A Framework for Ontological Description of Archaeological Scientific Publications

Andrea Bonomi, Glauco Mantegari and Giuseppe Vizzari
Department of Informatics, Systems and Communication
University of Milan–Bicocca

SWAP 2006 - Pisa - 20 December 2006
Outline

• ArcheoServer Project
• Digital Library
• Related Systems
• Application Scenario
• Content Description Approach
• Domain Ontology
• User Interface
• System Architecture
• Conclusion
ArcheoServer Project

• In 2005, the chair of Prehistory and Protohistory - University of Milan, the Department of Informatics, Systems and Communication - University of Milan-Bicocca and the Department of Archaeology - University of Bologna, have started a long-term project for the creation of a set of Web-oriented services aimed at supporting the sharing of knowledge on prehistory and protohistory in Italy.

• The main objective is the creation of a Web portal which will provide a collaborative platform for the exchange of scientific information among the communities of Italian archaeology researchers.
ArcheoServer Portal

• The project aims to realize a Web Portal with a set of advanced on-line services:

• Collaborative editing platform (wiki) for Archaeology researcher and students

• Web GIS of prehistorical sites in Italy

• Online database of ceramic type collections

• Digital Library helping users to finding information about scientific publications
Digital library: collections of bibliographical information + effective mechanism for the retrieval

The main requirement is to give the community itself the possibility of autonomously managing the contents by means of simple editing tools.

Bibliographic data will be entered manually by the students, while archaeology professors and researchers will supervise the work.

In most cases, archaeologists have just low-level technical competence.
Related Systems - DBLP

- e-Library is mainly inspired by DBLP (Digital Bibliography & Library Project).

- DBLP is a Computer Science Digital Library developed by the University of Trier.

- It allows searching a huge collection of bibliographic information (~ 800,000 publications) with a Web Interface.

- DBLP was started in 1993 (several years before the birth of the Semantic Web initiative) and it does not provide any form of semantic description of the publications.

http://www.informatik.uni-trier.de/~ley/db/
CiteSeer uses a different approach to populate the library: it actively retrieves new documents and automatic tags and links metadata information inherent in an academic documents syntactic structure.

“Search by keyword” only. (so, if you search for the papers of the “SWAP” conference, it also returns the publications about "Memory SWAP", “XOR SWAP”, etc...)

CiteSeer offers many interesting features, but since it is not an open source product, we cannot use it for the e-Library.
Related Systems - Bibster

- Bibster is a Peer-to-Peer system for exchanging bibliographic data among researchers.
- It exploits ontologies in data storage, query formulation, query routing and answer presentation.
- Bibliographic entries are automatically aligned to two common ontologies:
  - **Semantic Web Research Community Ontology**: It describes different generic aspects of bibliographic data.
  - **ACM Topic Hierarchy**: It describes specific categories of Computer Science domain.

2 main problems:
- not flexible
- extremely generic

http://bibster.semanticweb.org/
Application Scenario

- In order to effectively describe contents beyond a keyword based approach, and to support effective forms of information retrieval, users must have available a domain ontology whose elements can be selected as indicators of the topics treated in the publications.

- The usage of Semantic Web technologies and ontologies provide an added value - in fact it is almost a strict requirement.

Use Case Example

Searching for: "Sicily, Bronze Age"

The system will retrieve the article “The early bronze age Village of Mursia” annotated as “Mursia”, “Early Bronze Age”.
Content Description Approach

- The ontology T-Box has to be designed by archaeology professors and researchers with the aid of knowledge engineers.
- The e-Library supports the maintenance on the domain ontology A-Box component.
- Publications description is done by Archaeology students associating A-Box elements to the publications.
- End user can view ontology and bibliographic data
The Domain Ontology

Ontology T-Box

- Historical period
- Analysis
- Typology
- Social organization
- Site type
- Structure type
- Geographic Place
- Material
- Activity
- Figure
- Place
- Item

Details of the relations of “Item type” class

- Item type
  - subClassOf Typology
  - analyze
  - hasItemType
  - hasSuperType
  - period

Details of the relations of “Sword” individual

- Sword
  - instanceOf Weapon
  - hasSuperType
  - isBuildOf Metal
  - instanceOf Material
Ontology Visualization

- Ontology visualization is essential for the user to explore the available information and it also helps non-expert users to refine their search requirements, should they start with no specific requirement in mind.

- As shown in previous slide, ontology visualization is not an easy task.

- Our approach is to use a tree navigation and to visualize the details of the selected item only.

- The navigation tree is a **hierarchy of classes, and individuals** connected by `subClassOf`, `instanceOf` and other domain-specific properties.
Editor Interface

A-Box Editor
- Using this editor, an ontology maintainer can create new individuals, remove individuals, edit the related properties
- The aim of the navigation tree is to explore the A-Box and select the individual to edit

Publication Description Interface
- It allows the content editors to associate an ontology-based description to the publications.
- The publications descriptions are statements (i.e., subject-predicate-object triples) that associate a topic to a publication
End-User Interface

Semantic Query Interface

• Currently the Semantic Query Interface only allows searching for papers characterized by a specific topic.

• The interface allows selecting the requested topics from the ontology tree.

• The current implementation retrieves only the publications that satisfy all the specified criteria.

• A future extension may relax this constraint especially with reference to the number of retrieved publications (adapting the query to the result).
Semantic Navigation Interface

- The Semantic Navigation Interface will support users in the e-Library navigation (entirety e-Library or query results).

- The system may even suggest relevant contents which are semantically related to the ones the user is actually viewing on the screen.

- This system will suggest publications considering multiple strategies for making recommendations (e.g. similar treated topics, recently visited document, user interest, access frequency)

It will be object of future work
System Architecture

- The framework was developed according to a three-tier architectural approach.

The presentation layer is a Web-based user-interface.

The aim of the persistence layer is to store the topics ontology, the publications descriptions and bibliographic data.

OWL DL

RDF
Conclusion and Future Developments

• Preliminary results of the tests campaign showed that the proposed ontology visualization is useful for the users as a guide to describe the contents of publications.

• It helps users with no knowledge about ontologies to understand the relationship between the different topics and between the topics and the publications.

• Moreover new required features were expressed after the tests:
  • in particular, the users required the possibility to choose the property on which each tree is built on
  • another required feature is the ability to sort the tree items according to a given property

• Future works will be focused on the development and test of the Semantic Query Interface and the Semantic Navigation Interface, which will support users in the e-Library search and navigation