Ontologies as a Reference Framework for the Management of Knowledge in the Archaeological Domain

Oleg Missikoff
LUISS "Guido Carli" University
Centro di Ricerca sui Sistemi Informativi
omissikoff@luiss.it

Abstract

An ontology is a shared understanding of some domain of interest. An ontology entails some sort of world view with respect to a given domain. It contains a set of concepts (e.g., representing entities, attributes, processes), together with their definitions and their inter-relationships; this is also referred to as a conceptualisation. In other words, an ontology is an explicit, agreed specification about a shared conceptualisation.
An ontology may have different degrees of formality but, necessarily, it includes a vocabulary of terms with their meaning (definitions) and their relationships. According to [IDEF5], an ontology is a domain vocabulary containing a set of precise definitions, or axioms, that:
- provide the meaning of terms,
- enable a consistent interpretation of the terms defined in the vocabulary.

This document concerns the use of SymOntos (Symbolic Ontology System), a software prototype developed by LEKS (Laboratory for Enterprise Knowledge and Systems), at IASI-CNR in Rome, for the definition and management of archaeological domain ontologies.
The case study will be represented by the classification of Iron Age fibulae (brooches) from the cemetery of Quattro Fontanili near Veii (Rome).

Introduction

The Cultural Heritage (CH) sector is recently faced with new issues that need to be addressed quickly and competently. Among those we can list all problems related with the management, sharing and dissemination of the vast amount of digital resources now available, and the difficulties in transforming such data and information in knowledge.
For handling this ever increasing flow of information, supported by the diffusion of internet and ICTs, scholars are reconsidering studies on knowledge developed in the field of Artificial Intelligence (AI), almost 50 years ago. It is in fact in the summer of 1956 that Herbert Simon, Marvin Minsky, Claude Shannon and others meet at Dartmouth College of Hanover, New Hampshire, for discussing about the possibility of simulating human learning and reasoning processes by using rather recently invented machines: the computers. They think that time has come for this kind of studies to constitute an autonomous discipline; not without contrasts they give it the name of "Artificial Intelligence". AI provided the first, pioneer studies on knowledge and especially on its representation, producing a set of formalisms that would have paved the way for recent applications of Knowledge Management (KM) techniques in the business domain and for the state-of-art Semantic Web issues.

For more than thirty years, in fact, studies on knowledge have been confined in research labs, until the mid 1990s when the book by Nonaka and Takeushi, "The Knowledge-Creating Company" (Nonaka and Takeuchi 1995), opened to KM the doors of business and enterprise management domain. Through this contribution it became widely acknowledged that the competitive advantage of some of the world's leading companies was being carved out from those companies' knowledge assets such as competencies, customer relationships and innovation. Managing knowledge therefore, appeared a mainstream business objective as other companies sought to follow the market leaders. Technological solutions in mid 1990s, however, were still immature for permitting a successful and widespread implementation of KM techniques in enterprise management activities. After a few years of great interest, numerous theoretical publications and some failed implementation attempts, at the end of the nineties it looked as knowledge management was destined to be confined to the "management fad graveyard".

Meanwhile, the proliferation of contents and resources available on the internet has posed the problem of extracting meaningful information from an almost infinite repository: the world wide web. Again, a viable solution was spotted through the implementation of techniques and methods derived from the evolution of those AI pioneer studies on knowledge. This time the solution was called "The Semantic Web" and the proponent's name needs little presentation: Tim Berners-Lee (Berners-Lee, Hendler and Lassila 2001). The most remarkable advantages the semantic web should provide consist in the possibility to perform searches based on concepts instead of terms, therefore reducing the chances of confusion, and allowing software agents to carry out complex tasks for humans.

The Semantic Web, according to Berners-Lee, should substantially rely on well formed, interoperable and sharable contents. These conditions can be guaranteed by a recently developed knowledge organisation framework whose interest is rapidly growing in the academic community: Ontologies (Uschold and Gruninger 1996).

The aim of this contribution is to provide CH scholars with a practical understanding of the basic principles of ontologies, and of the possible advantages deriving from their application in their domain.

Ontologies and their Use

This section will be devoted to a brief description of the nature of ontologies together with the advantages that could derive from their adoption as a domain definition reference framework.

An ontology is an explicit, agreed and shared definition of a portion of reality by means of a conceptual model. This model may exist in someone's head or be embedded in a software or information system, in an object or in a process. The task of an ontology builder is to identify the model and make it explicit. This allows the model to be accessed by, or communicated to, a wider range of potential users, be they people, organisations or software agents.

With respect to a thesaurus, an ontology aims at describing concepts, whereas a thesaurus aims at describing terms. An ontology can be seen as an enriched thesaurus where, besides the definitions of, and relationships among, terms of a given domain, more conceptual knowledge is represented. With respect to a Knowledge Base (KB), an ontology can be seen as a KB whose goal is limited to the description of the concepts necessary for modelling domains. A KB, in addition, includes the knowledge needed to model and elaborate a problem, or to answer to queries about a domain.

An ontology is composed of:

- a set of concepts (e.g., entities, attributes, processes) regarding a given domain
- the definitions (´conceptualization´) of these concepts
- the relationships interconnecting entities within a given domain

Constructing an ontology implies a series of basic steps to be carried out, these are:

- examining the vocabulary that is used to describe the characteristic objects and processes of the domain
- developing rigorous definitions about the basic terms in that vocabulary
- characterizing the logical connections among those terms

For what concerns a practical use, at a higher level we can subdivide the space of uses for ontologies in the following four categories:

1. ´communication´ and ´cooperation´ among people
2. better institutions organization
3. ´interoperability´ among systems
4. ´system engineering benefits´(reusability, reliability, specification)

Applications in the Cultural Heritage Domain

In spite of the fact that, as we all know, the CH domain is normally rather slow in responding to pressures from "the outside world", this time new factors make think that there will be a faster reaction. Apart from the already mentioned Semantic Web and the need to organise the growing amount of data available, there is an increasing effort in the conversion of traditional cultural heritage resources in digital format.[1] These activities, strongly fostered by the European Commission in 1999 with the *eEurope* initiative[2] and in 2001 with the Lund meeting,[3] are calling for new researches on the conceptual organisation of digital resources that are already available or will be soon produced.

In the CH domain, an extensive contribution has been provided by the International Committee for Documentation of the International Council of Museums (ICOM-CIDOC), and is represented by the CIDOC CRM (Conceptual Reference Model). This model provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation:

"The CIDOC CRM is intended to provide a common language for domain experts and implementers to formulate requirements for cultural heritage information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives".[3]

Another valuable source of information has been provided by the "DigiCULT" FP5 IST Support Measure with its Thematic Issue No.3: "Towards a Semantic Web for Heritage Resources".[4]

In the next chapter, as a practical example, a case study will be presented: that is the classification of iron age fibulae (brooches), from the proto-historical cemetery of Quattro Fontanili near Veio (Rome).

The Methodological Foundation

This case study derives from a thesis dissertation concerning the classification of a set of archaeological objects, by means of a methodology for conceptual modelling developed in the Institute for Systems Analysis and Informatics (IASI) of the Italian National Research Council (CNR). At present a new prototype tool for symbolic ontology management, XML based, called *SymOntoX*[6] is being developed from a previous version (called Mosaico (Missikoff 1996)), and the knowledge base is being imported in this new tool.

SymOntoX is an Ontology Management System designed for supporting the construction and management of domain ontologies. It allows to manage several ontologies and supports different kinds of users with different access rights, it has multilingual capabilities and permits remote access through the Internet. SymOntoX is based on the OPAL (Object, Process, and Actor Language) metamodel for knowledge representation (Missikoff and Taglino 2002).

A metamodel is a set of definitions and rules which allow the definition of a modelling language. In this context the metamodel allows to specify the features to be filled and the rules to be followed during the definition of an ontology. The two main mechanisms used for classifying the material are the structural specification and the hierarchical organisation. The type structural specification consists in supplying the list of properties and, for each of them, the type corresponding to the values that they can assume.

<type-name> := [<prop-name> : <prop-type>
<prop-name> : <prop-type>]

In a Knowledge Schema, it is possible to organize the type definition within an ISA hierarchy. That is basically a generalization/specialization relationship between types. For instance we can declare that: "cup ISA vase", intuitively this statement shows that all the characteristics of vase are encountered in cup as well. Besides that, the latter could have additional characteristics which are not necessarily encountered in all the vases. This principle is often referred to as principle of inheritance because the type cup inherits all the characteristics of the type vase.

In the extensional level the ISA relationship turns into an inclusion relationship between classes. The example shows that the class of cups is contained in the class of vases. These qualitative considerations are rigorously described by the system, through strict criteria that guide the building of ISA hierarchies.

The ISA relationship implies that the type being defined be a specialization of the types appearing under ISA. Beside that, the principle of inheritance is also used to obtain a more compact schema description. Inheritance can be single or multiple, if in the ISA construction appear one or more supertypes. Instead we talk of absolute inheritance when the properties of the supertype are inherited without being modified.

Having given a type, the creation of a subtype is performed through specialization. The mechanisms of specialization must be always respected in defining a type using the ISA construct. Those mechanisms are of two basic sorts, specialization by specification and specialization by restriction:

- Specialization by specification. This mechanism of specialization requires the addition of new properties to those already defined in the supertypes (which, as stated above, are inherited by the subtype). If the supertypes are two or more and have properties in common, inconsistencies can arise. This is a critic point of the multiple inheritance, the problem has been already faced in the literature and there are different ways to solve it, but their description goes beyond the purposes of this work.

- Specialization by restriction. The mechanism of restriction allows to refine in the subtype one or more properties already defined in the supertype (this mechanism is called overriding). The overriding is performed essentially operating on the two property typing tools mentioned above: the explicit listing of allowed property values in the case of categorical variables and/or the range of allowed property values in the case of continuous variables.

The Case Study

The present work aims at describing the operational steps followed in analysing a group of artefacts which have traditionally presented relevant taxonomic problems. The case study is represented by the conceptual modelling of the "Fibula" type (en. brooch), and of some of its specializations, starting from the analysis of materials from surveys and excavations in the area of the villanovan cemetery of "Quattro Fontanili" near Veii (Southern Etruria). According to the Dizionari Terminologici of the Italian Institute for Cataloguing and Documentation (ICCD), the fibula is an "object of ornament used to fix parts of clothes. It is constituted by a bow ("arco" in Italian), a needle ("ago" in Italian) generally connected to the bow by means of a spring ("molla" in Italian), and a catch ("staffa" in Italian) in which the point of the needle is inserted."

A special interest for these materials consists in the fact that they have been previously and thoroughly analysed (Close Brooks 1965:53pp, Guidi 1993, Toms 1986:41pp), thus giving a chance for useful comparisons. The present classification has been performed starting from the pioneer work of J. Sundwall in her Die alteren italienischen Fibeln, published in 1943, and developed according to a hierarchy of main attributes that generates, as a consequence, a hierarchy of types progressively more specialized. The hierarchy of main attributes is the following: 1) shape of the bow (arco), 2) shape of the catch (staffa), 3) decoration.

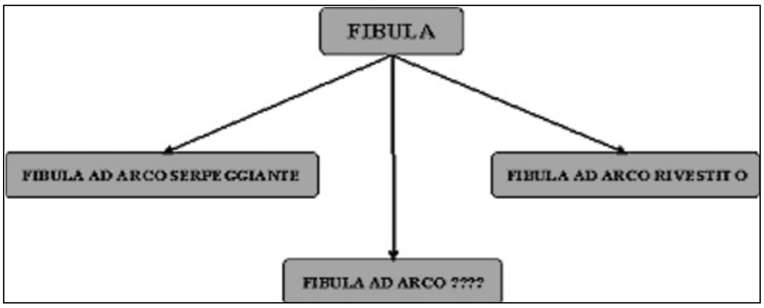

Figure 1

The use of SymOntoX and the OPAL methodology has, already at this level, stimulated a series of observations that have led to a partial redefinition of the hierarchical organisation for the Fibulae class. In fact, even if we can substantially agree in the definition of a first level based on the artefact function, that of fixing parts of clothes, in the definition of the second level a first problem arises. According to the traditional hierarchical organisation, the type "Fibula" is split, at the second level, in three subtypes: 1) Fibule ad arco serpeggiante, 2) Fibule ad arco rivestito, 3) Fibule ad arco. From a first intuitive observation it is possible to note that in cases 1 and 2 the "arco" is refined by the attributes "serpeggiante" and "rivestito", whereas in case 3 it is not refined at all. (Fig.1)

Figure 2a fibula serpeggiante

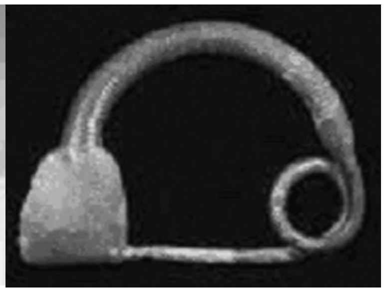

Figure 2b fibula ad arco

In a mathematical/logic context and according to the classification principles previously outlined, this automatically assigns the "Fibula ad arco" entity at a more generalised level. Furthermore, from a morphological point of view, to define the body of a "Fibula serpeggiante" as a "bow" can be misleading, being rather far from a standard definition of bow or arch that can be found in dictionaries or encyclopaedias which refer always to a curved line.

This observation is reinforced by the fact that, in archaeological deposits, the "Fibule serpeggianti" show a well defined social meaning being found exclusively in male burials, whereas the "Fibule ad arco" are characteristic of female burials. Having considered the above issues, we propose an organisation of the second level, defined as "generic shape", that gathers the concepts of "Fibula serpeggiante" (Fig.2a) and "Fibula ad arco" (Fig.2b). This last concept is further refined at the third level in "Fibula ad arco rivestito" and "Fibula ad arco decorato".

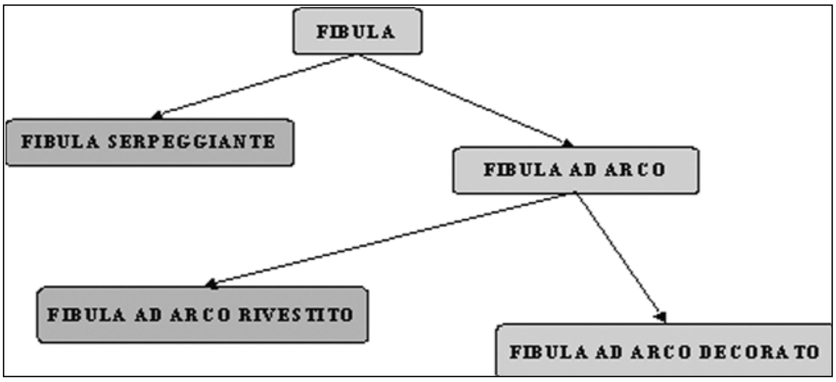

Figure 3

In the third level it is therefore defined the decoration technique of the "Fibule ad arco"; here the element of distinction is represented by the type of decoration chosen for enriching the objects.

This brief example clearly demonstrates that classifications performed on an intuitive base, without relying on a formal theoretical foundation, could easily lead to inconsistencies and produce an output which shows evident weaknesses from a logical point of view. The construction of domain ontologies can certainly provide a viable solution for these problems but, so far, it is evident a lack of real case studies on which stimulate a debate. The call for both scientific communities (artificial intelligence and archaeology) is to intensify collaborations for testing methodologies and tools on real case studies and, in the long run, forming "hybrid" researchers that could rely on a more integrated background.

Conclusion

In this contribution it has been proposed an ontological analysis of a real case study as general conceptual reference framework oriented towards the support in the definition of entities in the archaeological domain. Using ontologies could also facilitate the knowledge sharing among researchers for guaranteeing a more extensive reusability of results from scientific researches. For performing the analysis it has been used "SymOntoX", an ontology management system, XML based, developed in the LEKS (Laboratory for Enterprise Knowledge Systems) of the IASI/CNR.

1 See http://www.minervaeurope.org.

2 See http://www.cordis.lu/ist/ka3/digicult/eeurope-overview.htm.

3 See http://cidoc.ics.forth.gr.

4 See http://www.digicult.info/downloads/ti3_high.pdf.

5 See http://www.symontox.org.

References

BERNERS-LEE, T., HENDLER, J. and LASSILA, O., 2001. The Semantic Web. Scientific American May.

BERTOLAZZI, P., KRUSICH, C. and MISSIKOFF, M., 2001. An approach to the definition of a core enterprise ontology: CEO. In Luiss Edizioni (ed.), Open Enterprise Solutions: Systems, Experiences, and Organizations Workshop, Rome.

CLOSE BROOKS, J., 1965. Proposta per una suddivisione in fasi della necropoli veiente di Quattro Fontanili - Notizie degli Scavi:53pp.

GUIDI, A., 1993. La necropoli veiente dei Quattro Fontanili. Firenze.

GRUBER, T., 1993. A Translational Approach to Portable Ontologies. Knowledge Acquisition Vol. 5, no. 2.

MISSIKOFF, M. and TAGLINO, F., 2002. An ontology approach for Business and Enterprise modelling with SymOntoX. AICA XL Annual Conference, Bari.

MISSIKOFF, O., 1996. Application of an Object-Oriented approach to the formalization of qualitative and quantitative data. In Analecta Praehistorica Leidensia, Publications of the Institute of Prehistory, University of Leiden.

NONAKA, I. and TAKEUCHI, H., 1995. The Knowledge-Creating Company. Oxford University Press.

TOMS, J., 1986. The relative chronology of the Villanovan Cemetery of Quattro Fontanili at Veii. In AION Vol. 7:41pp.

USCHOLD, M. and GRUNINGER, M., 1996. Ontologies: Principles, Methods and Applications. The Knowledge Engineering Review Vol. 11, no. 2.