Dr. Oreste Signore
CNR-ISTI
Via G. Moruzzi, 1
56124 Pisa
Italy
Phone: +39 050 315 2995 (office)
Oreste.Signore@isti.cnr.it
www.weblab.isti.cnr.it/people/oreste/

Ontology Driven Access to Museum Information

Summary
Cultural heritage is rich in associations. Much emphasis is given to retrieval, while users tend to browse by association. If data are semantically annotated, an appropriate intelligent user agent aware of the mental model and interests of the user can support her/him in finding the desired information. The whole process must be supported by an ontology.

Introduction

Cultural heritage is very rich in variety of possible associations, either between documents themselves either with documents pertaining to other disciplines. Documents concerning history, economy, religion, ethnology, can easily contain information relevant for a scholar interested in archaeology, art history, architecture or any other specific field.

A lot of information is available on the web, and users need to access the complete universe of information, looking for any fragment of data that may be of interest.

Two of the long-term targets for the web are the *semantic web* and the *universal access*. As *semantic web*, we intend a document space where information is machine processable, so implementing a true "*universal information space*". This means to develop a software environment that permits each user to make the best use of the resources available on the Web. *Universal access* aims to make the Web accessible to all by promoting technologies that take into account the vast differences in culture, education, ability, material resources, and physical limitations of users on all continents. It is easily seen that the second goal is somehow making the first even more difficult. Processing information when we have to cope with different cultures or mental habits is really a big challenge.

In this paper we will describe a general, flexible architecture relying on web standards where intelligent agents help users in finding the appropriate information, making use of ontologies.

Accessing information on the Web
For effective access to information on the web, we must consider several issues: the importance of link mechanism and challenges in information integration.

Searching vs. Linking - In accessing information, much emphasis is upon query and retrieval. However, much information is conveyed by the links connecting different pieces of information. Links are the essence of the hypertext, but they are meaningful only if their semantics is clear, and users can perceive it. Most of the value in browsing the web comes from following associations coherent with user's scientific interests. Really, it is easy to realize as important are the links, just considering the typical "*search and link*" approach followed by the majority of users. In the typical usage scenario, the user starts with a fairly general query, and the search engine often returns a huge amount of records. The user behaviour is then to look to the most "promising" records. Once an interesting one has been found, the user tends to follow the links, so making use of the knowledge embedded in the document itself, as typically links are inserted by the designer to point to relevant connected information. Obviously, this brings up some old and very well known problems, inherent to the hypertext approach itself: the "*lost in the hyperspace"* syndrome and the *cognitive overhead*. It follows that the real need is for *adaptive* and *intelligent* systems, which should take care that navigation is consistent with the real user interests, that could be formalized by a representation of the user's mental model.

According to this approach, we can state that querying remains a very important issue, but we have to find a trade-off between effectiveness of the search (i.e. Precision and Recall) and complexity of query formulation, efficiency, etc.

We will show in the following that a possible way to support the user's mental model without imposing constraints on the data themselves is to implement navigation mechanisms based on a core ontology.

Information Integration - Information integration via a *common schema* appears in principle the simplest way, but experience shows that this approach will almost invariably fail. The main reason is that different schema exists as the heritage of well established cultural traditions and is very unlikely that one of them will accept to conform to the other. As a consequence, it is difficult, if not impossible at all, to agree on a single schema as a way to achieve effective querying.

Integration is often attempted at *metadata* level. In this approach, information is enriched by metadata, which permits to have a common reference schema. A typical example is the Dublin Core initiative. This approach has been adopted in many projects.

However, as also noticed in [Doerr2003], "the number of metadata vocabularies will continue to grow as individual communities seek to structure their own information for their own purposes", and "attempts to develop universal metadata vocabularies are misdirected, since "spoken" languages (those used by communities to actively describe content) will inevitably diverge (history is replete with failures to find common spoken languages [Eco1997])".
In addition, in our opinion, metadata by themselves cannot exploit the full richness of possible associations among different information items. The association mechanism remains in the mind of the user.

Semantics of links and documents
Especially following the wide adoption of XML technologies, documents on the Web are often deeply structured, and this can be the origin of incompatibility between different views of the same matters. As a consequence, it can be useful to have documents where some elements can be seen as "semantic items", useful to identify concepts that characterize the specific part of the document. Links have also their own semantics. This aspect has been often neglected, even if it was present since the inception of the Web, as we can easily see from the original proposal by Tim Bernes-Lee. Taking into account both the documents and links semantics can lead to a terrific enhancement of navigation possibilities, which will really support the association model that is the basis of the hypertext, and allow personalizing the presentation of documents, needed for a real adaptive hypertext environment.

The association model. When reading a book or a newspaper, our attention is often captured by some words (*anchor*s) as items leading our mind to other documents. In the Web context, documents, whatever will be their genesis, are seen as resources. We can model the association process in the following way:

the anchor leads to a concept
the concept is related to other concepts
the new concept is related to some resources.

This basic association mechanism is totally independent from document structuring. In the data space, documents are connected by *extensional* links. In the Semantic Web architecture's ontological level, associations among concepts implement *intensional* links among documents.
Now, two questions arise:

how can we implement the link from resources to concepts;
how concepts are linked together.

A simple and effective way to implement intensional links is to identify the *semantic items*. This can help in several cases (for example, "The French emperor" is an implicit reference to "Napoleon"). We can also characterize each semantic item with a specific semantic category (e.g. person, location, date, taxonomy) useful to tailor the document appearance to the specific user interests. A reader interested in space-time associations will get location and date items emphasized.

The second question directly leads to the *interaction metaphor* issue. Apart the case of taxonomic classifications, where we can make use of the well-known thesaurus techniques, very powerful association mechanisms are space and time. For example, a semantic item can have a spatial valence, then, using an interaction metaphor based upon space, we can both jump on other resources linked to the same place, or select a different place, and then find other resources related to this different place. This simple hyperlink association model can be implemented through a document, link and user model.

Document semantics. In XML documents we must clearly distinguish between *structural* and *semantic* information, which can be associated to elements or part of them. Documents on the web have different structures, which a wide variety of users should be able to share and understand. A way out is to *semantically annotate* both various parts of the documents and links. We can also specify a *weight*, stating the relevance of the concept in the document context.

Link semantics. Semantic qualification of explicit (or extensional) links identifies their meaning in the document and the role of involved resources. However, and probably a more important issue, two documents can be linked through an *intensional* link existing at the ontological level, even in absence of any extensional link specified in the document.

User Model. As a first approximation level, user mental model should be tightly related to the semantic model of documents and links.

Role of ontologies

Information integration – As it has been pointed out before, it is unlikely that information integration can be reached just converging on a single set of metadata, while a more useful effort is to attempt to formulate a language as a base for "understanding". This is what we can define to be a "core ontology" which incorporates basic entities and relationships common across the diverse metadata vocabularies.

Such a core ontology might then be useful for integrating information from heterogeneous vocabularies and uniform processing across heterogeneous information sources.

There is an important, even if subtle, difference between a core ontology and *core metadata*, such as Dublin Core. Even if both are intended for information integration, they differ in the relative importance of human understandability. Metadata is in general created, edited, and viewed by humans. Therefore, human factors, including limits on complexity, should play a primary role in its design. In contrast, a core ontology is an underlying formal model for tools that integrate source data and perform a variety of extended functions. As such, higher levels of complexity are tolerable and the design should be motivated more by completeness and logical correctness than human comprehension.

If must be stressed that ontology based information integration can be performed automatically by software agents.

Deriving knowledge – A *core ontology* is one of the building blocks to information integration. The goal of a core ontology is to provide both a global and extensible model into which data originating from distinct sources can be mapped and integrated, and base concepts that future metadata initiatives could build on when developing domain specific vocabularies. The canonical form of the model can then provide a single knowledge base for cross-domain tools and services (e.g., resource discovery, browsing, and data mining). A single model avoids the inevitable combinatorial explosion and application complexities that results from pairwise mappings between individual metadata formats and/or ontologies.

In accessing information ([Goble2001]) shared vocabularies give a little help in inferring new, previously undisclosed information about resources. Vocabularies based on ontologies that organize the terms in form that has a clear and explicit semantics can be reasoned over. For example, a metadata annotation about a page can be used to search for a resource related to a more general or specific concept, or having some relationship with the current one. This process is fundamental in enriching knowledge.

CIDOC-CRM – CIDOC Conceptual Reference Model represents an ontology for cultural heritage information as it describes, in a formal language, the explicit and implicit concepts and relations underlying the documentation structures used for cultural heritage. The primary role of the CRM is to serve as the semantic 'glue' needed to transform disparate, localized information sources into a coherent and valuable global resource. It is a conceptual model that can be used as a global schema in applications and for query mediation to heterogeneous sources, as well as a set of concepts to create common tagging schemes.

The CIDOC CRM is specifically intended to cover contextual information: the historical, geographical and theoretical background in which individual items are placed and which gives them much of their significance and value.
As a formal ontology, it can be used to perform reasoning (e.g. spatial, temporal).

Objectives
The main idea is to have an architecture where intelligent user agents can have access to the mental model expressing the interests of the user. The content can be tagged and semantically annotated using classes and properties defined in CIDOC-CRM. The agent can then perform reasoning,

linking the information the user is interested to, following the relevant associations.

Semantic annotation of documents – Documents can be annotated using a formal ontology, like CIDOC-CRM. The main advantage is in having a common frame of reference for all organizations, as a result of several years of effort by many scholars. In the peer-to-peer architecture that is the basis of the web, basic Semantic Web technologies allow semantic markup of content, in a fully decentralized way, without affecting existing data. In fact, a relevant issue is that the annotation can be done in RDF, and can reside on any place in the Web. This implies that semantic annotation must not necessarily be done by the owner of data, but any scholar could, in principle, co-operate in enriching the semantic of documents (quite obviously, security issues must be considered, and consistency with the role played by digital signature in the Semantic Web architecture is needed). Knowledge expressed by the markup can be used by intelligent software agents to make the best use of data, and perform reasoning, making use of an appropriate ontology.

The user mental model. – The user mental model can be expressed in terms of preferred interaction metaphors. Making reference to the ontology used as basis for semantic annotation, this means to specify the set of classes and properties the user can be interested in navigate.
A user interested in the *temporalContext* will be interested in classes like:
    E2 Temporal Entity
    E52 Time-span
and their subclasses, at various levels, like E3 Condition State, E4 Period., E5 Event.

The context can be expressed in a more precise way stating the properties the user is interested to navigate (e.g. P117 *occurs during*, P118 *overlaps in time with*, etc.) to build up the temporal interaction metaphor.

Identifying the properties the user is interested in can guide the agent to select the appropriate associations and perform the reasoning.

The architecture - The user agent (the **browser**) is enriched by two components: a **reasoner** and a **finder**, which accomplish the tasks of getting the semantic annotation of the current resource, looking to the user model, finding correspondences between user model and resource metadata, initiating a search following the properties the user is interested in.
The process can be summarized as follows:

1. **user** searches for something
2. the **browser** displays the list of returned records
3. the **user** selects one of them
4. the **browser** displays the **currentResource**
5. the **reasoner** parses the **userModel** and the **currentResource**, looking for matching of classes and properties the user is interested in

6. the **reasoner** looks at the **ontology** to find the kind of information it has to search for
7. the request is passed to the **finder**, which will query the Web
8. returned resources are passed by the finder to the **reasoner**, for further checking
9. the **reasoner** passes returned resources to the **browser**, which will display them as possible linked resources, selecting the appropriate interaction metaphor or suggesting a choice among several of them.
10.

A sample reasoning process – Suppose that the resource is describing a painting done in *1530*, describing an event pertaining to the *history of Christ*, by a painter of *Sicilian school*, and the user is interested in the *temporal* context, the **reasoner** can follow the properties relating year 1530 to:

historical or artistic period

events occurred in a suitable time interval around 1530.

To do this, it must ask the Finder to access reliable sources describing historical events and artistic schools.
Would the user be interested in the iconographical perspective, the reasoner should look at an iconographic autority (e.g. Iconclass) and afterwards the Finder will search for work of art describing the same subject, or perhaps, with lesser relevance, other works of art related to a similar subject.

The extent of the search will depend on the user preferences, as there can be a limit to the number of returned resources, at least at the first attempt.

Conclusion
Relying on a core ontology, web browsing can take advantage of the basic semantic web technologies (RDF, OWL) appropriately linking information according to the user preferred interaction metaphors, associating information on the basis of spatial, temporal, classification affinity, so greatly improving the access to the information and knowledge stored in museums.

The proposed framework intends to use metadata annotations to build and construct complex hypertextual associations. In this sense, this approach differs from the usual one, where metadata are used by an agent (a person or a machine) to perform an effective query, having back a list of returned records. In our approach metadata can be used not only to describe as to link to the resource, but also to indicate *where* and *why* you can go from the resource itself. In the whole architecture, a significant role is played by the search mechanism, as after all, the effective implementation of the interaction metaphor requires finding relevant resources.

References
[CIDOC] *The CIDOC Conceptual Reference Model*, http://cidoc.ics.forth.gr/

[Crampes2000] Michel Crampes and Sylvie Ranwez: *Ontology-Supported and Ontology-Driven Conceptual Navigation on the World Wide Web*, Hypertext 2000, San Antonio, TX

[DC] *Dublin Core Metadata Initiative*, http://www.dublincore.org/

[Digicult2003] *DigiCULT - Towards a Semantic Web for Heritage Resources* - Thematic Issue 3, May 2003, http://www.digicult.info/downloads/ti3_high.pdf

[Doerr2003] Martin Doerr, Jane Hunter and Carl Lagoze: *Towards a Core Ontology for Information Integration*, Journal of Digital Information, Volume 4 Issue 1, Article No. 169, 2003-04-09, (April 2003), http://jodi.ecs.soton.ac.uk/Articles/v04/i01/Doerr/

[Eco1997] U. Eco, *The search for the perfect language*. Oxford, UK ; Cambridge, Mass.:
Blackwell, 1997.

[Goble2001] Carole Goble, Sean Bechhofer, Les Carr, David De Roure, Wendy Hall: *Conceptual Open Hypermedia = The Semantic Web?*, Semantic Web Workshop 2001 Hongkong, China

[Martelli2002] S Martelli, O Signore: *Semantic Characterisation of Links and Documents*, ERCIM News, N. 51, Special: Semantic Web, October, 2002, http://www.ercim.org/publication/Ercim_News/enw51/signore.html

[OWL] *Web Ontology Language (OWL)*, http://www.w3.org/2004/OWL/

[RDF] *Resource Description Framework (RDF)*, http://www.w3.org/RDF/

[RDFMSS] O.Lassila, R.Swick: *Resource Description Framework (RDF) Model and Syntax Specification*,W3C Recommendation 22 February 1999, http://www.w3.org/TR/REC-rdf-syntax

[RDFSS] *Resource Description Framework (RDF) Schema Specification*, W3C Recommendation 03 March 1999, http://www.w3.org/TR/1999/PR-rdf-schema-19990303

[SemWeb] <http://www.semanticweb.org/>

[Signore1995a] Signore O.: *Issues on Hypertext Design*, DEXA'95 - Database and Expert Systems Application, Proceedings of the International Conference in London, United Kingdom 4-8 September 1995, Lecture Notes in Computer Science, N. 978, Springer Verlag , ISBN 3-540-60303-4, pp. 283-292

[Signore1995b] Signore O.: *Modelling Links in Hypertext/Hypermedia*, in Multimedia Computing and Museums, Selected papers from the Third International Conference on Hypermedia and Interactivity in Museums (ICHIM'95 - MCN'95), October 9-13, San Diego, California (USA), ISBN 1-88-5626-11-8, pp. 198-216

[Signore1996] O.Signore: *Exploiting Navigation Capabilities in Hypertext/Hypermedia*. In Proceedings of HICSS '96. Maui Hawaii, 3-6 January 1996.

[Signore1997a] O.Signore, R.Bartoli, G.Fresta: *Tailoring Web pages to Users' Needs,* In Proceedings UM97. Workshop on Adaptive System and User Modeling on the World Wide Web. Chia Laguna, Sardinia, ITALY, 2-5 June 1997, pp 85-90.

[Signore1997b] Signore O., Bartoli R., Fresta G., Loffredo M.: *Implementing the Cognitive Layer of a Hypermedia*, Museum Interactive Multimedia 1997: Cultural Heritage Systems Design and Interfaces – Selected papers from ICHIM 97 the Fourth International Conference on Hypermedia and InterActivity in Museums, Paris, France, 3-5 September, 1997, (Edited by David Bearman and Jennifer Trant) Archives & Museum Informatics (1997), p. 15-22, ISBN 1-885626-14-2

[Signore2001] O. Signore, *Culture across Cultures: a Quality Challenge*, Experts Meeting on European Cultural Heritage on the Web, Position paper, 17 luglio 2001, <http://www.w3c.it/papers/cultureAcrossCultures.pdf>

[Signore2002] O. Signore: *A Simple Architecture for Adaptive Hypertext*, Adaptive Hypermedia and Adaptive Web-Based Systems, Proceedings of Second International Conference, AH2002, Malaga, Spain, May 2002, Lecture Notes in Computer Science, n. 2347, Springer Verlag, ISBN3-540-43737-1, pp. 568-571

[Signore2003] Signore, Oreste: *Strutturare la conoscenza: XML, RDF, Semantic Web* - Clinical Knowledge 2003 (1st edition) - Udine, 20-21 September 2003 http://www.w3c.it/papers/ck2003.pdf, http://www.w3c.it/talks/ck2003/

[Signore2004] Signore, Oreste: *Representing Knowledge in Semantic Cultural Web* - EVA 2004 Jerusalem Conference on the Digitisation of Cultural Heritage - Jerusalem, 11-12 October 2004, http://www.w3c.it/talks/eva2004Jerusalem/

[Stevenson2003] A. Stevenson, M. Addis , M. Boniface, S. Goodall, P. Grimwood, S. Kim, P. Lewis, K. Martinez: *Semantic Web techniques for multimedia museum information handling*, http://eprints.ecs.soton.ac.uk/8920/01/SCULPTEUR_CIDOC_2003.pdf

[TBL1998] Tim Berners-Lee: *Semantic Web Road Map*, (1998), http://www.w3.org/DesignIssues/Semantic.html

[TBL1999] Tim Berners-Lee: *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*, HarperSanFrancisco (1999), ISBN 0-06-251587-X

[TBL2001] T. Berners-Lee, J. Hendler, O. Lassila: *The Semantic Web*, Scientific American, May 2001, <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>

[WorkshopSW]] Progetto Minerva: *1° Workshop: Rappresentazione della conoscenza nel semantic web culturale*, Roma, 6 luglio 2004, http://www.w3c.it/events/minerva20040706/