User participation in the structuring of cultural heritage data in a collaborative knowledge base

Ashraf Amad

Université Paris8
Laboratoire Paragraphe, équipe CITU
2 rue de la Liberté
93526 Saint Denis, France
ashraf.amad@citu.fr

ABSTRACT. *In this article, we present an approach which allows users to semantically represent their informational needs in a dynamic way. These representations are used to add user-interface input elements which have implicit semantics according to the CIDOC CRM ontology and which are used to insert cultural heritage data. To facilitate the indirect creation of ontological paths by the ordinary user, we exploit the techniques of mind mapping and we associate the terminological level, in the context of our application, with the conceptual one. Our objective is to find an effective solution to the situation where there is a lack of data in a particular cultural heritage context. The semantic technologies offer several advantages, but they also require several adaptations in order to allow the ordinary user participate effectively in the ontological representation.*

RÉSUMÉ. *Dans cet article, nous présentons une approche qui permet aux utilisateurs de représenter sémantiquement leurs besoins informationnels d'une façon dynamique. Ces représentations sont utilisées pour ajouter des éléments de saisie de l'interface utilisateur qui servent à insérer des données du patrimoine culturel et qui possèdent des sémantiques implicites selon l'ontologie CIDOC CRM. Pour faciliter la création indirecte de chemins ontologiques par l'utilisateur ordinaire, nous exploitons des techniques de représentation de cartes mentales et nous associons le niveau terminologique, dans le contexte de notre application, avec le niveau conceptual. Notre objectif est de trouver une solution efficace face au manque de données dans un contexte particulier du patrimoine culturel. Les technologies sémantiques offrent de nombreux avantages, mais exigent aussi plusieurs adaptations afin que les utilisateurs puissent participer efficacement à la représentation ontologique.*

KEYWORDS: *semantic web, cultural heritage, ontology paths, indexing, classification, collaborative platform, mind map, Wordnet, knowledge base*

MOTS-CLÉS : *web sémantique, patrimoine culturel, chemins ontologique, indexation, classification, plateforme collaborative, carte mentale, Wordnet, base de connaissances*

1. Introduction

The users of digital platforms can not only consume content, but also participate actively in these platforms through providing content and performing certain tasks that require human intelligence (Oomen *et al.,* 2011).

This is particularly interesting in the domain of cultural heritage where the participation of users increases their motivation and gives them the opportunity to engage and contribute to the collective memories that they care about and to the conservation of cultural heritage (Paraschakis, 2013).

This is very useful especially when there is a lack of data in a particular cultural heritage context, while there exists actually a rich environment of cultural heritage that lacks documentation (Hernández *et al.,* 2008). This situation corresponds to the Church of the Nativity in Bethlehem, which is a very important historical place that is considered as a UNESCO World Heritage Site.

However, this church has no knowledge base that enables its users to provide content and information they have about the church. Additionally, such a knowledge base should provide a decentralized and dynamic solution as it can't use a static schema since the information requirements in this domain are complex by nature and cannot be statically predefined. Moreover, the users are disperse worldwide with different information needs and perspectives. Our research aims to find an effective solution to this situation.

In fact, the acquisition of data is an essential step in the documentation of cultural heritage and the use of semantic technologies for collecting the data is an interesting approach since it provides, in a native way, a rich ontological structure, a high level of interoperability, and machine readability. All of that allows for better automatic processing.

The domain experts are used to add data using input elements of the graphical user interface such as form fields (Hohmann *et al.,* 2012) and the traditional methods of data acquisition use these statically predefined input elements. Therefore, the usage of these elements for the acquisition of semantic data is a logical choice to break the entry barrier for the ordinary user. However, with the available semantic technologies, should these interface elements be always statically predefined?

Certain research studies (Frischmuth *et al.,* 2012) (Vrandecic *et al.,* 2014) allow the users to enter the data using form fields and to use certain semantic elements statically (certain classes and properties) in order to add metadata values, but not to represent their informational needs in a dynamic manner especially in a rich and a complex domain such as cultural heritage.

(Scholz *et al.,* 2012) is a research study that allows the definition of ontological paths by an administrator who has sufficient knowledge about semantic technologies, ontological modelling, and the construction of form fields based on

ontological paths. In such an environment, each input field is associated with an ontological path that is predefined by the administrator. This allows the data to be added as instances of the used ontology classes in each path.

This same idea of ontological paths is presented in different articles such as (Nussbaumer *et al.,* 2009) in a project called BRICKS which uses a metadata schema that should be mapped to the CIDOC CRM ontology in order to ensure semantic interoperability.

The authors have found that the corresponding representation in the CIDOC CRM ontology was always an ontological path that connects certain classes using certain properties in a particular sequence to produce the desired semantics that correspond to the original metadata schema used in the BRICKS project. The authors have proposed a mapping approach based on the adaptation of the relational database model to the RDF graph based model.

However, the missed advantage in these projects is that the ordinary users are not able to represent their dynamic information needs in a way that is associated with precise semantics which can be used to create input elements for inserting cultural heritage data.

Therefore, our research concentrates on the possibilities to exploit the semantic technologies in order to allow the ordinary users participate in the representation of their information needs to structure cultural heritage content and data without exposing them to the complexity of ontologies and the techniques of formal knowledge representation.

Our hypothesis is that the ordinary users are able to represent semantically their informational needs if we give them the appropriate tools and if we apply certain principles that we will present during this article.

In our approach to deal with the problem of the lack of data in a particular cultural heritage context, we adapt a semantic wiki data (Scholz *et al.,* 2012) and we also benefit from several resources of structured and semi structured content that already exist such as Wikipedia and Wordnet to avoid starting our platform from scratch and to ensure dynamic interaction with the semantic web, which enriches the data in our knowledge base (Merti, 2012).

We also benefit from the social networks which give us the possibility to interact with an already existing community of users (Rosemberg *et al.,* 2012) (Bozzon *et al.,* 2013) having hypermedia resources (text, images, videos, URLs, etc.) published in their profiles that we semantically analyze and index in order to improve and personalize the user experience, increase the participation, and ensure a higher quality of added data and content.

2. Our approach

Our approach starts by studying the informal semantic representation by ordinary users using mind maps and how can we benefit from the principles and techniques of mind-mapping in order to give the users the possibility to participate indirectly in the formal semantic representation of constructing ontological paths

However, the user would not be required to deal directly with the abstract conceptual classes of our base ontology. This had been made possible after we associated the terminological level in the semantic sphere of our particular cultural heritage context with the conceptual level of the CIDOC CRM ontology classes.

The users will interact with the terminological level which represents the lexical knowledge and which contains the linguistic terms that users are familiar with. The advantage is that these terms are connected to the ontology which represents the domain knowledge.

2.1. *The informal semantic representation by ordinary users*

In our approach, we have benefited from the principles of mind mapping since they are well known by the ordinary users. However, there are main differences between the semantic representation using mind mapping and ontologies that we had to take into consideration in our approach.

Semantic representation using mind mapping has the following main characteristics:

– The users "suppose" certain semantics (in their mind) without representing them explicitly (Eppler, 2006).

– There is no formal model or syntax that represents the mind maps (Siochos *et al.,* 2011).

– There is no separation between the level of concepts and the level of instances (Sarker *et al.,* 2008).

– The users have a tendency and a preference to use the instances and the more concrete concepts as they correspond to a clearer mental image in their mind (Schopenhauer, 2011) (Sarker *et al.,* 2008).

– There is a common usage of complementary resources such as URLs, images, documents, colors, etc. (Siochos *et al.,* 2011).

– The representation is concentrated around a central idea "a main topic" which is connected with sub ideas or topics which constitutes a functional division. (Eppler, 2006).

2.2. *The ontological representation*

One of the widely cited definitions of ontology is “formal specification of a shared conceptualization” (Borst, 1997). Hence, ontologies are explicit in their representation. For example, the CIDOC CRM that we use in our platform as a base ontology is an event-centric ontology (Doerr *et al.*, 2006) where most of the semantic representations are centered around pertinent events such the class E12_Production which represents an event for the production of physical man-made things and the class E67_Birth which represents the birth of a human being, and so on.

An ontological path is a sequence of ontology classes and properties ordered in certain way in which it represents particular semantics. An example of an ontological path to represent that a painter drew a particular painting is done with the following classes and order:

– E84_Information_Carrier (to describe to the painting)

– P108i_was_produced_by

– E12_Production (to describe the production of the painting)

– P14_carried_out_by

– E21_Person (to describe the painter)

– P1_is_identified_by

– E82_Actor_Appellation (to describe the name of the painter)

2.3. *The indirect creation of ontological paths by the users*

In order to create the full ontological path, we ask the user to only choose the initial and final concept (without thinking about any intermediate events or steps for the moment). Our algorithm tries to construct the full path which represents the desired semantics.

In order to reduce the possible choices when constructing a certain path, we apply the idea of functional division which was noticed in mind maps. However, this time we use this idea with our base ontology, which facilitates its management and processing.

This functional division is already available in the web site of CIDOC CRM ontology (CIDOC, 2014) in HTML format. Therefore, we have developed a parser that extracts certain text elements from HTML tags and then transforms them into RDF data according to a predefined model where certain text elements that represent the ontology classes (starting with capital E) and properties (starting with capital P) belong to a certain functional division as provided in the website of the ontology.

We have found this functional division very useful to reduce the choices proposed to the user in order to represent the desired semantics in a particular semantic sphere and to reduce the required analysis and processing.

For example, if the user wants to represent the production, the modification, or the used material of an artifact, he would choose "Material and Technique Information".

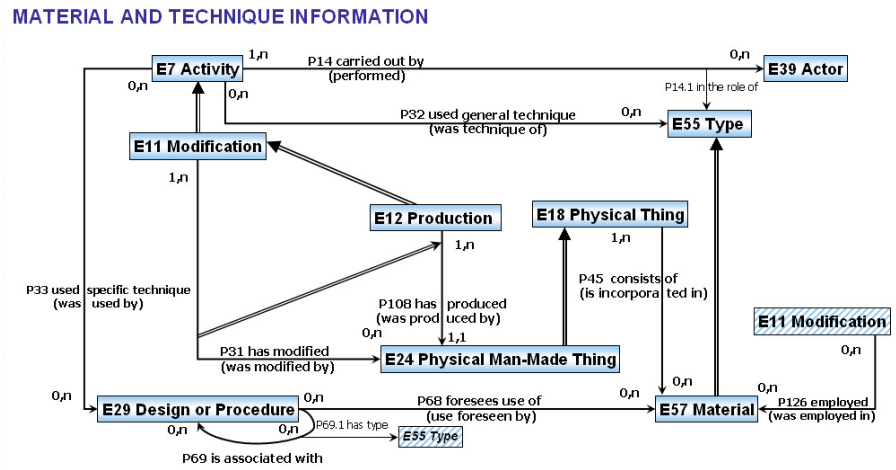


Figure 1. *The funcional division "Material and Technique Information" of the CIDOC CRM ontology.*

We find the shortest path between the initial and final concept inside the functional division chosen by the user. In the simplest case, we use the findShortestPath method of Jena open source semantic web framework for java. If the found path does not represent the semantics desired by the user, our recursive algorithm proceeds from the current concept (currently the initial) towards all its possible properties (where the current concept is the rdf:domain for these properties) and then towards all the possible range concepts of these relations (where the concept is rdf:range for these properties).

The algorithm stops when it finds the final concept. We use a stack to trace the concepts in the path in order to avoid visiting the same concept two times to ensure that we do not enter inside a loop.

To remove the semantic ambiguity and to select a particular path among the found paths, we use the intermediate events in the found paths to perform the semantic disambiguation, so we ask the user if he is describing the production or the modification of the paining. The user's answer will remove the semantic ambiguity and will favor a particular ontological path over the others.

Additionally, instead of asking the user to choose among the abstract concepts of the CIDOC CRM ontology, which are hard to understand for an ordinary user, the user chooses a concrete concept or even a precise instance that we have associated automatically to the abstract classes of the CIDOC CRM ontology. For example, the user can start with “painting” instead of “E84_Information_Carrier” or “nativity church” instead of “E22 Man-Made Object”.

This ensures an easier and a higher quality semantic representation. This is because that the lexical terminological level is much more intuitive for the user than the ontological conceptual level (Schopenhauer, 2011). However, the produced path includes uniquely the conceptual level (the ontology classes), but applies to any corresponding terminology.

2.4. *The association of the terminological and conceptual level in our context*

Certain research studies such as (Vlachidis, 2012) which realizes certain associations between the terminological level as metal, gold, and copper with the conceptual level of the CIDOC CRM ontology as E57_Material class, using the available taxonomies and thesauri which hold the terms (the instances) corresponding to certain CIDOC CRM classes. However, this approach is limited to certain classes which have a direct terminological correspondence in a taxonomy or a thesaurus.

In order to associate the terms in our particular cultural heritage context with the classes of the ontology, we used the supervised machine learning. For the training data, we started by the terms that are provided by the ontology CIDOC CRM as examples of every class and which are available in the web site of the CIDOC CRM ontology.

These terms are already classified according to the classes of the ontology and we increase the quantity of training data using Wordnet as we will show in the next section 2.5.

We find the Wikipedia articles that correspond to these terms. For example, Mona Lisa, Rosetta stone, etc. We represent the text of each of these articles as a vector by calculating the weight of each term in each article using the TFIDF (term frequency–inverse document frequency) (Hocine, 2011).

The machine learning algorithm that we use is NaiveBayes and this classical approach of learning gives us an F-Measure of 81.3222%. However, our choice of NaiveBayes is because of its simplicity and performance especially for short documents and when using reduced number of descriptors.

Therefore, to improve the results of the classical approach and to reduce the quantity of necessary training data, we have reduced the dimensionality of the vectors representing the documents according to the Zipf law. We did so by

reducing the number of the used descriptors. We have realized that by the creation of "semantic aspects" where each aspect represents an agglomeration of distinguishing terms not for a certain class, but for a precise semantic aspect. For example, the aspect of biological process is composed of certain terms such as breath, sleeping, digestion, birth, etc.

These terms are taken from the lexical database Wordnet which groups the terms of the English language into synsets. Each synset is a set of synonyms that are connected to other synsets through various linguistic semantic relations. Each semantic aspect corresponds to a particular Wordnet synset (such as biological process) and the agglomeration of terms for that aspect includes all the terms that are found by following the hyponym relations (the more specific synsets) below that synset. This constitutes a semantic sphere for that particular aspect.

For each term that belongs to a particular semantic aspect (agglomeration of terms), we calculate its TFIDF in each document, and we calculate a score for each aspect. The score S_a of a certain aspect *a* for *n* terms that belong to that aspect in a document *d* is:

$$S_a = \frac{\sum_{n=1}^{n} TFIDF\ term\ n}{n} \qquad [1]$$

The training data are represented as a matrix where for each document that we know its class, we add the score of each aspect:

biological process aspect	Physical object aspect	aspect *a*	Semantic aspect / Document class
S_1= 0.65	S_2= 0.88	S_a= 0.35	E18_Physical_Thing
S_1= 0.92	S_2= 0.58	S_a= 0.79	E39_Actor
S_1= 0.21	S_2= 0.94	S_a= 0.49	class *c*

Table 1. *The matrix that represents the learning data*

If the machine learning algorithm has to learn all the distinguishing terms for each class directly from the text by itself, a lot of training data would be necessary. Our method allows the algorithm to learn from the results of the analysis that we proposed above, and this would reduce the quantity of required training data because the learning is more focalized with reduced number of dimensions for the vectors representing the documents.

To find the terms that represent the semantic sphere of our particular cultural heritage context and which we need associate and classify according to the CIDOC

CRM ontology using our learned classifier, we searched the web for the documents (such as the Wikipedia articles) that contain certain keywords that we have defined manually and which correspond to our application, such as: Palestine, Bethlehem, Nativity Church, Jesus, etc. We extend these keywords automatically using their corresponding synsets in Wordnet and following certain semantic relations with other synsets such as hyponyms and hypernyms, which yields more related keywords that improve the results of our search. Among the found documents, we chose the documents that have a high cosine similarity between their vectors and the vector of our query.

We perform natural language processing using a part of speech tagger to find the nouns NN and the proper nouns NNP in the text. Moreover, to find the nominal phrases (which are sequences of words that play the role of nouns and that have a noun as their head word) such as “Nativity Church” or “Jesus Christ”, we used a chunker and a parser which analyze the grammatical structure of the phrase in order to find the nominal phrases NP in each phrase (Vadas *et al.,* 2011).

To improve these results, we also used an NGram Tokenizer which tokenizes the text in N words sequences, such as bigram (two words sequence) and we compare the found sequences between all the documents in order to extract the correct sequences, which are repeated as a single unit (same sequential order) in several documents. To find the terms that are particular to our cultural heritage context, we choose the nouns, proper nouns, and nominal phrases that have a high TFIDF in the corpus of documents that we found.

To be in harmony with our learning approach, we find the Wikipedia articles which correspond to the terms that are particular to our context and we represent the text of each article according to the same semantic aspects and in the same way that we followed in our learning approach. Finally, we implement the classification of these articles according to the classes of the CIDOC CRM ontology using our trained classifier.

2.5. *Increasing the learning examples*

We have also used Wordnet to increase semi automatically the learning examples. As we mentioned in section 2.4, we started with some examples of the ontology classes which are already classified and provided in the definition of the conceptual model of the CIDOC CRM ontology presented in its website, such as:

Disk	rdf:type	E84_Information_Carrier
Coliseum	rdf:type	E22_Man-Made Object
Gold	rdf:type	E57_Material

Table 2. *Instances of the CIDOC CRM ontology classes*

For example, “disk” is provided by the ontology definition as an example of E84_Information_Carrier, our algorithm obtains its direct hypernym (the more general concept of disk) which is «memory device, storage device, a device that preserves...» and links that more general concept as subclass of E84_Information_Carrier, then we obtain the full hyponym (the instances) of the more general concept of disk (the sister terms of disk). These instances constitute a lot of examples of the E84_Information_Carrier class.

- S: (n) memory device, storage device (a device that preserves information for retrieval)
  - ***direct hyponym*** / *full hyponym*
    - S: (n) acoustic storage (a storage device consisting of acoustic delay lines)
    - S: (n) auxiliary storage, external storage, secondary storage (a data storage device that is not the main memory of a computer)
    - S: (n) buffer, buffer storage, buffer store ((computer science) a part of RAM used for temporary storage of data that is waiting to be sent to a device; used to compensate for differences in the rate of flow of data between components of a computer system)
    - S: (n) magnetic disk, magnetic disc, **disk**, disc ((computer science) a memory device consisting of a flat disk covered with a magnetic coating on which information is stored)
    - S: (n) magnetic tape, mag tape, tape (memory device consisting of a long thin plastic strip coated with iron oxide; used to record audio or video signals or to store computer information) *"he took along a dozen tapes to record the interview"*
    - S: (n) memory, computer memory, storage, computer storage, store, memory board (an electronic memory device) *"a memory and the CPU form the central part of a computer to which peripherals are attached"*
    - S: (n) optical disk, optical disc (a disk coated with plastic that can store digital data as tiny pits etched in the surface; is read with a laser that scans the surface)

Figure 1. *Using Wordnet to increase the training examples*

We also search the Wikipedia articles that correspond to the Wordnet examples that we obtained, then we search the categories of these articles in a similar procedure to what we did with Wordnet. These categories enable us to access new more examples (the children articles of these categories) which increases the training data.

3. Results

The CIDOC CRM ontology is an ontology of high level of abstraction. Therefore, depending on the level of analysis, certain terms discussed above, such as church and painting, can be themselves considered subclasses of the CIDOC as they contain instance level terms such as “Nativity Church” and “Mona Lisa”. This helped us to specialize the CIDOC CRM ontology for our context in the cultural heritage domain. For example, we have added the following subclasses:

E24a_Church	rdfs:subClassOf	E22_Man-Made Object
E12a_Church_Construction	rdfs:subClassOf	E12_Production
E84b_Painting	rdfs:subClassOf	E84_Information_Carrier

Table 2. *An example of specializing the CIDOC CRM ontology to the context of our application*

Specializing the ontology classes to our context ensures a more precise semantic representation and a more intuitive user experience. This describes the domain in a better way and reduces the semantic gap between high level concepts and instance terms.

Moreover, our approach of classifying the terms according to the classes of the CIDOC CRM ontology had an F-Measure of 86.6667 % (better than the classical approach). For example, in one of the tests that we did, the results of the following evaluation measures for this sample of CIDOC CRM classes were as follows:

Precision	**Recall**	**F-Measure**	**Class**
1.000	0.800	0.889	E84_Information_Carrier
0.667	0.667	0.667	E24_Physical_Man-Made_Thing
0.833	1.000	0.909	E26_Physical_Feature
1.000	1.000	1.000	E39_Actor

Table 3. *A sample of results of the evaluation measures*

In fact, associating the terminological terms with the conceptual classes of the CIDOC CRM ontology produces a rich semantic network that can be used not only for constructing ontology paths but also for various tasks in information retrieval, natural language processing, and the semantic web especially that the terms in the

text of the cultural heritage content in the context of our application are now connected to the ontology classes.

4. Discussion and conclusion

The participation of the user in the semantic representation seems very advantageous because this user is given the possibility to structure the information according to his dynamic needs. However, this requires several adaptations and should be applied in a particular context to give the user the opportunity to interact with the lexical terms of that particular context instead of upper level abstract classes.

We benefited from the principles of mind map representation since mind mapping is well known and intuitive. The two principles that found fundamental for the realization of our objective to enable the end user participate indirectly in the construction of ontological paths are that event-centric representation is much less intuitive than the object-centric one for the end user and that the concepts of high level of abstraction are also an entry barrier for end users to ontology based semantic representations.

To help the represent his desired semantics, there is a need to traverse the RDF graph of the ontology with the constraint of not entering inside a loop. The simplification of the graph traversal problem by dividing the ontology into functional divisions and the progressive construction of a complex ontological path was very useful for our approach and made certain tasks easier.

To give the user the possibility to interact with the concrete concepts and instance terms in our cultural heritage context, we have benefited from external sources such as Wikipedia and Wordnet. In order to associate these terms with our base ontology, we used the supervised machine learning. However, our approach to augment the training data was necessary in order to obtain reasonable results.

The reduction of vector dimensionality was also an essential step especially with NaiveBayes classifier. The introduction of our method of “distinguishing semantic aspects” has improved the F-Measure from (81.32 %) to (86.6667%). This has helped us in our general approach in order to realize our objective which corresponds to our initial hypothesis.

The user-interface input elements that are based on the created ontological paths allow users to add pertinent data about the church. All of this data is saved to our knowledge base and can be processed in different useful ways, such as searching, navigation, visualization, and inference. This invokes more informational needs that encourage users to perform more semantic representations.

In fact, we can conclude by saying that the ordinary user is capable to represent his informational needs and participate actively in a semantic knowledge base. The

participation of the user in such an environment increases his engagement and interest in a dynamic way especially in the cultural heritage domain because it gives him the opportunity to participate in preserving the cultural heritage that he cares about and allows him go beyond simple data entry to the structuring of cultural heritage content and data.

5. References

Borst, Willem Nico. 1997. « Construction of engineering ontologies for knowledge sharing and reuse.» University of Twente, Enschede Netherlands. http://purl.utwente.nl/publications/17864.

Bozzon, Alessandro, Marco Brambilla, Stefano Ceri, Matteo Silvestri, Giuliano Vesci, et Politecnico Milano. 2013. « Choosing the Right Crowd : Expert Finding in Social Networks Categories and Subject Descriptors. » In *EDBT '13 Proceedings of the 16th International Conference on Extending Database Technology*, 637-48. Genoa, Italy: ACM Press.

CIDOC. 2014. « Graphical representation of CIDOC - version 5.0.1. » *ICOM*. http://www.cidoc-crm.org/cidoc_graphical_representation_v_5_1/graphical_representation_5_0_1.html.

Paraschakis, Dimitris. 2013. « Crowdsourcing cultural heritage metadata through social media gaming. »

Doerr, M., et A. Kritsotaki. 2006. « Documenting Events in Metadata. » In *The 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST*. Cyprus. http://www.cidoc-crm.org/docs/fin-paper.pdf.

Eppler, Martin J. 2006. « A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. » *Information Visualization* 5 (3): 202-10. doi:10.1057/palgrave.ivs.9500131.

Frischmuth, Philipp, Michael Martin, Sebastian Tramp, Thomas Riechert, et Sören Auer. 2012. « OntoWiki – An Authoring , Publication and Visualization Interface for the Data Web. » *Semantic Web Journal – Interoperability, Usability, Applicability* 1 (5): 1-25.

Hernández, Francisca, Luis Rodrigo, et Jesús Contreras. 2008. « Building a cultural heritage ontology for cantabria. » In *Annual Conference of CIDOC*, 1-14. Athens. http://cidoc.mediahost.org/archive/cidoc2008/Documents/papers/drfile.2008-06-28.pdf.

Hocine, Matallah. 2011. « Classification automatique de textes: approche orintée agent. » Universie de Tlemcen.

Hohmann, Georg, et Mark Fichtner. 2012. « Embedding an ontology in form fields on the web. » *Semantic Web Journal – Interoperability, Usability, Applicability*.

Merti, Georgia. 2012. « Roman Necropolis : An Ontology for the Semantic Web. » University of Patras.

Nussbaumer, Philipp, et Bernhard Haslhofer. 2009. « Putting the CIDOC CRM into Practice. » *Universitat Wien*. Vienna, Austria.

http://www.cs.univie.ac.at/upload///550/papers/putting_the_cidoc_crm_into_practice-covered.pdf.

Oomen, Johan, et Lora Aroyo. 2011. « Crowdsourcing in the Cultural Heritage Domain : Opportunities and Challenges. » In *Proceedings of the 5th International Conference on Communities and Technologies*, 138-49. Brisbane, Australia. doi:10.1145/2103354.2103373.

Rosemberg, Carlos, Maia De Carvalho, et Elizabeth S Furtado. 2012. « Wikimarks : an approach proposition for generating collaborative , structured content from social networking sharing on the web », 95-98.

Sarker, Biplab K., Peter Wallace, et Will Gill. 2008. « Some observations on mind map and ontology building tools for knowledge management. » *Ubiquity* 2008: 1-9. doi:10.1145/1366313.1353570.

Scholz, Martin, et Guenther Goerz. 2012. « WissKI: A Virtual Research Environment for Cultural Heritage. » *Frontiers in Artificial Intelligence and Applications* Volume 242: 1017-18.

Schopenhauer, Arthur. 2011. *Le vouloir-vivre. L'art et la sagesse*. Édité par Claude Khodoss et Jean Laubier. 9e éd. Presses universitaires de France. http://www.amazon.fr/vouloir-vivre-Lart-sagesse-Arthur-Schopenhauer/dp/213059039X.

Siochos, Vasilis, et Christos Papatheodorou. 2011. « Developing a Formal Model for Mind Maps. » In *First Workshop on Digital Information Management*, 39-44. Corfu, Greece: Ionian University. http://eprints.rclis.org/15842/.

Vadas, David, et James R. Curran. 2011. « Parsing Noun Phrases in the Penn Treebank. » *Computational Linguistics* 37 (4): 753-809. doi:10.1162/COLI_a_00076.

Vlachidis, Andreas. 2012. « Semantic Indexing via Knowledge Organization Systems: Applying the CIDOC-CRM to Archaeological Grey Literature. » University of Glamorgan. http://hypermedia.research.southwales.ac.uk/media/files/documents/2013-07-11/Andreas-Vlachidis_Thesis_print_ready.pdf.

Vrandecic, Denny, et Markus Krotzsch. 2014. « Wikidata : A Free Collaborative Knowledge Base. » *Communications of the ACM* 57 (10): 78-85. doi:10.1145/2629489.