

Automating the Mapping Process of Traditional Malay Textile Knowledge Model with the Core Ontology

Syerina Azlin Md Nasir and Nor Laila Md Noor

Department of System Science, University Technology MARA, 40000, Shah Alam, Malaysia

Abstract: Problem statement: 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. **Approach:** Manual mapping was conducted to find similar or overlapping concepts which are aligned to each other in order to achieve ontology similarity. This is achieved after TMT Knowledge Model already undergone transformation process to match with CIDOC CRM structure. **Results:** 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. **Conclusion/Recommendations:** Future research will be focused on the construction of Batik Heritage Ontology by using the mapping result obtained in this study. Further testing, evaluation and refinement by using the real collections of cultural artifacts within museums will also be conducted in the near future.

Key words: Automated mapping, common view, cultural heritage, core ontology

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; Mika, 2007; Noy and McGuinness, 2001; McGuinness *et al.*, 2000). As such, definitions are simplified and described as a set of defined vocabulary with a meaning that constrains the vocabulary to describe concepts (Doan *et al.*, 2004; Noy, 2004; Shvaiko and 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. The mutual agreement on the terms depicted assists the process of integration, interoperability, knowledge sharing and reuse. For this reason, ontology's offer the conceptual foundation for allowing the semantics of metadata machine-understandable.

In textile domain, there are a number of available ontology's 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)

(www.AsIsKnown.org) is created which aims to collect the product data from diverse textile producers in one system. The system covers customer consulting, ordering and overall analyzing of customer behaviors. To support the system, AsIsKnown knowledge flow requires two domain ontology's 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) (www.seamless-eu.org) is another project which is based in Europe. It is designed to improve communication and 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 and C GLOB). Both global ontology's are relevant to the identified population of Craft and Trade companies which present some overlapping areas (geotextiles) and are also critical for the European industry (EU Technology Platforms in both sectors).

Corresponding Author: Syerina Azlin Md Nasir, Department of System Science, University Technology MARA (UiTM), 40000 Shah Alam, Selangor, Malaysia

Other than that, 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.*, 2009).

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 ontology's such as ABC Model (Lagoze and Hunter, 2001), MPEG-7 (Hunter, 2002), Functional Requirements for Bibliographic Records (Doerr and 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.

The growing number of ontology's has created predicament where the disseminated nature of ontology development has led to disparate ontology's 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 to unify and enrich the body of knowledge. 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 ontology's. Among identified problems is when ontology's are different in terms of context and background knowledge, it will bring to failure in discovering some correct mappings (Aleksovski *et al.*, 2006; Sabou *et al.*, 2006). Nonetheless, when tools are used for mapping two schemas or even ontology's, there is high possibility of missing information because not all concepts are mapped between them (Doan *et al.*, 2002). In another aspect, the computer cannot make decisions and possess vocabulary understanding like a human can (Lilac and Al-Abdullatif, 2010).

This study 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 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 study 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 study is to provide a base for common views in automating the process of mapping between TMT Knowledge Model and CIDOC CRM.

MATERIALS AND METHODS

Aim: The aim of the study is to provide commonality of content between revised TMT Knowledge Model and CIDOC CRM via common conceptualization. This

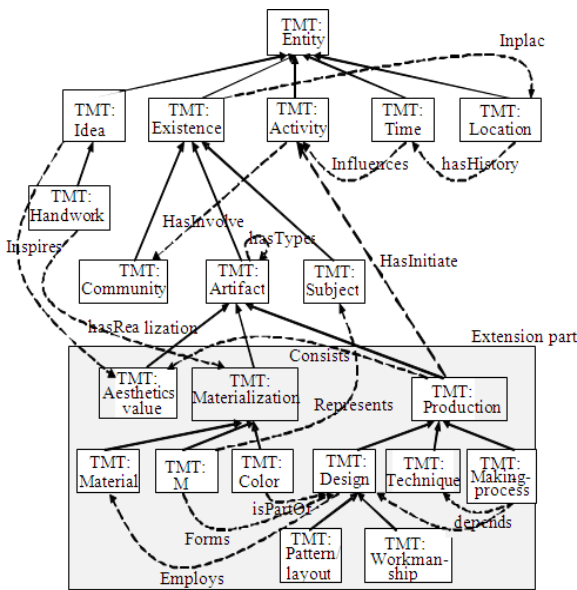


Fig. 1: TMT classes and properties

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.

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

- 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. The study 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 Fig. 1 (Nasir and Noor, 2010)
- CIDOC CRM: The study was initiated by the International Committee for Documentation (CIDOC) of the International Council of Museums

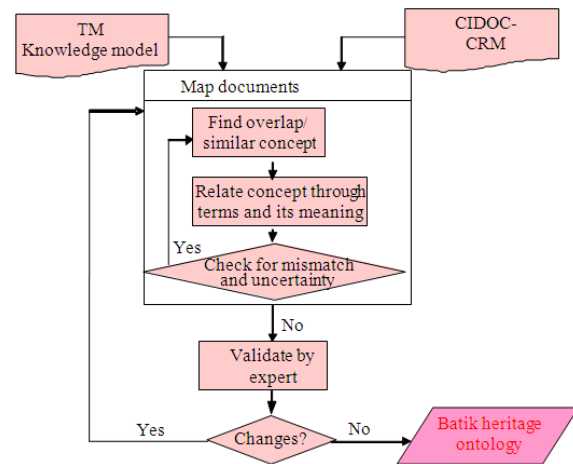


Fig. 2: Flowchart of the manual mapping process

(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 individualized 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 (Boeuf, 2003; Doerr and LeBoeuf, 2007)

Mapping process: The study by Doerr *et al.* (2003) 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 ontology’s are in the form of documentation, therefore, this research will adapt the study by McGuinness *et al.* (2000). The documents will be reviewed and compared in order to find similar or overlapping concepts. This is to ensure both ontology’s 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 ontology's, 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 (http://cidoc.ics.forth.gr/docs/cidoc_crm_version_5.0.1_Nov09.pdf).

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

RESULTS

This section presents the result of automated 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

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: Colour	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

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

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

DISCUSSION

Figure 3 below illustrates the outcome 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 Onto Clean methodology (Oltamari *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.*, 2003)

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 and Musen, 2001).

CONCLUSION

This study 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. It is found out that human interventions in the mapping process is crucial in analyzing the mapping results and understand the characteristics of the source ontology's. Therefore, the next step is to use the mapping result in constructing Batik Heritage Ontology in order to preserve batik-related information.

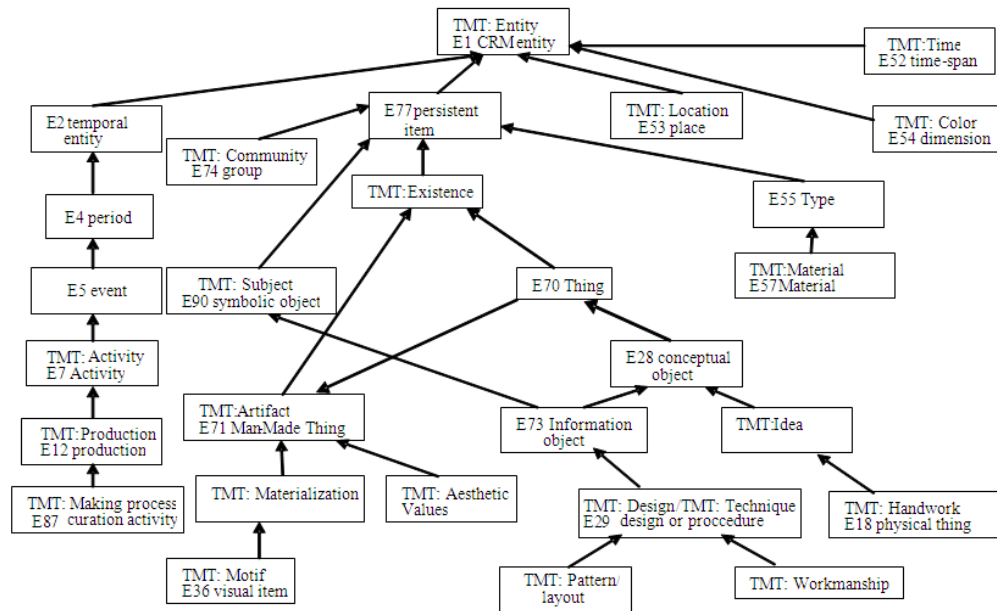


Fig. 3: The mapped and merged TMT and CIDOC CRM class hierarchies

REFERENCES

- Aleksovski, Z., M. Klein, W. Kate and F.V. Harmelen, 2006. Matching unstructured vocabularies using a background ontology. *Manag. Know. World Networks*, 4248: 182-187. DOI: 10.1007/11891451_18
- Arnold, D.Y., S. Helmer and R.V. Arando, 2009. Towards building a knowledge base for research on andean weaving. *Dataspace: Final Frontier*, 5588: 180-188. DOI: 10.1007/978-3-642-02843-4_18
- Binding, C., K. May and D. Tudhope, 2008. Semantic interoperability in archaeological datasets: Data mapping and extraction via the CIDOC CRM. *Res. Adv. Technol. Digital Libraries*, 5173: 280-290. DOI: 10.1007/978-3-540-87599-4_30
- Doan, A.H., J. Madhavan, P. Domingos and A. Halevy, 2002. Learning to map between ontologies on the semantic web. *Proceeding of the 11th International Conference on World Wide Web, (WWW'02)*, ACM New York, NY, USA., pp: 662-673. DOI: 10.1145/511446.511532
- Doan, A.H., J. Madhavan, P. Domingos and A. Halevy, 2004. Ontology matching: A machine learning approach. In: *Handbook on Ontologies in Information Systems*, Staab, S. and R. Studer (Eds.). Springer-Verlag, pp: 397-416.
- Doerr, M. and P. LeBoeuf, 2007. Modelling intellectual processes: the FRBR - CRM harmonization. *Digital Libraries: Res. Dev.*, 4877: 114-123. DOI: 10.1007/978-3-540-77088-6_11
- Doerr, M., J. Hunter and C. Lagoze, 2003. Towards a core ontology for information integration. *J. Digital Inform.*, <http://portal.acm.org/citation.cfm?id=1358619>
- Eide, O., A. Felicetti, C.E. Ore, A.D. Andrea and J. Holmen, 2008. Encoding cultural heritage information for the semantic web. *Proceeding of the Procedures for Data Integration through CIDOC-CRM Mapping, EPOCH Conference on Open Digital Cultural Heritage*, pp: 1-7.
- Garshol, L.M., 2004. Metadata? Thesauri? Taxonomies? Topic maps! making sense of it all. *J. Inform. Sci.*, 30: 378-391. DOI: 10.1177/0165551504045856
- Gruber, T., 1993. A translation approach to portable ontology specifications. *Know. Acquisition*, 5: 199-220. DOI: 10.1006/knac.1993.1008
- Hunter, J., 2002. Combining the CIDOC CRM and MPEG-7 to describe multimedia in museums. *Proceeding of the International Conference on Museums and the Web, Apr. 17-20, Boston, MA.*, pp: 16-16.
- Jiayi, P., C.P.J. Cheng, G.T. Lau and K.H. Law, 2008. Utilizing statistical semantic similarity techniques for ontology mapping-with applications to AEC standard models. *Tsinghua Sci. Technol.*, 13: 217-222. DOI: 10.1016/S1007-0214(08)70152-4
- Kakali, C., I. Lourdi, T. Stasinopoulou, L. Bountouri and C. Papatheodorou *et al.*, 2007. Integrating Dublin Core metadata for cultural heritage collections using ontologies. *Proceeding of the International Conference on Dublin Core and Metadata Applications, (DCMI'07)*, Dublin Core Metadata Initiative, United States, pp: 128-139.

- Lagoze, C. and J. Hunter, 2001. The ABC ontology and Proceeding of the International Conference on Dublin Core and Metadata Applications, Oct. 24-26, National Institute of Informatics, Tokyo, Japan, pp: 160-176.
- Lilac, A.E.A.S. and N.A.O. Al-Abdullatif, 2010. Educational advertising ontology: a domain-dependent ontology for semantic advertising networks. *J. Comput. Sci.*, 6: 1070-1077. DOI: 10.3844/jcssp.2010.1070.1077
- Liu, H.Z., 2007. Global ontology construction for heterogeneous digital museums. Proceeding of the 6th International Conference on Machine Learning and Cybernetics, Aug. 19-22, Hong Kong, pp: 4015-4019. DOI: 10.1109/ICMLC.2007.4370848
- McGuinness, D.L., R. Fikes, J. Rice and S. Wilder, 2000. The Chimaera ontology environment. Proceeding of the 17th National Conference on Artificial Intelligence, Aug. 3-3, Austin, Texas, pp: 1123-1124.
- Mika, P., 2007. Ontologies are us: A unified model of social networks and semantics. *Web Semantics: Sci. Services Agents World Wide Web*, 5: 5-15. DOI: 10.1016/j.websem.2006.11.002
- Nasir, S.A.M. and N.L.M. Noor, 2010. Analysing the effectiveness of COMA++ on the mapping between Traditional Malay Textile (TMT) knowledge model and CIDOC CRM. Proceeding of the International Symposium in Information Technology (ITSim), June 15-17, Kuala Lumpur, pp: 1-6. DOI: 10.1109/ITSIM.2010.5561323
- Noy, N.F. and D.L. McGuinness, 2001. Ontology development 101: A guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report. <http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html>
- Noy, N.F. and M.A. Musen, 2001. Anchor-PROMPT: Using non-local context for semantic matching. Proceeding of the workshop on Ontologies and Information Sharing at the International Joint Conference on Artificial Intelligence, Aug. 4-5, Stanford University, Stanford, CA., pp: 63-70.
- Noy, N.F., 2004. Semantic integration: A survey of ontology-based approaches. *ACM SIGMOD Record*, 33: 65-70. DOI: 10.1145/1041410.1041421
- Oltramari, A., A. Gangemi, N. Guarino and C. Masolo, 2002. Restructuring WordNet's top-level: The OntoClean approach. Proceeding of the International LREC2002 (OntoLex workshop), (LREC'02), Las Palmas, Spain, pp: 17-26.
- Sabou, M., M. d'Aquin and E. Motta, 2006. Using the semantic web as background knowledge for ontology mapping. Proceeding of the International Workshop on Ontology Matching, Athens, (IWOMA'06), <http://eprints.aktors.org/571/>
- Shvaiko, P. and J. Euzenat, 2008. Ten challenges for ontology matching. *On the Move to Meaningful Internet Systems: OTM.*, 5332: 1164-1182. DOI: 10.1007/978-3-540-88873-4_18
- Sinclair, P., S. Goodall, P.H. Lewis, K. Martinez and M.J. Addis, 2005. Concept browsing for multimedia retrieval in the SCULPTEUR project. Proceedings of the 2nd European Semantic Web Conference, May 29-1 June, Heraklion, Crete, <http://eprints.ecs.soton.ac.uk/10913/>