WikiBridge: a Semantic Wiki for Archaeological Applications

Pascale Chevalier, Eric Leclercq, Arnaud Millereux, Christian Sapin, Marinette Savonnet

WikiBridge: a Semantic Wiki for Archaeological Applications

Chevalier, P.[12], Leclercq, E.[3], Millereux, A.[3], Sapin, C.[2], Savonnet, M.[3]
[1] University Blaise-Pascal, Clermont-Ferrand, France
[2] ARTeHIS Laboratory, UMR CNRS 5594, University of Burgundy, France
[3] Le2I Laboratory, UMR CNRS 5158, University of Burgundy, France
care@u-bourgogne.fr

This paper details the main concepts and the architecture of WikiBridge, a semantic wiki, developed for the project CARE (Corpus Architecturae Religiosae Europeae – IV-X saec.). The aim of the project CARE is the constitution of integrated corpus of the European Christian buildings dated from the 4th to the beginning of the 11th century. A semantic wiki, called WikiBridge, is developed in order to: 1) allow collaborative work of researchers involved in the project and 2) open the corpus to a large public. WikiBridge combines collaborative and traceability aspects of wiki with semantic consistency and query capabilities. Semantics is guaranteed by an application ontology based on CIDOC-CRM.

Keywords: Semantic Wiki, Archaeological Corpus.

1.Introduction

The aim of the international project CARE (Corpus Architecturae Religiosae Europeae – IV-X saec.) is the constitution of integrated corpus of the European Christian buildings dated from the 4th to the beginning of the 11th century. This corpus will greatly facilitate work of comparisons, exchanges and discussions with numerous foreign researchers and specialists. This wide-ranging European program was introduced by the IRCLAMA (International Research Center for Late Antiquity and Middle Ages) at the University of Zagreb (Croatia). Several countries among which Italy, Spain, Czech Republic, Slovakia, Poland and Croatia began 3 or 4 years ago to work on the preliminary documentations of this ambitious project; moreover, Hungary has just joined the Middle-European group; Benelux, United Kingdom, Ireland and Greece are interested, as well as Albania for a common volume with Kosovo and Montenegro; Switzerland and Germany already have published complete catalogs in recent years. The programs of corpus in every country progress on different pace and reveal disparities inherent to sources. For example, Mediterranean regions are often more concerned with the analysis of Early Christian monuments. The most advanced work concerns Northern Italy and Croatia who fall into this category of analysis.

The project has been launched in France on January 1st, 2008 after acceptance of the French National Agency for Research. Managed by Christian Sapin and Pascale Chevalier (UMR 5594-ARTeHIS of the CNRS, Dijon), the project will last 4 years (2008-2011). More than sixty researchers from about twenty universities, diverse research institutions and heritage management institutions are working on. Various categories of staffs are involved: field archaeologists, historians, art historians, draftsmen, topographers, etc. They are collecting and analysing data concerning approximately 2700 monuments. Each of the 22 French regions will form a task force before 2011, 9 of them are already active. With new studies and recent excavations relating to all the period, the French team defines protocols covering all the buildings included in the diverse chronological periods. The accent is placed on the 7th-8th centuries and the decades around the year 1000. The corpus of multimedia documents (including texts, maps, and photographies) concerning every known building will be gradually published in the form of classic books (one for each modern region).

A Web 2.0 application is developed in order to: 1) allow collaborative work of researchers involved in the project and 2) open the corpus to a large public with a little knowledge on the European religious culture. Data driven web application technologies can be used to generate dynamic web content by using databases, but have some major drawbacks in terms of collaboration and traceability. Our approach is based on a combination of collaborative and traceability aspects of wiki with semantic consistency and query capabilities that database can provide. This part of the project is developed by the CNRS LE2I laboratory (Dijon). Some geomaticians of the Social Sciences and Humanities Research Institute of Dijon will conduct specific spatial analysis by providing GIS tools from end of 2010.

The rest of the paper is organized as follow: Section 2 presents our motivation, Section 3 describes the state of art and Section 4 presents our architecture. Finally, Section 5 concludes the paper.

2.Motivation

The CARE corpus needs a spatial and temporal specific models as well as a representation of domain knowledge.

Ontologies and data models have similarities for knowledge representation. They both offer means of description based on concepts and relationships between these concepts. In both cases, knowledge represents a consensus among applications that cooperate. Knowledge representation in an information system may be considered as two dimensions introduced by Spear (SPEAR 2006):

- horizontal dimension (or relevance) aims at determining the extent of information that should be included in the representation of knowledge. For example, if we represent knowledge in the field of archeology, relevance is the choice whether to specify knowledge on liturgical installations, on construction techniques, on religious environment such as diocese (figure 1.a);

- vertical dimension (or granularity) aims at determining the degree refinement of knowledge representation. In the archaeological area, the granularity is the choice whether to include a building description from walls structure to decor elements, pavements, etc.

It is problematic to include in a single data model general description for elements and fine details for others, except to take the risk of building a data model difficult to read and to maintain. In contrast, a data model can use multiple sources of knowledge representation, and can therefore adjust the extent of knowledge it covers.

Furthermore, ontologies offer a great freedom in managing the granularity (vertical dimension) of knowledge representation, albeit they cover a limited area. In figure 1.b, each sub-tree from THING may be considered as an ontology of a particular domain. To cover a wider field of knowledge it should compulsorily consider the relationships between several sub-trees. Grenon et al. (GRENON et al. 2004) propose the definition of three kinds of relationships:

- intra-ontology: relationship with two concepts of same part of an ontology;

- trans-ontology: relationship with a concept of a sub-tree and a concept of another sub-tree. For instance, a building is consecrated to a saint, in DL we can write: `Building ∃isConsecrated.-Saint`

- meta-ontology: relationship with a concept of an ontology and another ontology (considered as a whole).

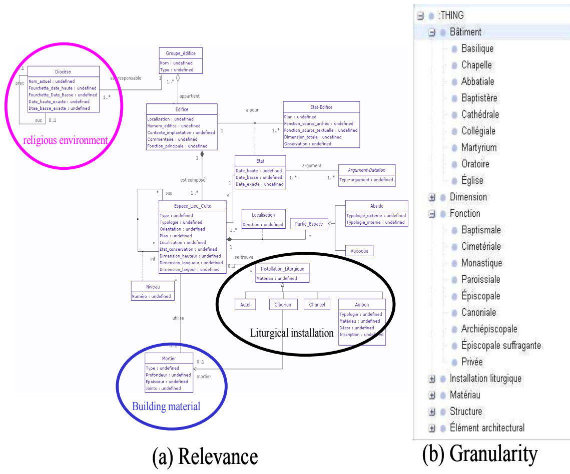

Figure 1: *Horizontal and vertical dimensions of knowledge representation.*

Moreover, if data models as well as ontologies allow representations based on concepts and relations, their organization in terms of relationships is different. Ontologies focus on specialization relations and strictly control the other relations being used. The data models in turn leave a great freedom in the choice of relations to use.

In short, data models and ontologies for knowledge representation can be combined to mix their specificities: the scope of coverage for the models, the granularity for ontologies. Ontologies are conceptualization of a domain, data models specify an implementation of structure and behaviour according to stated functionality requirements (SPYNS et al. 2002).

Data models are used to implement relational Data Base Management System (DBMS) such as Oracle or PostgreSQL. Semantic data consistency in DBMS is carried out by embedded controls such as triggers or stored procedures. Developing a typical stand alone application based on a database has two major drawbacks: 1) archaeologists have a purely document-based approach, away from the concepts of atomic decomposition and fully structured information imposed by database; 2) research conducted by archaeologists requires an open environment which allows to aggregate knowledge produced by different teams involved in the project. This environment must also provide sharing, exchange and evolution capabilities. The knowledge evolution leads to a dynamic evolution of database schema. It is difficult to subsequently modify a previously defined database schema and its content. A high degree of flexibility is then required. These two drawbacks make the construction of such an application at reasonable cost, difficult (BONOMI et al. 2008).

The request of a web application with a collaborative component led us to choose a solution based on a wiki. Despite the power of wiki (free input, rich user-interface, traceability, bi-directional links between pages, etc.), it is difficult to answer a specific query because of the purely textual information stored. One way to address this problem is to implement social tagging. Social tagging, a key characteristic of Web 2.0, allows users to index contents by their own keywords. Moreover, users do not understand an annotation schema since they decide themselves what keywords to use. The list of commonly used keywords are called a folksonomy. Nevertheless, keywords have no explicit links between them (hierarchy, similarity, synonym) and management of ambiguity and heterogeneity of keywords is not made. In our context, social tagging cannot provide enough semantic quality. Consequently, a semantic annotation approach of content based on ontologies is more relevant. The Semantic Web that can represent a complex knowledge is based on languages (RDF, OWL), tools, reasoners but requires knowledge experts. We generally consider that Semantic Web and social web are competitors. Some authors suggest to combine these two approaches (ANKOLEKAR et al. 2007). In addition, we believe that requirements for interoperability and data exchange (connection with other communities such as historians) must be taken into account since the design phase of the application. The Semantic Web thereby provides such kind of solutions by increasing the expressiveness of data representation, and by allowing reasoning tools and semantic search.

Wiki engines provide tools to manipulate document in a collaborative environment and ensure the cost of developing and maintaining reasonable. Our proposal is to use MediaWiki to develop a numerical corpus by integration of individual contributions. We have extended MediaWiki with some DBMS capabilities: form based acquisition interface, annotations, query engine. The form based acquisition interface allows users to add data with a global structure (specific fields such as location and free text based fields such as the description of liturgical installations). We use annotations to make links between semi-structured data manipulated by MediaWiki and structured data necessary to query engine. Annotation semantics is guaranteed by an ontology which allows to describe concepts and their relations. As stated by (ABITEBOUL et al. 1999) one of the strengths of semi-structured data is "... the ability to accommodate variations in structure". Our dual approach allows to cope with evolution of knowledge by modifying the ontology and annotations dynamically without modifying database schema. Moreover, ontologies can represent concepts at different levels of abstraction (granularity). For example, for some archaeological remains the type of mortar can be described, for others only the presence of marks is recorded.

Figure 2 presents an overview of the interactions between different types of users and our system. Yellow arrow describes data capture, red arrows present semantic queries and blue arrows symbolize links between semantics and semi-structured data.

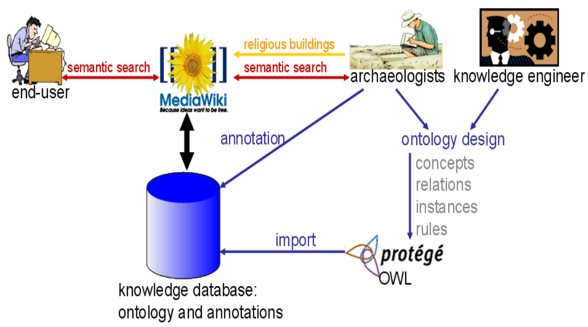

Figure 2: *Outline of WikiBridge users interaction.*

3.State of art

In traditional wiki, semantics is not explicit, but is implicitly described by links between pages and by the context of the link (surrounding text). In the following subsections we give a short overview of semantic wiki and their theoretical background.

3.1.Semantic wiki engines

A semantic wiki is a wiki which includes semantic web technologies to cope with domain knowledge generally represented by ontologies. Semantic wikis can be built on top of existing wiki or created from scratch. In (BUFFA et al. 2008) authors have identified two approaches of wiki based on their relationship to the ontology: 1) wiki centric approaches use the wiki to organize knowledge i.e ontology emerges from the wiki through categories and links or 2) ontology based approaches allow to import an existing ontology and use it in the annotation process.

However, we consider that this classification is too restrictive and we propose to define a third category of approaches which combines the first two ones.

Platypus, Rise and Rhizome fall in the first category. The first system was probably Platypus (CAMPANINI et al. 2004) which has focused on the creation of RDF meta-data. Rhizome (SOUZIS 2005) allows to edit content, structure and meta-data in RDF format. URL represent elements such as structural components, abstract entities, and relationships. Rise wiki (DECKER et al.

2005) allows user to create and edit the ontology with wiki pages (concepts) and links (relations). Moreover the same approach has been used in numerous wiki engines. For example in MediaWiki, Category is the simplest form of annotation used to classify wiki pages. Semantic MediaWiki (KRÖTZSCH et al. 2006) extends MediaWiki and provides new features such as: 1) Relations are used to describe relationships between two pages by assigning annotation to existing links and 2) Attributes allow users to specify relationships between pages and literals. Table 1 gives a concrete example of a page in MediaWiki using links and categories and the same page using Semantic MediaWiki capabilities.

`The city of Moulis is located on the [[Medoc region]]. The building of the High [[Middle Ages]] was discovered in 1993 under the present parish church, largely Romane, surrounded by a parish cemetery until 1901, then transformed in the public square...` `[Category:ArchaelogicalSite]}`
`The city of Moulis is located on the [[region::Medoc]] region. The building of the High [[Middle Ages]] was discovered in [[date::1993]] under the present [[building-type::parish church]], largely Romane, surrounded by a parish cemetery until [[date::1901]], then transformed in the public square...` `[Category:ArchaelogicalSite]`

Table 1: *Concepts, relationships and attributes in MediaWiki and Semantic MediaWiki.*

In (VRANDEČIĆ et al. 2006), the authors propose an equivalent representation between OWL concepts and Semantic MediaWiki constructs (Table 2). This approach mainly produces assertions which correspond to ABox statements.

OWL	Semantic MediaWiki
OWL individual	normal article page
owl:Class	article in namespace Category
owl:ObjectProperty	article in namespace Relation
owl:DatatypeProperty	article in namespace Attribute

Table 2: *OWL Concepts and Semantic MediaWiki constructs.*

In short this first category of semantic wikis can be used to present knowledge by structuring concepts through pages, categories and links.

Makna (DELLO et al. 2006) and BOWiki (BACKHAUS et al. 2007) are two examples of the second category. In Makna, users can create semantic content using RDF statements referencing pre-existing ontologies. They are provided with an extended Wiki syntax and with assistant tools to simplify ontologies based annotation process. A specific application MannWiki has been developed for sharing knowledge on micro-array in LifeScience domain and allows to make references to GeneOntology. BOWiki is used to collaboratively create knowledge base in biological domains. Moreover, BOWiki allows to access to several ontologies like the GeneOntology and ontologies about cell types or anatomy. This second category of semantic wikis based on pre-existing ontologies can be used as a platform to build applications that require a global consensus over knowledge in order to maintain the quality of data.

The third category of semantic wiki is an hybrid approach that can be used both to build ontologies in a collaborative ontology engineering process or to import pre-existing ontologies to annotate documents. SweetWiki (BUFFA et al. 2006) allows users to tag pages, (called social tagging) but also integrate external ontologies. The set of users tags generates a folksonomy. In addition, SweetWiki adds a WYSIWYG editor for managing content and meta-data, a reasoning engine used for querying the wiki content. Semantic MediaWiki (KRÖTZSCH et al. 2006) is an hybrid engine which enables to load ontology and to consult it as wiki pages. Semantic MediaWiki does not provide yet a complete set of tools in particular for semantic constraints management.

Our approach of semantic wiki is directed towards scientific application domains which contribute to produce knowledge. This kind of application relies on core ontologies that act as a consensus. Knowledge is enhanced by querying and analyzing data, new concepts can emerge and new constraints can be found out. As a result, ontologies can be modified dynamically and semantic checks are necessary to find inconsistent annotation with regards to an ontology version.

3.2.Theoretical background

Semantic web technologies such as RDF and OWL ontologies are based on well founded theoretical background. RDF is based on conceptual graphs and semantic networks (SOWA 1984). OWL is based on description logics (BAADER et al. 2003). Some features of description logics make it difficult to use for validating data or annotations through integrity constraints: 1) OWL-DL works in open world assumption; 2) OWL does not use the unique name assumption. Finding inconsistent annotations requires to evaluate OWL rules in a closed world assumption to detect violation. Some compelling solutions are described in (SIRIN et al. 2008).

4.WikiBridge's architecture

In the next section we present WikiBridge's architecture and detail some key features for archaeological application.

4.1.Acquisition

Two types of acquisition form have been created: a form for entering a record corresponding to atomic building and a form corresponding to a group of buildings. These two forms are a simplified version of paper forms filled by archaeologists to publish their research results. Electronic forms are created by using Semantic Forms extension for MediaWiki. It allows users to fill in fields through a model (figure 3). A non-expert in archeology can then easily feed the wiki from paper forms already made.

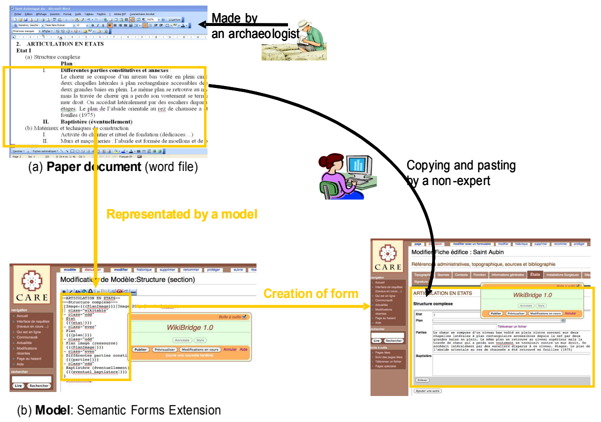

Figure 3: *Methodology for entering a form.*

Wiki allows to enrich the description of a form by using: 1) multimedia content (photographies, maps, sound, video), 2) links (i.e. internal links to handle the case of group of buildings or external links to the URL of a museum) and 3) external services such as geolocation (figure 4).

Figure 4: *Multimedia enrichment.*

4.2.Annotation

To improve quality of search, we expanded MediaWiki with a semantic component. The semantic component consists of annotations made by experts, that are guaranteed by an application ontology.

WikiBridge restricts the access to ontological knowledge management to a predefined set of Wiki users: we argue that implementing such functionality without adequate process-level support might have uncontrolled consequences on the operation of the overall wiki system. Knowledge engineers interacting with archaeologists create the ontology with standard tools like *Protégé*. Ontology contains concepts of the domain, instances and rules. This assertion is considered as a rule in the ontology. Knowledge engineers can test consistency of the ontological representation with reasoners such as Racer or Pellet.

Within the cultural heritage domain, the CIDOC Conceptual Reference Model (CIDOC) has emerged as a domain ontology. CIDOC CRM deals with concepts at a high level of generality. Its scope encompasses the general culture heritage domain and it is envisaged as "semantic glue" useful for exchange among diverse information sources. Application ontologies contain all the definitions that are needed to model the knowledge required for a particular application. Typically, application ontologies are a mix of concepts that are taken from domain ontologies and specific application. We have developed an application ontology as a CIDOC CRM extension covering the Christian European buildings (figure 5).

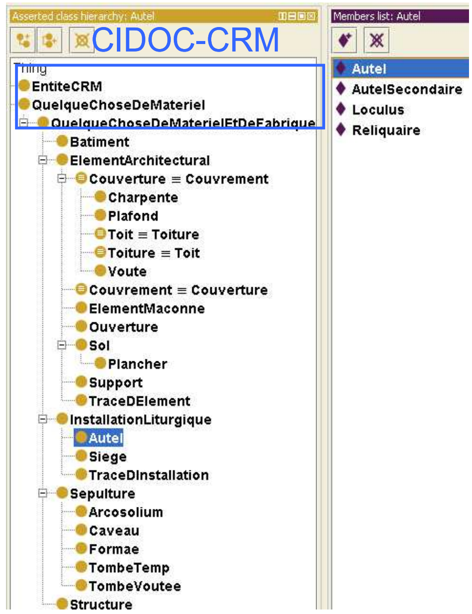

Figure 5: *CARE ontology (extract).*

The CARE ontology contains concepts and relations about:

1) Man made objects (type of buildings, architectural elements, liturgical installations)

2) temporal data and

3) others entities that are specific to Christian buildings.

For instance, building is a concept, it has liturgical functions which is another concept. Cathedral and episcopal are instances of building and liturgical functions. A cathedral has one function, that function can be episcopal or archiepiscopal.

Experts directly enter and modify annotations through an extension of the wiki's editing interface (figure 6). The consistency of annotation in relation to the context (field in the form) is checked by specific application module. At this stage of development, this functionality is directly implemented in the application. As a result, annotations are stored in an ad-hoc RDF triple store. Ontology is then stored in a relational database and queried by PHP programs to fill in the annotation wizard.

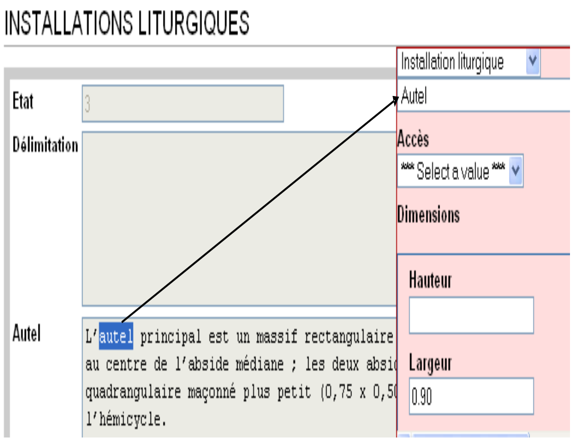

Figure 6: *Annotation interface.*

4.3.Query engine

Although, most of wikis includes by default an engine that can only query full-text. The aim of our engine is to provide semantic search by filling in parameters associated with ontology concepts. Three types of interfaces (figure 7) for building semantic queries are developed: 1) a wizard lets users to specify search parameters to engine; 2) users can create query models that are then stored; and 3) user can navigate through an ontology tree.

Three kinds of results can be displayed: 1) results can appear in a list containing links to articles, at the right annotation place, so where the information is given. User can then manually navigate through articles interlinked; 2) factbox, each article is displayed with its annotations and 3) users can select annotation to be displayed in the result. From this result, users can obtain the list of the articles in which have the same annotation. This third kind of display is a mix of result list and factbox and allows more sophisticated analysis.

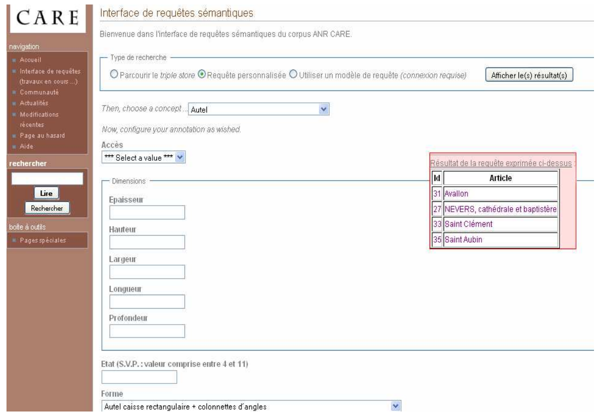

Figure 7: *Query Interface.*

5.Conclusion

A feasible combination of wiki and Semantic Web technologies should preserve the key advantages of both technologies: the simplicity of wiki systems as shared content authoring tool, the power of Semantic Web technologies with.respect.to. structuring and retrieving knowledge. In this article, we have demonstrated that flexibility required by scientific applications can be achieved by using wiki with semantic web technologies. At this stage of development, data quality is maintained by ad-hoc programs.

Ontologies can also include logical rules representing the domain constraints. Reasoners are used to verify the semantics contained in ontology. For example, the following constraint "a building can be consecrated to a saint only if the construction date is later than the death of the holy person" is represented by the following rule:

$$\texttt{isConsecrated(?b,?p)} \leftarrow \texttt{hasConstructionDate(?b,?d1)} \wedge \texttt{hasDateDead(?p,?d2)} \wedge \texttt{d1} \geq \texttt{d2}$$

In the next version of WikiBridge, automated verification of integrity constraints will be performed by a reasoning tool. For spatial analysis, we are developing in parallel web services to export data to PostGIS.

Acknowledgements This work is supported by the ANR (ANR-07-CORP-011).

References

ABITEBOUL S., BUNEMAN P., SUCIU D., 1999. Data on the Web: from relations to semistructured data and XML. *Morgan, Kaufmann Publishers InC.* San Fransisco, CA, USA.

ANKOLEKAR A., KRÖTZSCH M., TRAN T., VRANDEČIĆ D., 2008. The two cultures: Mashing up Web 2.0 and the Semantic Web. *Journal of Web Semantics* 6(1), pp. 70–75.

BAADER F., CALVANESE D., McGUINNESS D., NARDI D., PATEL-SCHNEIDER P.F. Edts, 2003. The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press.

BACKHAUS M., KELSO J., 2007. BOWiki – a Collaborative Annotation and Ontology Curation Framework. Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC).

BONOMI A., MOSCA A., PALMONARI M., VIZZARI G., 2008. Integrating a Wiki in an Ontology Driven Web Site: Approach, Architecture and Application in the Archaelogical Domain. Proceedings of the 3rd Semantic Wiki Workshop (SemWiki), pp. 106-118.

BUFFA M., GANDON F., ERETEO G., SANTER P., FARON C., 2008. SweetWiki: A semantic wiki. *Journal of Web Semantics* 6(1), pp. 84-97.

CAMPANINI S. E., CASTAGNA P., TAZZOLI R., 2004. Platypus Wiki: A Semantic Wiki Wiki Web. Proceedings of Semantic Web Applications and Perspectives (SWAP) - 1st Italian Semantic Web Workshop.

CIDOC Conceptual Reference Model (CRM) http://cidoc.ics.forth.gr/

DECKER B., RAS E., KLEIN B., HOECHT C., 2005. Self-organized Reuse of Software Engineering Knowledge Supported by Semantic Wikis. Proceedings of 4th International Semantic Web Conference (ISWC).

DELLO K., PASLARU E., SIMPERL B., TOLKSDORF R., 2006. Creating and using Semantic Web Information wih MAKNA. Proceedings of the first Workshop on Semantic Wikis – From Wiki to Semantics (SemWiki).

GRENON P., SMITH B., 2004. SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition and Computation: An Interdisciplinary Journal* 4(1), pp. 69-104.

KRÖTZSCH M., VRANDEČIĆ D., VOLKEL M., 2006. Semantic MediaWiki. Proceedings of 5th International Semantic Web Conference (ISWC 2006), vol. 4273 of LNCS, pp. 935-942.

SIRIN E., SMITH M., WALLACE E., 2008. Opening, Closing Worlds – On Integrity Constraints. Proceedings of 5th OWLED Workshop on OWL : Experiences and Directions.

SOUZIS A., 2005. Building a Semantic Wiki. *IEEE Intelligent Systems* 20(5), pp. 87-91.

SOWA J., 1984. Conceptual Structures Information Processing in Mind and Machine. *Addison Wesley.*

SPEAR A.D., 2006. Ontology for the twenty First century: An Introduction with Recommendations. Technical report, INFOMIS, Sarrbrück, Germany.

SPYNS P., MEERSMAN R., JARRAR M., 2002. Data modelling versus Ontology engineering, *SIGMOD Record* 31(4), pp. 12-17.

VRANDEČIĆ D., KRÖTZSCH M., 2006. Reusing Ontological Background Knowledge in Semantic Wikis. Proceedings of the first Workshop on Semantic Wikis – From Wiki to Semantics (SemWiki).