DATA RETRIEVAL FROM A CULTURAL KNOWLEDGE DATABASE

G. Marcos, H. Eskudero, C. Lamsfus , M.T. Linaza

Asociación Centro de Tecnologías de Interacción Visual y Comunicaciones, VICOMTech, Spain

ABSTRACT

This paper describes a domain ontology which has been designed within the European art-E-fact project to represent the associated knowledge and to enable the description, exchange and sharing of multimedia added-value content for the creation of artistic expressions both within and between cultural institutions. Within the art-E-fact project, authors access and retrieve all the content stored. Since authors do not mainly know how to deal with a database management, a mapping method has been developed. This paper presents the details of the developed mapping and content retrieval methodology through the domain ontology implemented for the project.

1. INTRODUCTION

Semantic Web technologies can enable the heritage sector to make its information available in meaningful ways to researchers, its own curators, the didactics departments and even the general public. Content generators will benefit from the ability to define an exhibition and have the entire process from the identification of the pieces to be shown in the exhibition to the final production of the exhibition.

Delivering these technologies to the heritage sector depends on the syntactical and semantic mark-up of content, the development of better knowledge analysis and modelling tools, widespread adoption of interoperable knowledge representation languages and the implementation of suitable ontologies. In most of this, the heritage sector is lagging behind.

For the semantic technologies to succeed in the field of information exchange and interoperability between cultural institutions, there is a great need not only to gain interoperability using standard ontologies, but to provide the cultural heritage community with user-friendly tools for inferencing and semantic information retrieving of the content stored in cultural databases. One should always keep in mind that the personnel from cultural and artistic institutions have not a technical background and most of them are not very familiar with dealing with databases.

This paper presents a contribution concerning a simple methodology for non-technical users in order to map or relate the cultural content with the storing databases. Section 2 describes in more detail different aspects of the art-E-fact ontology, including a brief description of the project, the scope of the ontology, the description of the metadata and the applied form. Section 3 presents the methodology applied to the mapping and finally, Section 4 provides some conclusions and further works.

2. THE ART-E-FACT ONTOLOGY

The art-E-fact ontology is part of the outcome of the art-E-fact project. Since the target of the art-E-fact project is to create stories about artworks and thus create art, we have to be taught by the experience that was gained the last 4000 years of civilization.

The domain metalevel ontology conception leads the author to assimilate the internal world of the creator of an artwork, create and tell stories. This is not just a conception of the experts performing the scientific diagnosis, but it is also a tool for artists, authors and content generators. Artists using the ontology have to create stories or experiences concerning one or more selected artworks, including its main features, technical data, historical context, etc. All this information is included within the Cultural Content concept. The domain metalevel ontology conception leads them to assimilate the internal world of the creator of an artwork, and create and tell stories.

In order to build the art-E-fact ontology, the main focus has been put on the target group that are using the Generic platform, that is, the artists. Among the three possible alternatives to define the classes, a combined development process has been used. This means that we have first defined the most representative concepts, and then we have generalised and specified them appropriately in order to get an accurate representation of the knowledge stored in the database.

Scoping the ontology has been mainly based on two brainstorming sessions with the artists and the content providers. Having these brainstorming sessions allowed us to produce most of the potentially relevant terms and phrases. At this stage, the terms alone represented the concept, thus concealing significant ambiguities and differences of opinion.

A clear issue that arose during these sessions was the terminology differences among different art styles, between the Greek traditional iconography and the traditional European painting schools. The concepts listed during the brainstorming sessions were grouped in areas of work corresponding to naturally arising sub-groups. Most of the important concepts and many terms were identified. The main work of building the ontology was then to produce definitions. During scoping, most of the important concepts and many terms have been identified.

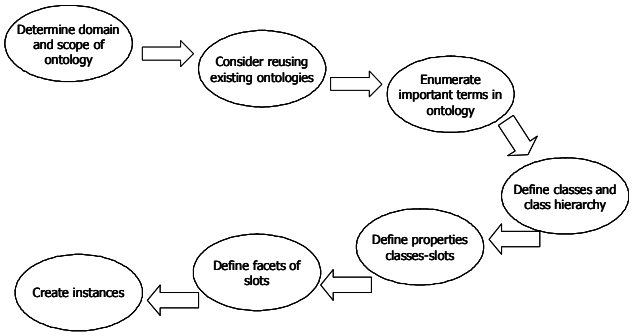

Figure 1 Methodology for the development of the ontology.

The methodology of the ontology building process has also included information about mapping the structure of the content database and the representation of the conceptual schema represented in the database.

Due to the nature of the project, we decided that we had to build an ontology from scratch. We were working in the Cultural Heritage context and although there is a very well known and established ontology (the CIDOC CRM) this did not cover the purpose of our commitment.

Protégé 2000 has been used as the ontology editor. It is a very powerful ontology editor that generates RDF and RDF(S) codes that can later be used in order to manage the ontology. Moreover, it also has the possibility to add a great variety of plug-ins that enormously increments its usability as an ontology development tool.

One of the main requirements was the management of the ontology during the development process in order to ensure that all relations and possibilities are taken into account. RDF(S) files have been used for this purpose, storing the information in a MySQL database.

A major advantage of these two programming languages is that their syntax is based upon XML, which has been widely adopted as a standard to share information on the web.

On the other hand, RFD and RFD(S) are a recommendation by the W3C to formulate metadata in the web. The basic structure of these languages is organized in triples, i.e. Subject (S), Predicate (P) and Complement (C), usually represented as P(S, C).

As stated previously, it is very important to manage the knowledge during the building process. Therefore, we need to semantically query the ontology. In order to achieve this objective using Sesame, the information of the ontology is transferred from Protégé to a MySQL database.

The RQL (Resource Query Language) is then used to query the database and check if the ontology really fulfils the requierements. One very important advantage of the RQL language is that its syntax is based on SQL, being very easy to be used. The following figure represents the query flow or process:

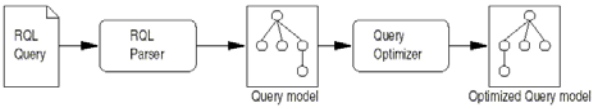

Figure 2 Parsing and optimization of the model.

The final result of the ontology can be graphically seen in the following picture. The key concept within the ontology is *Artwork*. The rest of the concepts depend on this, and all the information can be browsed through the graphical ontology representation developed for this project.

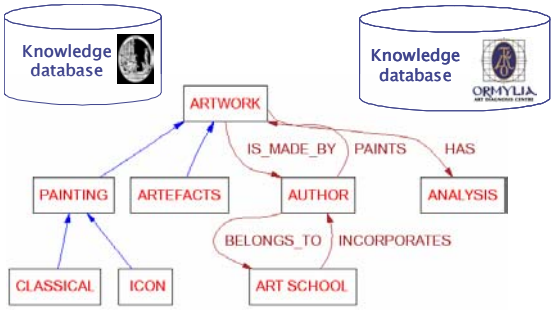

Figure 3 Graphical outlook of the ontology.

3. MAPPING THE ART-E-FACT DATABASE USING THE SINGLE CONSTRAINT MAPPING METHODOLOGY

The task of integrating heterogeneous information sources put ontologies in context. They cannot be perceived as standalone models of the world, but should rather be seen as the glue that puts together information of various kinds. Consequently, the relation of an ontology to its environment plays an essential role in information integration. In this paper, we use the term mapping to refer to the connection of an ontology and the information it describes [1].

3.1. The single constraint condition

One necessary and sufficient condition for the database to be mapped through the ontology is that there must be a field in one of the tables within the structure of the database (called hereby "DB entry point") which has to access to all of the fields of the tables that form the dataset.

In this way, it is possible to guarantee that once this field is correctly identified, queries to retrieve information from each of the fields in every table of the database are allowed.

3.2. The database architecture

The main idea concerning the design of the architecture of the ArtWorks Database (AWDB) system corresponds to the conception of the ontology. In this architecture, the main levels of knowledge are treated as main "thematic entities" for the database. The entity of the identification-work is the table "ITEM" (DB entry point) and all the tables are related to this table. There can never be a table with no connection to at least one more table, since that would automatically mean that there is some data that is not reachable by the querying engine.

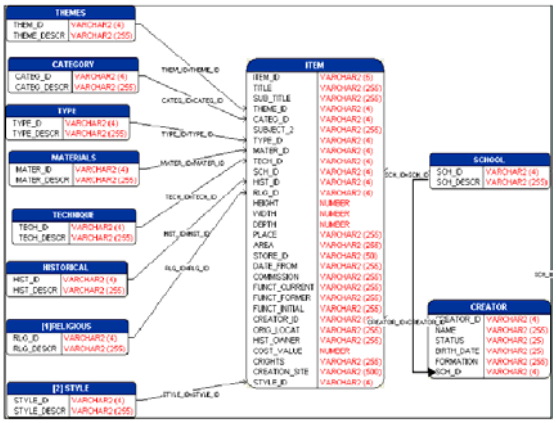

Figure 4 View of the relation database used in art-e-fact.

The thematic entities defined in the conception of the database are the following ones:

- Identification of the artwork, including general historical information about the artwork as subject, title, category, type, dimensions, etc.
- Description containing details about the item in order to better understand the context in which the artwork was created.
- Aesthetic Appearance, which concerns basically with plastic elements that provide the appreciation of the style and appearance of the artwork, manner, composition, colours, drawing style and so on.
- Tehcnical Issues, related to both techniques and materials used in the creation of the artwork such as support, preparatory layers of painting, underdrowings, painting materials, varnishes, stratigraphy, conservation treatments and diagnosis history.
- Interpretation, which compares the artwork with analogous or totally unlike artworks, such as thematic relationships, persons, symbols, styles or techniques.

3.3. The mapping

The most obvious application of mapping is to relate the ontologies to the actual contents of an information source. The art-E-fact ontology is related to the cultural content provided by the cultural participants within the art-E-fact project. Among the different approaches used to establish the connection between ontologies and information databases, we have chosen a definition of terms. In order to make the semantics of terms in a database schema clear, it is not enough to produce a copy of the schema. The definitions of the ontology do not correspond to the structure of the database. These are only linked to the information by the term that is defined.

The ontology is not a "mirror" of the structure of the database. Moreover, it should act as a "semantic index" representing the data and information stored in the database. Therefore, a class of the ontology can not be identified with a table in the database. The mapping has been carried out at the property level, so that a property in the ontology represents a field on a table of the database.

To map the database through the ontology, properties can be divided into two different groups:

- Properties made up by "Class" and "Instance". As this group is used to define relations between concepts, the associated properties cannot be used for the mapping, since they only show information about the structure of the ontology; and
- Properties made up by "Literals" (e.g. string, boolean, float, etc.). This second group can store information about the information stored in the database (e.g. the name of a person), so it is used to take part on the mapping.

The properties of each Class in the ontology have to be matched with a field on the database. As the aim of this mapping is to facilitate the access to the content of the database, we can have two properties of a Class pointing at the same field, but not vice versa, since it could cause some conflicts when trying to retrieve data from the database.

To do so, we have designed a "middle" ontology that records the information of the structure of the ontology, the database and the relations between them.

The mapping tool has been implemented as a part of the art-E-fact project. The problem is that most of the cases the artist is not an expert him/herself on the topic, since the profile of the user is more of a story-teller or an artist rather than that of an expert in History of Art.

The artists could just explore and get the information of the database with an interface, or even directly. Since this is a very complex and tough task, the mapping tool provides an easier way of accessing to the mysterious internal world of artists.

4. CONCLUSIONS AND FURTHER WORK

The art-E-fact project aims at developing a generic platform for interactive storytelling in Mixed Reality that allows artists to create artistic expressions in an original way. This platform should facilitate the access to knowledge bases which content will be used as an inspirational material for them.

The domain ontology that has been implemented within the art-E-fact project gives artists a general overview of the content stored in the database provided by the other members of the consortium without having to deal with a rough structure of a database.

The methodology for the mapping engine has been developed in a generic way, as well as both the database and the ontology. The mapping engine constitutes an easy way of accessing the interchanged content without the need of any other interfaces nor software, just through the knowledge representation (i.e. ontology).

As further improvement of the system, we are developing a multiple constraint mapping methodology. This means that we are having more than one only DB Entry Point. This is rather more complex than the single constraint method, but it will fasten the speed at which the access to the information on the database is achieved.

5. REFERENCES

[1] H. Wache, T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann and S. Hübner, “Ontology-based Integration of Information- A Survey of Existing Approaches”, *Proceedings of IJCAI 2001 Workshop 'Ontologies and Information Sharing'*, 2001.

[2] I. Iurgel, “From another point of view: art-E-fact”, *Proc. TIDSE’04*, Vol. 1, pp. 26-35, 2004.

[3] H. Stuckenschmidt, F. van Harmelen, A. de Waard, T. Scerri, R. Bhogal, J. van Buel, I. Crowlesmith, Ch. Fluit, A. Kampman, J. Broekstra, E. van Mulligen, “Exploring Large Document Repositories with RDF Technology: The DOPE project”, IEEE Intelligent Systems 19 (2004) 34-40.

[4] R. Studer, R. Benjamins, D. Fensel, “Knowledge Engineering”, DKE 25 (1998) 161-197.