understood in a different way. Some authors have given their own definitions about the use of "context" in different domains. We will enumerate briefly the different works in which this concept has been focused. The context can be described and used as:

- "Alternative worlds", being a "world" the environment in which the data can obtain meaning (Stavrakas *et alii*, 2004).
- A set of statements, within a domain, that depends on a joint of rules. Thus, a context is represented by the means of an ordered triple that contains definitions, rules and values of the definitions. (Zhu *et alii*, 2004).
- A part of the ontology that contains concepts, relations and instances from a local model, not shared by a community (Bouquet *et alii*, 2004).
- "any information about the circumstances, objects or conditions surrounding a user that is considered relevant to the interaction between the user and the ubiquitous computing environment" (Ranganathan & Campbell, 2003, p. 353).
- can be used as "an automated environment to solve semantic heterogeneity between Federate Systems" (Buccela *et alii*, 2004).

In our case, we have been developing a mechanism that let us to retrieve semantic information from ontologies. So, we propose that context can be described as "any information about the circumstances, objects or conditional surrounding all the ontologies from a given domain, which will let to manage them jointly". Thus, our prototype "ArqueOnto" is able to retrieve different elements of the ontology from the Knowledge Base depending on the interests of the researcher who uses the application, as it is described later. In fact, we can understand "the context" as a set of instances of OC which refers to the Knowledge Base. In order to describe a context by the means of the ontology, we need to create instances in the Ontology of Context (OC), which define the point of view of the Ontology of Archaeology.

The use of an ontology that explicitly defines these contexts let us use a shared vocabulary and the possibility of reasoning about the terms of that vocabulary. When designing the ontology, we have employed the same structure of any phrase

from natural language. Thus, we have obtained a set of concepts and relations between them, as shown in figure 1 and table 1, where we have represented the equivalence of natural language and the concepts and relations. By their relevance on the context definition, we have emphasized the prepositional phrases of time and space.

Table 1 – Relations between terms from natural language and the OC ontology.

Natural language	**Ontology of the Context**
Subject	"**Agent**" that produces an "**action**"
Verb	"**Action**" produced by the "**agent**"
Direct object	"**Matter**" on which an "**action**" is produced
Indirect object	"**Agent**" on which an "**Action**" is produced
Prepositional phrase (time)	"**Time**" during an "**action**" is produced
Prepositional phrase (space)	"**Space**" at which an "**action**" is produced
Prepositional phrase (others)	"**Matter**", "**Action**" or "**Agent**" inform of any circumstance about "Action"

An important character of this structure is the easiness achieved when we answer questions as "what?", "who?", "when?" or "where?". These questions are answered by the means of a relationship between concepts defined in the OC, as shown in table 2.

Table 2 – Example of use of basic questions determined by their relations between concepts.

Basic question	**Determinant relationship**
Who?	Produced_by
What?	Act_on
When?	During
When?	Is_at, pass_at
How?	How
Why	Caused_by,
With how?	Presence_of
With what?	With

In the following section we will represent the context by means of instances of the OC concepts. The relations in OC, and in the ontology of our Knowledge Base, allow using it jointly to deduce new knowledge. For example, if we have a set of ontologies with sites containing information about coordinates, artefacts, etc, we can store other kind of information such us the author, or the time in which ontologies were created to use it of a shared form. If in the bibliography we found papers which state that the first evidences of agriculture in the inner basin of Tagus were about 3500 cal B.C., and we know, that after 2003 studies, the C^{14} dates demonstrate that agriculture was known earlier (5200 cal B.C.); then we can deduce that the sources used are previous to 2003. In this case, the time in which the ontology was defined is essential to share knowledge.

5. ARQUEONTO: A PROTOTYPE OF APPLICATION

The described system has been applied in a domain of ontologies to be used in Archaeology called "ArqueOnto". In one hand, this application allows us to maintain successive version of the evolution of an ontology. On the other hand, it eases the integration with other ontologies from the same domain. In order to complete the system, a filter of semantic information has been created which will select the proper ontologies for a context.

To make queries based on the context in ArqueOnto, we have defined a very simple language. Queries are done from a list of n-terms, separated by a slash, where:

- The first n-1 terms are related to instances from OC and define the context of the search.
- The n^{th} term, called "term of search" tells us what we are searching on the context of the n-1 first terms, which can be related to concepts, instances or relations of Knowledge Base.
- The terms of context and the term of search are optional. In figure 1, we show ArqueOnto process of working, which is divided in three steps:
  - Dividing the search path into "context" and the "term searched"
  - Extracting the knowledge defined by the user from the Knowledge Base, using the OC.
  - Retrieving the knowledge associated to the term searched from the selected ontologies.

In figure 2, we present the environment of the application and the way of presenting data.

6. CONCLUSIONS AND FUTURE WORK

The main conclusion if that we have an efficient framework to start developing more complex, and useful applications. Moreover, the logical structure of this ontology is plenty of linguistic possibilities to analyse human behaviour, so we can express any action which is performed by an agent during a time, in a given space and on a matter. Nevertheless, it is still a formal language which does not let the use of rhetoric and other metaphors as it is usual in contemporary Archaeology (Hodder, 1993).

In the case of ontologies like plane A and B that are described based on the five concepts of the context, we can automatically extract the point of view of these ontologies. On the contrary, the context of UJI ontology was extracted in a manual way.

Many characteristic can be improved. We are working to develop a multilingual environment which can include ontologies in other languages, for example, the CRM-CIDOC. Moreover, since data is stored in relational databases it would be interesting to use SQL queries to retrieve data from them. The last step would be as necessary as the yet mentioned, the design of a collaborative environment to be used for different users.

REFERENCES

BARRETT, J. C. (2001) – Agency, the duality of structure, and the problem of the archaeological record. In Hodder, I., (ed.), *Archaeological theory today.* Cambridge: Polity, p. 141-164.

BOUQUET, P., *et al.* (2004) – Contextualizing ontologies. *J. Web Sem.* 1(4), p. 325-343.

BUCCELA, A.; CECHICH, A.; RODRÍGUEZ BISABROA, N. (2004) – *An Ontology-Based Environment to Data Integration.* IDEA2004.

CROFTS, N., *et al.* (2004) – *Definition of the CIDOC Conceptual Reference Model.*

GRUBER, T. (1993) – Toward Principles for the Design of Ontologies Used for Knowledge Sharing, *Technical Report KSL-93-04*, Knowledge Systems Laboratory, Stanford University, CA.

HODDER, I. (1986) – *Reading the Past: Current Approaches to Interpretation in Archaeology.* Cambridge. Cambridge University Press.

HODDER, I. (1993) – The narrative and rhetoric of material culture sequences. *World Archaeology* 25(2). P. 268-282.

RANGANATHAN, A.; CAMPBELL, R. H. (2003) – An infrastructure for context awareness based on first order logic. *Personal Ubiquitous Computing*, 7(6), p. 353-364.

STAVARKAS, Y., *et al.* – Representing and Querying Histories of Semistructured Databases Using Multidimensional OEM. In *Information Systems journal*, Vol. 29, Issue 6, p. 461-482.

ZHU, H.; MADNICK, S. E.; SIEGEL, M. E. (2004) – Reasoning about Temporal Context using Ontology and Abductive Constraint Logic Programming. *MIT Sloan Working Paper* No. 4498-04; *CISL Working Paper* No. 2004-15.

FIGURES

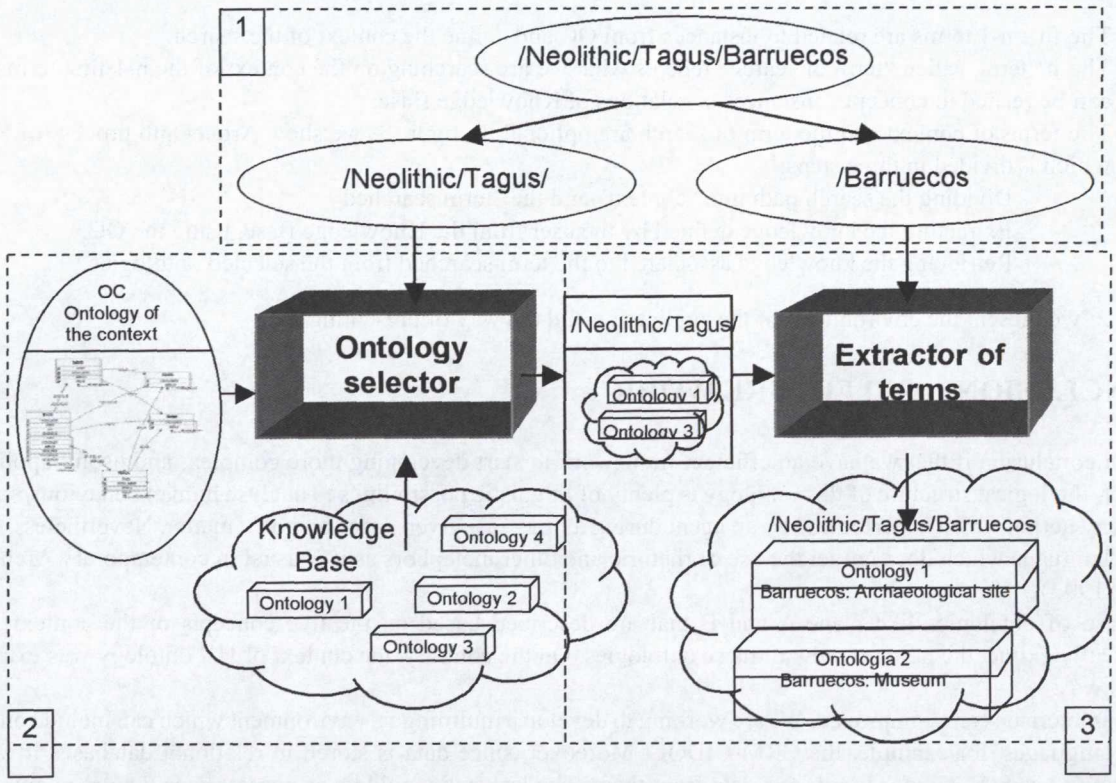

Fig. 1 – ArqueOnto, scheme of work.

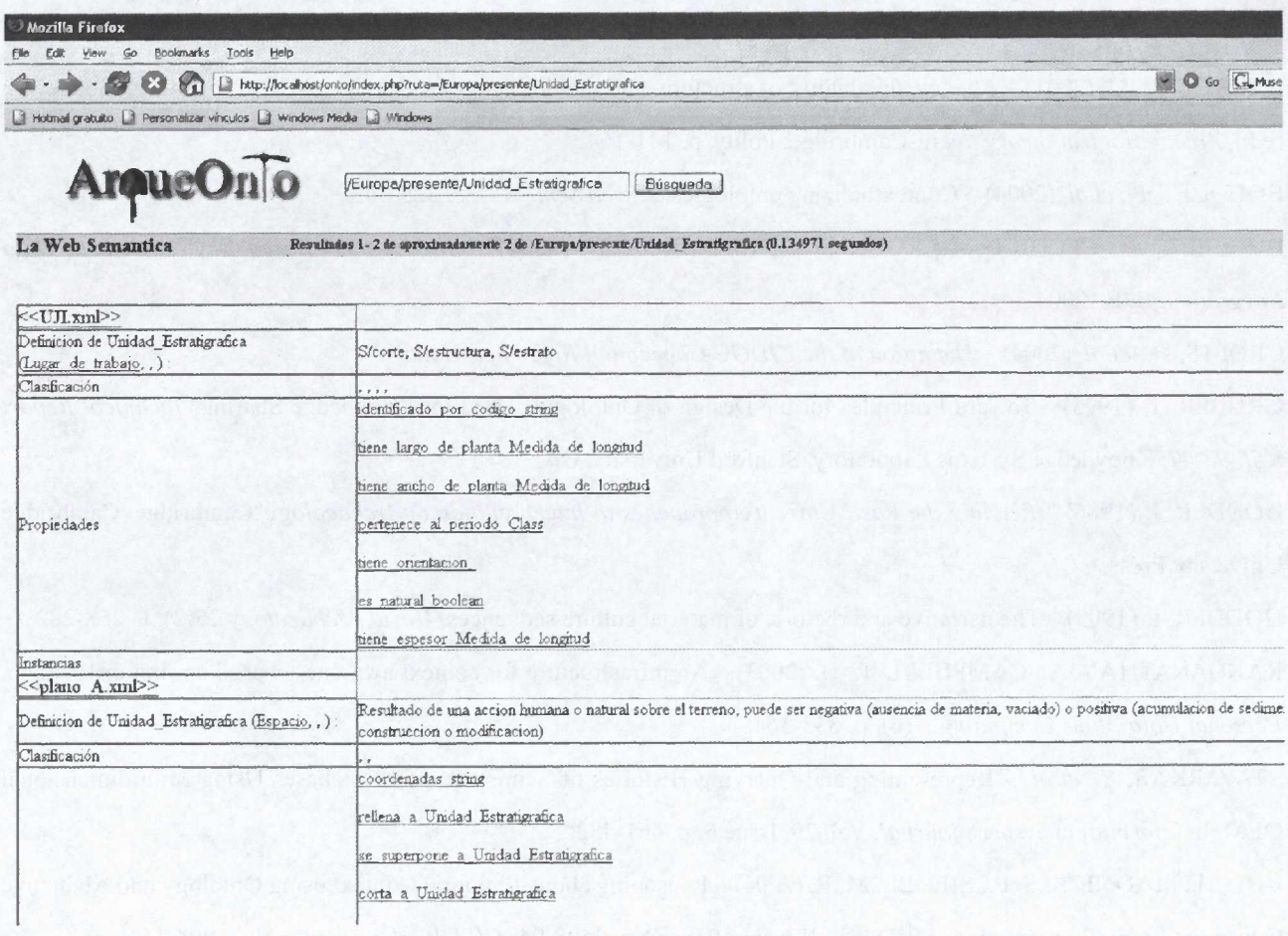

Fig. 2 – Screenshot of a query in ArqueOnto.