Representation and management of ontologies in cultural heritage domain

Cristina Ghiselli[1], Loris Bozzato[1], Alberto Trombetta[1]

[1] Dipartimento di Informatica e Comunicazione,
Università degli Studi dell'Insubria,Via Mazzini 5,
21100 Varese, Italy
{cristina.ghiselli, alberto.trombetta}@uninsubria.it

Abstract. The aim of this work is to present a complex, web-based virtual museum application, integrating several tools for flexible management of heterogeneous and highly structured knowledge. All the used tools are compliant to W3C's standards. In particular, the complex network of associations and relations among concepts and objects (as typically found in a virtual museum environment) has been faithfully represented adopting W3C's Semantic Web standards. A formal representation of these concepts and relations to enable inference requires rigorous formalism as ontologies. Then, the proposed ontology allows for constraining, expressing and analyzing the intended meaning of the shared vocabulary of concepts and relations in the project domain of knowledge. As a valuable byproduct, this allows the virtual museum's visitor to interact by means of highly expressive queries.

1 Introduction

As the next step in the evolution of the World Wide Web, the Semantic Web supports the users (not only human readers, but software agents also) to find accurate information and combine easily related pieces of information in new objects [1]. To achieve these objectives the content of the Web has to be better structured and marked up to allow a semantic processing of information. The relevant information has to represent in a declarative and semantically way through the generation of ontology-based metadata [19]. An ontology is a formalism to define relationships among different concepts and objects used to describe and represent an area of knowledge. In particular, an ontology tries to capture and represent the meaning, or semantic, of a domain of knowledge, simulating explicitly the human mental model of that domain. Then, the expressiveness of the ontology representation can be used for enabling inferences through reasoning procedures [4].

While implementing the Semantic Web on the Internet is still a vision, the building blocks for the Semantic Web are being deployed in small domains and prototypes [2].

In this work we explore the possibilities of using Semantic Web tools for representing and querying the complex relationships occurring among data in a cultural heritage domain. These tools are used on the top of a complex, web-based virtual museum application which is part of cultural heritage project called Domus Naturae [17].

The project was promoted to study, classify and store information about natural species painted on the walls of the about 60 rooms of Palazzo Arese Borromeo in Cesano Maderno (a small town near Milan). As a traditional "virtual museum" the hypermedia system is composed of logically related collection of digital objects and provides various points of information access, giving the visitors a first dynamic and multidisciplinary approach to the collection [15]. In the next section the system architecture based on new technological standards of the WWW Consortium, like XML (Extensible Markup Language) and XSL (Extensible Stylesheet Language), is briefly described.

In order to transform the project into a Semantic Web prototype, we define an ontology describing relationship among artistic, botanic and zoological multimedia data by means of OWL (Ontology Web Language), part of the growing stack of W3C recommendations related to the Semantic Web [12]. In fact, the more perceived need to make machine-processable the wide information residing on the Web has acted as a strong impetus for the development of ontology languages. Representing in a web knowledge representation language like OWL, the ontology is backed by a rigorous formal logic, which makes the ontology semantically interpretable by the system without direct human involvement.

The most relevant feature of this semantic Web-based application is its capability of expressing powerful queries that are difficult or impossible to answer using SQL. The main reason for this augmented expressive power comes from the possibility offered by the deployed Semantic Web tools of directly representing ontologies as derived from the conceptual map describing the application's overall scenario and from the reasoning mechanisms provided by such tools [13].

As a negative remark, we experiment the extreme computational inefficiency of the available Semantic Web tools, even in the execution of the simplest queries. In order to alleviate the efficiency problem, we focus on the ontology development process phase, in which we aim to deploy a well methodology.

Nowadays there is not a well-established design methodology for this kind of data organization comparable to the ER methodology in the relational model setting.

As a starting point, we follow a top-down approach in order to create first the main classes, then their properties and then adding more details to the scheme adding restrictions.

Section 3 explains the applied web semantic techniques in more detail and then the features needed to support the additional scenario. In particular we present the ontology development process and introduce how this ontology allows the visitor to interact by means of highly expressive queries.

Another feature of the project is to support collaboration between users [1]. The system, indeed, provides a community space for field experts, which may help grow and/or amend the knowledge stored in the hypermedia system, e.g., by means of annotations. The basic idea is that the users/visitors can't only browse, but can also create shared information. In the section 4 is presented the problem of extending these annotation tools from simple instruments that enhance collaboration within a group of experts to complex instruments that permit authoring data by using ontology guidance. Finally Section 5 highlights current research and implementation directions.

2 Scenario: Domus Naturae project

Domus Naturae project is the result of a collaborative work between Computer Science and Communication Department and Biology Department of Insubria University of Varese (Italy), coordinated by Prof. Andrea Spiriti. The project is a team effort for integrating different disciplines, ranging from zoological, botanic to historical and artistic knowledge. In fact, it consists of a study about natural species painted on the walls of the rooms of the 17th century Palazzo Arese Borromeo.

In our approach, the project Domus-Naturae is considered as a laboratory in which new technological tools supporting cultural needs can be proposed and experimented. Consequently, the infrastructure must keep pace with the trend of technological progress in order to understand the usefulness and applicative, or methodological, innovation of the instruments proposed and to see whether they can be adapted and reused in different environments, or whether further developments can be proposed.

The role of technologists and designers is also to create a link in continuous tension between speculation and pragmatism: while considering cultural needs, they are also introducing the most abstract of ideas into the operating laboratory to conceive solutions based on available and experimental infrastructures, methodologies, and languages.

The core of the Domus Naturae project is a hypermedia system allowing the access to the application data and their scientific analysis through different keys of reading and communication, like guided walkthroughs or highly expressive queries over the multimedia data, as well.

The hypermedia is articulated on various levels of navigation in order to yield an enhanced museum experience: from a panoramic virtual visit through the floors and rooms to the dynamic visualization of particulars of a wall, to the analysis of a particular with the list of its occurrences in the rooms.

According to new technological standards of the WWW Consortium, the system is based on XML (Extensible Markup Language) for representing the structure of information and on XSL (Extensible StyleSheet Language) for its presentation. Graphic representations of dynamic maps are realized with SVG (Scalar Vector Graphics).

The architecture of the application is divided into two basic components. One component concerns the automatic and interactive generation of XML-based data with XMLSchema validation that represent the knowledge structures of the application, while the other component concerns the XSLT-based transformation of the data in the result document format, like XHTML or SVG (fig. 1).

In the application the user interaction with the classification of particulars painted on the wall is improved by the use of SVG as language for graphical interfaces. With SVG vector graphics, images and text are grouped, styled and transformed with filter effects and it is possible to perform advanced animation and dynamic graphics.

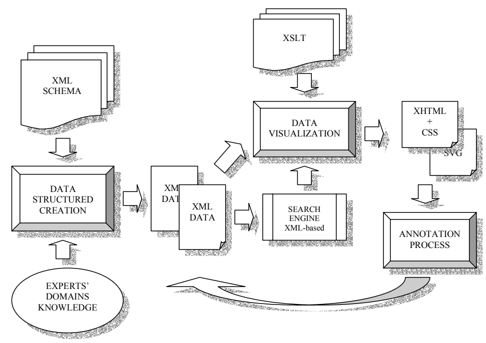

Fig. 1. System Architecture

The application is based on Apache Tomcat Web Server. It is part of an open-source project by Apache Software License that develops and distributes solutions for the Java platform. Tomcat is not only a common server that manages requests through the HTTP protocol, but is a servlet container used for implementation of JavaServlet and Java Server Pages technologies. The application also includes an open-source native XML database (eXist) allowing traditional interrogations and the annotation tools that is presented in Section 4.

At this moment the system prototype operates in pilot form within an Intranet.

3 Semantic Web techniques

Semantic Web tools are used on the top of the hypermedia system for representing and querying the complex relationships occurring among the above mentioned data.

In the Semantic Web scenario, as already said, relationships among different concepts and objects are represented by means of ontologies. In our work the domain of interest of the project, ranging from artistic to botanic and zoological data, is described as a formal, shared conceptualization through ontology. In particular, the ontology allows for constraining, expressing and analyzing the intended meaning of the shared vocabulary of concepts and relations in the cultural heritage environment.

In general, an ontology can range from the simple notion of a taxonomy (knowledge with minimal hierarchic or parent/child structure), to a thesaurus (words and synonyms), to a conceptual model (with more complex knowledge), to a logical theory (with very rich, complex, consistent, meaningful knowledge) [2]. The representation of the ontology should be used for intelligent reasoning and be computationally efficient. The language used for knowledge representation determines the kind of reasoning. Then language with limited expressivity cannot directly be used for automatic reasoning methods that require more complex expressiveness [11].

We have adopted OWL Lite as the language for the representation of ontology and we investigated the tools and technologies already available that can be used to create knowledge from existing and new project information. In particular we adopt Protégé-2000 as OWL Lite editor and the framework HP Jena for using and querying the developed ontology into the web application [16]. The queries are expressed through the (far less standard) ontology query language RDQL (RDF Data Query Language) [18].

However, we have tested the extreme computational inefficiency of the available Semantic Web tools, even in the execution of the simplest queries. In order to alleviate the efficiency problem, we focus on the ontology development process phase, in which we aim to deploy a methodology inspired by those employed in the database community.

3.1 Ontology development process

A very relevant issue regards what methodology to adopt when developing an ontology. There are many proposals and none of them has been widely adopted. As such, we chose to employ an "ad-hoc" methodology comparable to the well-known ER methodology in the relational model setting.

Such process has undergone through many steps. We started from a conceptual map and from data samples about the species represented on the paintings. Being defined by content experts, the conceptual map allows identifying easily the classes and subclasses, properties and relations between classes of the domain. For example, the animal taxonomy (of interest to our context) is modeled through a hierarchy of is-a relationships (fig. 2).

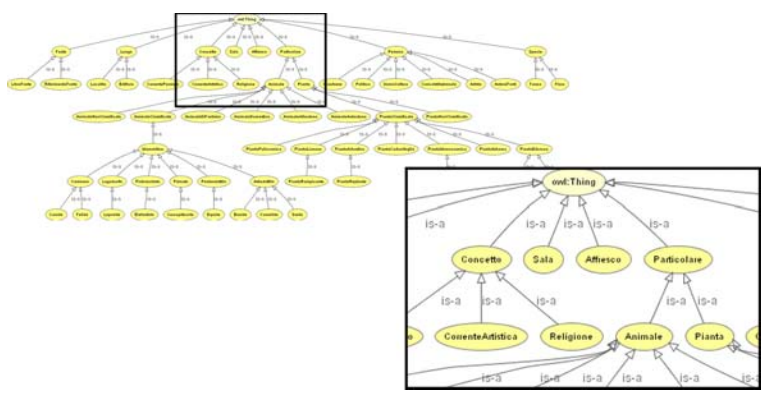

Fig. 2. Ontology schema describing the domain of the application.

From the conceptual map, we developed the specifications of the prototype. We followed a development based on the ER model development for relational databases.

So we developed a schema of the ontology and then converted it on the classes and properties taxonomy, adding also restrictions on classes and properties to better describe their features (fig. 3).

Once done this formalization of the ontology schema and having further refined it, we implemented the schema in the OWL ontology by means of the ontology editor Protégé-2000 [7].

Then we added part of the data defining instances (individuals) of the ontology classes and properties. The consistency checking has been performed using the RACER reasoner, as a plugin to Protégé.

We implemented the queries on the semantic search application (using the RDQL ontology query language) on the top of it.

Nowadays there is no well established ontology development methodology. As a simple solution, we followed a top-down approach in order to create first the main classes, then their properties and then adding more details to the scheme adding restrictions. The corresponding OWL file generated by Protégé has been included into the application and, through the HP Jena framework and the Tomcat server, the query interface has been defined. The graphical interface has been implemented through a Java applet.

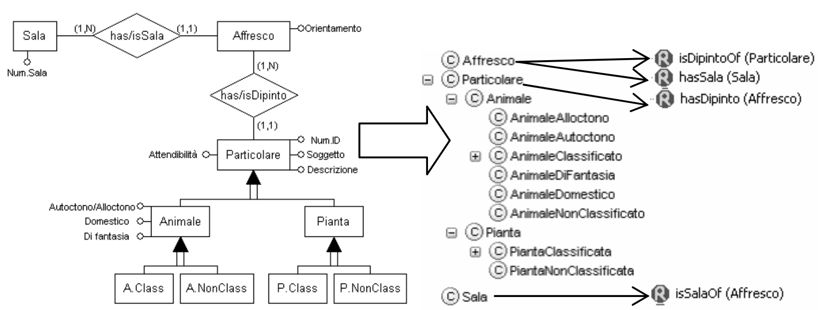

Fig. 3. Example of schema development process

3.2 Query processing

As a positive remark, deploying such technologies has allowed us to express far more expressive queries than those usually possible within an RDBMS-based web application.

As already said, the most relevant feature of our semantic Web-based application is its capability of expressing queries that are difficult or impossible to answer using SQL as query language. Such enhanced expressive power comes from the modeling capabilities offered by Protégé and from the reasoning capabilities offered by tools as RACER.

Now we present some examples of powerful queries based on the developed ontology. As examples of interesting queries based over information represented as RDF triples, we mention finding all the relations between two given persons or looking for all the relations of a person. In the following, we present a query RDQL requesting a relation (and not an object) related to a particular person.

```
SELECT ?relation, ?related

WHERE (?person dn:name_person "[nameperson]")

      (?relation rdfs:subPropertyOf ?super)

      (?person ?relazione ?related)

AND   (?super eq dn:isLinkOf || ?super eq dn:hasLink)
&& !((?relation eq dn:hasLink)||(?relation eq
dn:isLinkOf))

USING dn FOR
<http://domusnaturae.dicom.uninsubria.it/ontologia#>
```

The query about an instance of the class person retrieves all the connected individuals of the classes of the ontology (identified by the URI http://domusnaturae.dicom.uninsubria.it/ontologia# and the namespace dn). The query returns also the relations (or properties of the ontology) that are only subproperty of the general relations hasLink or isLinkOf. One or more of these relations can be the input for the following query retrieving only the related individuals connected to the input person by the relation.

```
SELECT ?related

WHERE (?person dn:name_person "[nameperson]")

      (?person dn:"[relazione]" ?related)

AND   USING dn FOR
<http://domusnaturae.dicom.uninsubria.it/ontologia#>
```

Figure 4 shows like this example is implemented in the web application. The web ontology-based interface enables to search and shows graphically the connections between the instances of the ontology, as also shown in figure 5. The search of the user's selected instances in the hypermedia system can be transferred to the search engine used in the application.

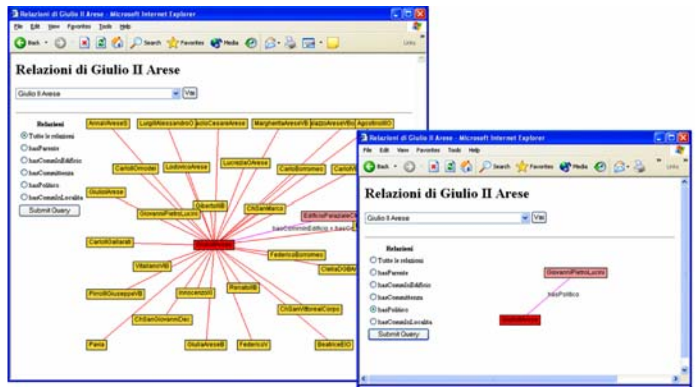

Fig. 4. Example of a search in the web interface. First, the query “Giulio Arese II” retrieves all instances (like persons, place, buildings, …) that are connected (by relations like kinship, political partnership, artistic partnership, cultural interests, ..). Then, for example, it is possible to search and visualize only the instances that have political partnership.

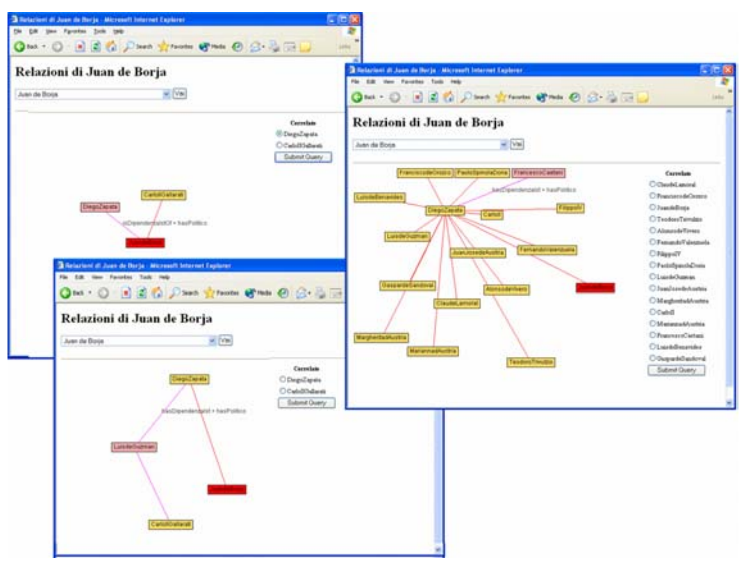

Fig. 5. Example of query interface for visualization of the user’s walkthroughs over the ontology. The first query retrieves all instances connected by relations. Then, the user can select one of these instances and repeat again the search keeping the previous connections. Then the user obtains the graphical network of individuals’ connections.

In general, the queries that allow searching sub-relations or subsumption between classes are based on the ontology structure. Moreover it is more interesting to make queries based on the reasoning techniques. For example, the query for searching all the relatives of a person, uses the implicit symmetric and transitive relation defined for kinship.

To further underline the difference from “usual” SQL expressions, a query asking (for example) for all the places in which an artist has worked and vice versa all the artists who have worked in a given place are expressible in SQL at the cost of explicitly defining complex joins. On the other side, using an ontology-based approach, such queries are answered directly by the system, using its information representation and reasoning capabilities.

4 Annotation problem

The proposed web ontology-based application represents a first step towards a more complete formalization of the relationship occurring among cultural heritage objects in a virtual museum. As we said, the effective search of the information about the user’s selected ontology instances in the hypermedia system can be transferred to the search engine utilized in the application.

The information could be handled with meta-information about ontology source and about user assertions. Additional types of information may be required if users need to understand the meaning of terms or implications of query answers. [6][9]

Then the application requires tools that allow introducing new information. In order to support collaboration between users the system provides a community space for field experts, which may help grow and/or amend the knowledge stored in the hypermedia system, e.g., by means of manual annotations. The basic idea is that the users/visitors can't only browse, but can also create shared information.

At this moment the developed tools are examples of instruments for textual annotation. In annotation process notes and commentaries are added to resources and the characteristic of this approach is that it is aimed at human readers. In the application, for example, this process is useful in cases of uncertain interpretations of natural species painted on the walls of the palace. As registered experts visit the web application and find unclear classifications, they can use annotations to share general comments each other or point out interesting perspectives. Not registered users can only see the presence of annotations and read them.

It is possible to extend these annotation tools from simple instruments that enhance collaboration within a group of experts to complex instruments that permit authoring data by using ontology guidance [8].

A solution is to perform semantic annotations of the documents or fragments of documents selecting the classes and sub-classes or instances in an ontology-based graphical interface. As a result, a list of semantic annotations is generated for the respective document. Then, it is possible define semantic queries via the construction of a combination of pattern classes, restrictions, and relations in order to get the relevant fragments or documents that match the query definition, basing on the adopted application search engine.

5 Conclusion

This work is an investigation of the tools and technologies already available also for the web that can be used to manipulate ontologies and gives some indication on how a semantic web for virtual museums might work. In fact, the work shows that ontologies and open hypermedia architecture can work together to support a better searching and retrieval phase. In order to get over the extreme inefficiency of the available Semantic Web tools, we are working to standardize an ontology design methodology for various kind of data organization. Another task is to define a framework more flexible enough to allow user annotation tool to be used for increasing the ontology dynamically.

Finally, this experimental prototype of the multimedia system, developed and maintained by Computer Science and Communication Department of Insubria University of Varese, is useful to prove how the synergy of new technologies, communication and specific disciplines can activate processes of knowledge, exploiting the interdisciplinary nature and the potentials for interaction, which is a distictive mark of the Web.

Acknowledgment

The authors are grateful to Prof. Andrea Spiriti for his assistence in providing the conceptual map. We gratefully acknowledge the work of the Domus Naturae Project team of Computer Science and Communication Department and Biology Department of Insubria University of Varese (Italy).

References

1. Berners-Lee, T.: Weaving the Web. Harper San Francisco, San Francisco (2002)
2. Daconta, M. C., Obrst, L.J., Smith, K.T.: The Semantic Web: a guide to the future of XML, Web Services, and Knowledge Management. Indianapolis, Wiley Publishing (2003).
3. Fikes, R., Hayes, P., Horrocks, I.: OWL-QL. A Language for Deductive Query Answering on the Semantic Web, Stanford Knowledge Systems Laboratory Technical Report KSL-03-14, Stanford, CA (2003).
4. Guarino, N., Giaretta, P.: Ontologies and Knowledge Bases: Towards a Terminological Clarification. In: Mars., N., Towards very large knowledge bases: knowledge building and knowledge sharing, IOS Press, Amsterdam (1995), 25-32
5. Gruber, T. R.: A translation approach to portable ontology specifications. In Knowledge Acquisition n. 5 (1993) 199-220
6. Handschuh, S., Staab, S.: Annotation of the Shallow and the Deep Web. In: Handschuh, S., Staab, S. (eds.) Annotation for the Semantic Web, IOS Press (2003), 25-45
7. Horridge, M.: A Practical Guide To Building OWL Ontologies With The Protégé-OWL Plugin. Edition 1.0. http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf. University of Manchester, Manchester (2004).

8. Kiryakov, A., Popov, B., Terziev, I., Manov, D., Ognyanoff, D.: Semantic annotation, indexing, and retrieval, Web Semantics: Science, Services and Agents on the World Wide Web 2 (2004) 49–79
9. Koivunen, M., Swick, R.: Collaboration through Annotations in the Semantic Web. In: Handschuh, S., Staab, S. (eds.) Annotation for the Semantic Web, IOS Press (2003), 46-60
10. McGuinness, D., Pinheiro da Silva, P.: Explaining answers from the Semantic Web: the Inference Web approach. In Web Semantics: Science, Services and Agents on the World Wide Web 1 (2004) 397–413
11. McGuinness , D.: Ontologies Come of Age. In: Fensel, D., Hendler, J., Lieberman, H. and Wahlster, W. (eds.): Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press (2003)
12. McGuinness, D., van Harmelen, F.: OWL Web Ontology Language Overview, W3C Recommendation del 10 febbraio 2004. last updated: 10-02-2004. consulted May 2005 http://www.w3.org/TR/2004/REC-owl-features-20040210/
13. Patel-Schneider, F., Fensel, D.: Layering the Semantic Web: Problems and Directions. In: Proceedings of the First International Semantic Web Conference, Lecture Notes In Computer Science vol. 2342, Springer-Verlag, Heidelberg (2002)
14. Schuler, D.: Community Networks: Building a New Participatory Medium, Communications of the ACM, 37 (1) (1993) 39-51.
15. Schweibenz, W.: The "Virtual Museum": New Perspectives For Museums to Present Objects and Information Using the Internet as a Knowledge Base and Communication System. In Zimmermann, Harald H./Schramm, Volker (eds): Knowledge Management und Kommunikationssysteme. Workflow Management, Multimedia, Knowledge Transfer. Proceedings des 6. Internationalen Symposiums für Informationswissenschaft (ISI '98) Prag, 3.-7. November 1998. UKV. S., Konstanz (1998)185-200.
16. Seaborne, A.: RDQL. A Query Language for RDF, W3C Member Submission del 9 january 2004. last updated: 09-01-2004. consulted May 2005 http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/
17. Spiriti, A.: Domus Naturae. Arte e scienza a Palazzo Arese Borromeo: la flora e la fauna della grande boscareccia, Comune di Cesano Maderno, Cesano Maderno (2003)
18. Stanford Medical Informatics (2004). The Protégé Ontology Editor and Knowledge Acquisition System. Version 29-12-2004. (http://protege.stanford.edu/).
19. Swartz, A.:The Semantic Web In Breadth. last updated: 01-05-2002. consulted May 2005 http://logicerror.com/semanticWeb-long. (2002)