Construction of Batik Heritage Ontology through Automated Mapping

Syerina Azlin Md Nasir[1], Nor Laila Md Noor[2]

[1,2]Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA (UiTM),
40000 Shah Alam, Selangor, MALAYSIA
syerina@kelantan.uitm.edu.my[1], norlaila@ftmsk.uitm.edu.my[2]

ABSTRACT

The wave of ontology has spread drastically in the cultural heritage domain. The impact can be seen from the growing number of cultural heritage web information systems, available textile ontology and harmonization works with the core ontology, CIDOC CRM. The aim of this study is to provide a base for common views in automating the process of mapping between revised TMT Knowledge Model and CIDOC CRM. In this study, manual mapping was conducted to find similar or overlapping concepts which are aligned to each other in order to achieve ontology similarity. Although there are several problems encountered during mapping process, the result shows an instant view of the classes which are found to be easily mapped between both models.

Keywords

Automated mapping, common view, cultural heritage, ontology

1.0 INTRODUCTION

Building an ontology has become common in many domains in comprehend with the demand of Semantic Web. Many definitions of ontology have been reported by different authors ((Garshol, 2004; Gruber, 1993; Guarino, 1998a; Mika, 2007; Noy & McGuinness, 2005). As such, definitions are simplified and described as a set of defined vocabulary with a meaning that constraints the vocabulary to describe concepts (Doan et al., 2004a; Noy, 2004; Shvaiko & Euzenat, 2008). It is the process of discovering the fundamental categories by cataloguing the terms used in the domain and defining the rules governing the mapping of the terms into which the concept naturally falls (Benjamin et al., 1994). The mutual agreement on the terms depicted assists the process of integration, interoperability, knowledge sharing and reuse. For this reason, ontologies offer the conceptual foundation for allowing the semantics of metadata machine-understandable.

The growing number of ontologies has created predicament where the disseminated nature of ontology development has led to disparate ontologies which do not fully understand each other (Doan et al., 2004). Therefore, ontology mapping is seen as the key to the dilemma and this continuously triggers research and development. Works on ontology mapping is viewed by Noy (2004) into two directions involving shared ontology and the emergence of different kinds of automation tool. Debate on both architectures has been long discussed and reported in many works. The growing number of related studies indicates the importance of ontology mapping in realizing the idea of making the communication on the web more meaningful. However, both approaches are looking at the same element which is to find similarity or common view between ontologies. Among identified problems is when ontologies are different in terms of context and background knowledge, it will bring to failure in discovering some correct mappings (Aleksovski, 2008; Sabou et al., 2006). Nonetheless, when tools are used for mapping two schemas or even ontologies, there is high possibility of missing information because not all concepts are mapped between them (Doan et al., 2002).

This paper proposes an approach that could make mapping process becomes easier through the use of automated mapping process which requires the existence of common views. Common view, as explained by Bekiari et al. (2008) provides *"interoperable information systems for those users interested in accessing common or related content"*. Meanwhile, automated mapping can be defined as the process of creating, editing and manipulating concepts and properties of ontology accordingly beforehand to achieve automatic process of mapping afterwards. In this case, automated mapping is achieved after all concepts and properties of TMT Knowledge Model have undergone refinement process to reach certain similarities with standard ontology in cultural heritage domain, CIDOC CRM. This work is aggravated in order to detain and embody the underlying semantic of Malay Textile information for ease of integration and exchange between communities in e-museum applications. Therefore, the objective of this paper is to provide a base for common views in automating the process of mapping between TMT Knowledge Model and CIDOC CRM.

The remainder of this paper is organized as follows. Section 2.0 summarizes the previous work closely related to textile ontology and harmonization works. The description of the approach used is explained in Section 3.0. The preliminary results are presented in Section 4.0. The paper concludes with Section 5.0 that discusses on the anticipated future work directions.

2.0 RELATED WORK

There are a number of available textile ontologies developed, either part of a project or based on individual research being reported. For instance, a semantic-based knowledge flow system for the European home textiles industry (AsIsKnown)[1] is created which aims to collect the product data from diverse textile producers in one system. The system covers customer consulting, ordering and overall analysing of customer behaviours. To support the system, AsIsKnown knowledge flow requires two domain ontologies which are Home Textile Ontology and Multimedia Ontology which will be then mapped to the DOLCE ontology. The mapping process is done manually before evaluations are made by the experts who are partners of the project.

Small Enterprises Accessing the Electronic Market of the Enlarged Europe by a Smart Service Infrastructure (SEAMLESS)[2] is another project which is based in Europe. It is designed to improve communication & collaboration processes between companies belonging to the textile sector. To enable communication and collaboration, the global ontology of the textile sector (TEX GLOB) is constructed by extending the SEAMLESS core ontology. The result was mapped with global ontology of the building and construction sector (B&C GLOB). Both global ontologies are relevant to the identified population of Craft & Trade companies which present some overlapping areas (geotextiles) and are also critical for the European industry (EU Technology Platforms in both sectors).

The project on Andean Weaving was initiated with the goal to construct a knowledge base that collects and stores 3D Andean textiles patterns and preserve the rich and historical information about the subject. Andean textile heritage knowledge model extended the CIDOC CRM as a reference model to illustrate not only the images but also the weaving techniques and the productive processes. At the same time, it allows weavers to document and protect their rights on the designs. The project relied heavily on the human experts such as ethnographer-linguists, archaeologists, museum curators, weavers, and computer scientist to make it happen (Arnold et al., 2008).

Apart from that, there are many cultural heritage web information systems developed on the Semantic Web. The SCULPTUER project handles museum multimedia collections by mapping the museum's partner legacy system with CIDOC-CRM for cross collection searching (Sinclair et al., 2005). The Archive Mapper for Archaeology (AMA) project aims to create tools for semi-automated mapping from archaeological archive materials, reports, catalogues and databases to CIDOC-CRM (Eide et al., 2008). The English Heritage Centre for Archaeology Ontological Model (CRM-EH) is an extension of the CIDOC-CRM which aims for effective search across multiple different databases and their associated controlled vocabularies (Binding et al., 2008). The project on building global ontology for distributed digital museums employs CIDOC-CRM to identify and classify the semantics of data derived from local museums (Liu, 2007). All these projects utilize CIDOC-CRM as a common standard either as a global or extensible model.

Moreover, harmonization works are also carried out between CIDOC CRM and other ontologies such as ABC Model (Lagoze & Hunter, 2001), MPEG-7(Hunter, 2002), Functional Requirements for Bibliographic Records (Doerr & LeBoeuf, 2007) and Dublin Core (Kakali et al., 2007). ABC ontology is a model for the exchange and integration of digital library information. The combining effort between MPEG-7 and CIDOC CRM metadata models resulted in the creation of a standardized model for describing and managing museum multimedia content. FRBR is a formal ontology designed to capture and represent the underlying semantics of bibliographic information. Dublin Core is well accepted and widely used by all digital libraries. All these projects aim to create a single ontology that represents the conceptualization of reality in the domain area.

3.0 METHODOLOGY

3.1 Aim

The aim of the study is to provide commonality of content between revised TMT Knowledge Model and CIDOC CRM via common conceptualisation. This is done to attain better descriptions and understanding of their concepts through shared common entities. The mapped model will be a result of complementary between both models to achieve broader coverage especially pertaining to batik-related information.

3.2 Data Description

This study will use two sets of data sources for mapping task as described below:

[1] www.AsIsKnown.org

[2] www.seamless-eu.org

3.2.1 Revised TMT Knowledge Model

Originally, TMT Knowledge Model is created which solely focused on the historical factors and limited to the description of textile from the Malay Peninsula (Suriyati et al., 2009). The work was motivated to capture and represent the underlying semantic of Malay textile information. By reviewing the previous harmonization works with CIDOC CRM as exemplars, TMT Knowledge Model has been redesigned to reach a common view with CIDOC CRM. All facets in the model are re-created by transforming into RDFS classes and properties. The model is extended to further classify artefacts to capture the details of the textile-making in terms of techniques and productive process. As a result, it encompasses 21 classes and 15 properties as shown in Figure 1 (Syerina et al., 2010).

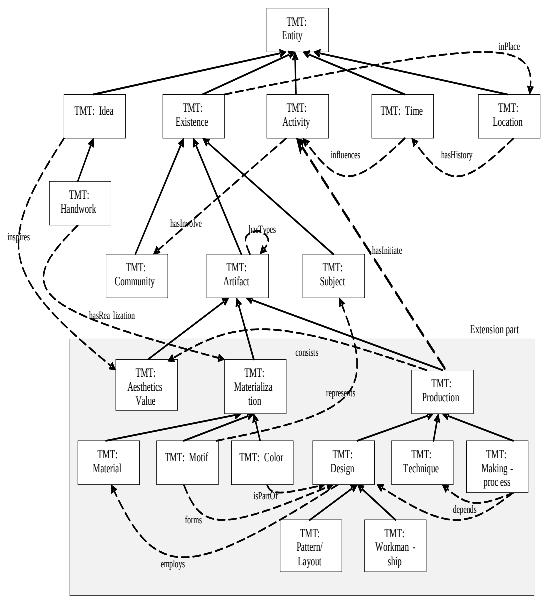


Figure 1: TMT Classes and Properties

3.2.2 CIDOC CRM

The work was initiated by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM) since 1996 and further developed until it has been accepted as ISO standard (ISO21127) in September 2006 with 90 classes and 148 properties, representing the semantics of hundreds of schemata. The "CIDOC object-oriented Conceptual Reference Model" (CRM) is a core ontology for information exchange in the cultural heritage and museum community. It provides the semantic connection in creating a high quality global resource due to the need of transforming distinct and individualised information sources. The aim of this initiative is to provide a formal ontology intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation, and interchange of bibliographic and museum information (Bekiari et al., 2008; Boeuf, 2003; Doerr et al., 2007).

3.3 Mapping Process

The work by Doerr et al. (2003b) has inspired the study. In this study, the mapping was done manually between TMT-Knowledge Model and CIDOC CRM. Basically, there are two main steps that will be performed to develop the Traditional Malay Textile Ontology (TMTO). Firstly, since both ontologies are in the form of documentation, therefore, this research will adapt the work by McGuiness et al. (2000). The documents will be reviewed and compared in order to find similar or overlapping concepts. This is to ensure both ontologies are aligned to each other in order to achieve ontology similarity. At the same time, preserving the meaning of each concept will be the heart of the process. This leads to the next step which is to relate concepts through the terms and its meaning between both ontologies, TMT Knowledge Model and CIDOC CRM. Then, the output will be checked for any mismatches or uncertainties during the course. If any, the process will be repeated until the desired outcomes are obtained. This task is demanding due to the specification of CIDOC CRM which is composed of 80 classes and 132 relationships[3].

Lastly, the process requires validation by an expert in the area such as curators from Muzium Negara and professionals on CIDOC. The whole process will be repeated if there are any changes identified during this process as shown in Figure 2 below. The resulting ontology will be known as Batik Heritage Ontology.

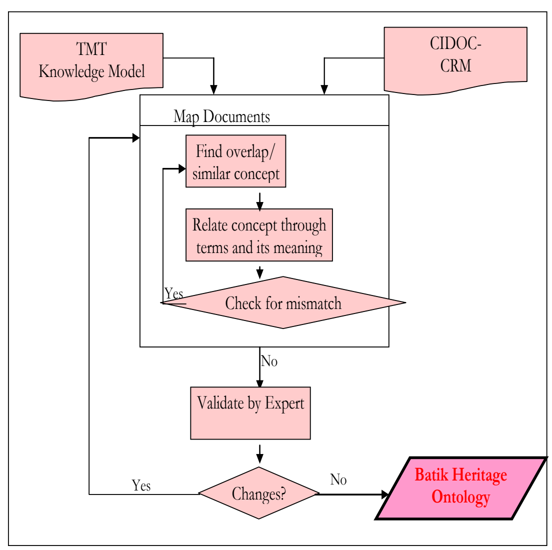


Figure 2: Flowchart of the manual mapping process

[3]http://cidoc.ics.forth.gr/docs/cidoc_crm_version_5.0.1_Nov09.pdf

4.0 PRELIMINARY RESULTS AND DISCUSSIONS

This section presents the result of manual mapping and merging between two models. Table 1 shows the harmony between two models on the agreement on class definition:

Table 1: The agreement between both models

TMT:Entity	is equivalent to	CRM:Entity
TMT:Activity	is equivalent to	CRM:Activity
TMT:Time	is equivalent to	CRM:Time-Span
TMT:Artifact	is equivalent to	CRM:Man-Made Thing
TMT:Production	is equivalent to	CRM:Production
TMT:Material	is equivalent to	CRM:Material
TMT:Design	is equivalent to	TMT:Design or Procedure

The remainder of the classes were compared and combined between both models and revealed the following outcomes (Table 2).

Table 2: The agreement based on common view

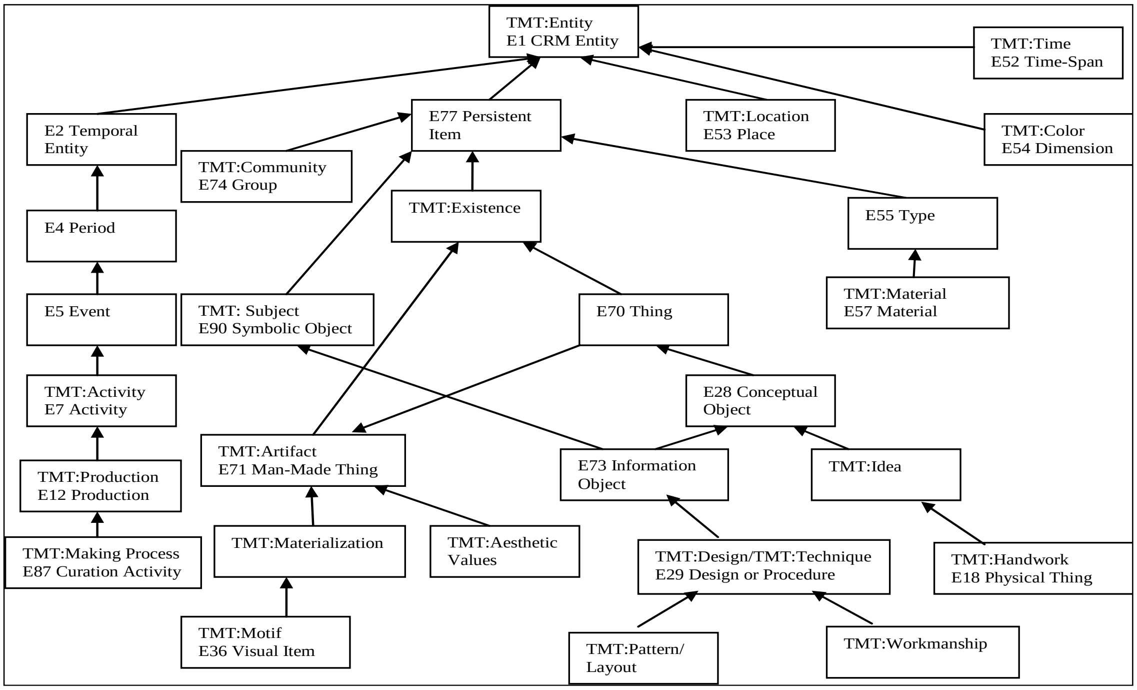
TMT:Location	is similar	CRM:Place
TMT:Community	is similar	CRM:Group
TMT:Existence	is equivalent	CRM:Persistent Item
TMT:Idea	is subclass of	CRM:Conceptual Object
TMT:Handwork	is equivalent to	CRM: Physical Man-Made Thing
TMT:Materialization	is subclass of	CRM:Information Object
TMT:Making-process	is equivalent to	CRM:Curation Activity
TMT:Motif	is equivalent to	CRM:Visual Item
TMT:Subject	is equivalent to	CRM:Symbolic Object
TMT:Color	is equivalent to	CRM:Dimension
TMT:Aesthetic Values	is subclass of	CRM:Man-Made Thing
TMT:Production	is equivalent to	CRM:Production
TMT:Design	is similar	CRM:Design or Procedure
TMT:Technique	is equivalent to	CRM:Design or Procedure
TMT:Pattern/Layout	is subclass of	CRM:Design or Procedure
TMT:Workmanship	is subclass of	CRM:Design or Procedure

This process is conducted based on the common understanding of the representation of classes and an acknowledgement of semantic consistency.

Figure 3 above illustrates the result of automated mapping between TMT Knowledge Model and CIDOC CRM. There are several limitations identified during the mapping process as described below:

- Some of the classes were present in TMT Knowledge Model but missing from CIDOC CRM; like TMT: Materialization, TMT:Aesthetic Values, TMT:Idea, TMT:Pattern/Layout and TMT: Workmanship. However, this is expected since all these concepts are extension parts of TMT Knowledge Model.
- This is a groundwork that needs further testing, evaluation and refinement by using the real collections of cultural artifacts within museums. Therefore, the first step is to adopt the OntoClean methodology (Oltramari et al., 2002) to detect missing clarity and rigidity of class definitions, to justify subsumption relations and to detect wrong subsumption declarations (Doerr et al., 2003b).

Despite all these, the result shows that by transforming all classes into TMT Knowledge Model beforehand helps to generate automated mapping with CIDOC CRM. This is due to the common entities and enhanced version which are aligned to the standard ontology. The result of the mapping is an instant view of the classes which are found to be easily mapped between both models. Hence, this approach is not in agreement with some of the claims made that manual mapping is laborious, time consuming, error prone, difficult to maintain and update (Jiayi et al., 2008; Noy & Musen, 2001).



5.0 CONCLUDING REMARKS

This work is an attempt to provide a basis for ontology developers as a way of transforming unstructured information into a format that machines could understand especially when it comes to align with any standard ontology. In addition, the mapping process itself raises several issues of its own. However, it is found that the textual explanations and formal definitions of the concepts on both models were very useful to provide insights for mapping task. Therefore, the next step is to use the mapping result in constructing Batik Heritage Ontology in order to preserve batik-related information.

REFERENCES

Aleksovski, Z. (2008). *Using background knowledge in ontology matching.* Ph.D Thesis, Vrije Universiteit Amsterdam.

Arnold, D. Y., Helmer, S., & Arando, R. V. (2008). *Towards Building a Knowledge Base for Research on Andean Weaving.* Research Technical Reports: Birkbeck, University of London.

Bekiari, C., Doerr, M., Le Boeuf, P., Aalberg, T., Barthélémy, J., Boutard, G., et al. (2008). International Working Group on FRBR and CIDOC CRM Harmonisation. *FRBR Object-Oriented Definition and Mapping to FRBR ER.(version 0.9 draft)*.

Benjamin, P. C., Menzel, C. P., Mayer, R. J., Fillion, F., Futrell, M. T., deWitte, P. S., et al. (1994). IDEF5 Method Report. *Knowledge Based Systems, Inc*.

Binding, C., May, K., & Tudhope, D. (2008). *Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction Via the CIDOC CRM* (pp. 280-290): Springer.

Doan, A. H., Madhavan, J., Domingos, P., & Halevy, A. (2002). *Learning to map between ontologies on the semantic web*.

Doan, A. H., Madhavan, J., Domingos, P., & Halevy, A. (2004a). Ontology matching: A machine learning approach (pp. 397-416): Springer.

Doan, A. H., Madhavan, J., Domingos, P., & Halevy, A. (2004b). *Ontology matching: A machine learning approach* (pp. 397-416): Springer.

Doerr, M., Hunter, J., & Lagoze, C. (2003b). Towards a Core Ontology for Information Integration. *Journal of Digital Information, 4*(1), 169.

Doerr, M., & LeBoeuf, P. (2007). Modelling intellectual processes: the FRBR-CRM harmonization. *LECTURE NOTES IN COMPUTER SCIENCE, 4877*, 114.

Eide, O., Felicetti, A., Ore, C. E., Andrea, A. D., & Holmen, J. (2008). Encoding Cultural Heritage Information for the Semantic Web. Procedures for Data Integration through CIDOC-CRM Mapping. *Paper presented at the EPOCH*

Conference on Open Digital Cultural Heritage Systems.

Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic Maps! Making sense of it all. *Journal of Information Science, 30*(4), 378-391.

Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199-220.

Guarino, N. (1998a). Formal Ontology and Information Systems. *Proceedings of FOIS, 98*(1998), pp. 81-97.

Hunter, J. (2002). Combining the CIDOC CRM and MPEG-7 to Describe Multimedia in Museums. *Museums on the Web, Boston, April*.

Jiayi, P., Cheng, C. P. J., Lau, G. T., & Law, K. H. (2008). *Utilizing Statistical Semantic Similarity Techniques for Ontology Mapping—with Applications to AEC Standard Models* (Vol. 13, pp. 217-222): Elsevier.

Kakali, C., Lourdi, I., Stasinopoulou, T., Bountouri, L., Papatheodorou, C., Doerr, M., et al. (2007). *Integrating Dublin Core metadata for cultural heritage collections using ontologies*.

Lagoze, C., & Hunter, J. (2001). The ABC Ontology and Model. *Journal of Digital Information, 2*(2), 2001-2011.

Liu, H.-Z. (2007). Global Ontology Construction for Heterogeneous Digital Museums. *Paper presented at the Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong*.

McGuinness, D. L., Fikes, R., Rice, J., & Wilder, S. (2000). *The Chimaera ontology environment*.

Mika, P. (2007). Ontologies are us: A unified model of social networks and semantics. *Web Semantics: Science, Services and Agents on the World Wide Web, 5*(1), 5-15.

Noy, N. F. (2004). Semantic integration: a survey of ontology-based approaches. *ACM SIGMOD Record, 33*(4), 65-70.

Noy, N. F., & McGuinness, D. L. (2005). Ontology Development 101: A Guide to Creating Your First Ontology.

Noy, N. F., & Musen, M. A. (2001). *Anchor-PROMPT: Using non-local context for semantic matching*.

Oltramari, A., Gangemi, A., Guarino, N., & Masolo, C. (2002). Restructuring WordNet's top-level: The OntoClean approach. *LREC2002, Las Palmas, Spain*.

Sabou, M., d'Aquin, M., & Motta, E. (2006). Using the Semantic Web as Background Knowledge for Ontology Mapping. *International Workshop on Ontology Matching, Athens, GA.*

Shvaiko, P., & Euzenat, J. (2008). *Ten challenges for ontology matching*.

Sinclair, P., Goodall, S., Lewis, P. H., Martinez, K., & Addis, M. J. (2005). Concept browsing for Multimedia Retrieval in the SCULPTUER Project. *Paper presented at the 2nd European Semantic Web Conference.*

Suriyati, R., Norlaila, M. N., & Wan Adilah, W. A. (2009). Towards Sustainable Heritage Information System: Conceptualization of Community based e-Museum. *Paper presented at the International Conference on e-Commerce, e-Administration, e-Society and e-Communication (e-CASE).*

Syerina, A. M. N., Nor Laila, M. N., & Suriyati, R. (2010). The Traditional Malay Textile (TMT) Knowledge Model: Transformation towards Automated Mapping. *Paper presented at the International Conference on Knowledge Management* (to be appear).