

Developing and Validating the Socio-Technical Model in Ontology Engineering

Mesnan Silalahi^{ab1}, Dana Indra Sensuse^{a2}, Yudho Giri Sucahyo^{a3}, Izzah Fadhilah Akmaliah^{a4}, Puji Rahayu^{a6}, Elin Cahyaningsih^{a7}

^ae-Government and e-Business Research Lab, Faculty of Computer Science, University of Indonesia

^bResearch Center for S&T Development, Indonesian Institute of Science, Jakarta, Indonesia

Abstract. This paper describes results from an attempt to develop a model in ontology engineering methodology and a way to validate the model. The approach to methodology in ontology engineering is from the point view of socio-technical system theory. Qualitative research synthesis is used to build the model using meta-ethnography. In order to ensure the objectivity of the measurement, inter-rater reliability method was applied using a multi-rater Fleiss Kappa. The results show the accordance of the research output with the diamond model in the socio-technical system theory by evidence of the interdependency of the four socio-technical variables namely people, technology, structure and task.

Keywords: Socio-Technical, Ontology Engineering, Model, Validation, Meta-Ethnography

1. Introduction

Building ontology is initially regarded as a craft rather than an engineering activity [17]. Some of the methodologies developed in the engineering of ontology were initially developed based on the experiences in building ontology whereby mostly based on the principle of common senses such as OTK methodology [19]. The experiences were organized and structured systematically to become a guideline for ontology engineering (OE). Several other methodologies are developed based on the theory that exist in a particular context or an approach with specific orientation, such as HCOME [18], UPON [15], Onto-Eng.[1], and DILIGENT [16]. The developed methodology partially focuses on either technical approach or social and human approach. There is little known about methodology which are technically focused while still considering the social aspect equally in the approach. This study bridges this gap by building a methodology that equally treat the technical and social aspect in ontology engineering, based on a systematic review of the literature in ontology engineering and an empirical investigation in case studies in the development of an ontology. The review includes analyses on methodologies in building ontology, the tools, the environments and the organization, and the study of principles of design from the perspective of socio-technical system. Based on experiences in building ontology a systematic design method is compiled and proposed in ontology engineering methodology with socio-technical approach that provides a comprehensive guideline to building an ontology. This study has two research



questions that is “how to build a methodology in OE based on social-technical approach and how to validate the OE model”.

2. The Socio-Technical Approach in Ontology Engineering

Ref. [7] formulated four interdependent variables in the diamond model of socio-technical system namely: task, people, technology, and structure as shown in Fig. 1 by the arrows, so that changes to one variable will affect other variables.

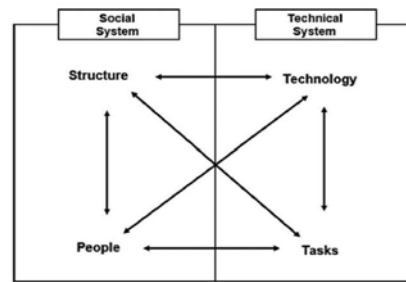


Figure 1. A diamond model in a socio-technical system

Technology is regarded as a tool that assists organizations in completing the work through the mechanisms of transformation from inputs into outputs. The socio-technical ontology engineering perspective would require all activities that support social subsystems (a human capital with its competencies and preferences in a social process) as well as the technical subsystem (engineering functions such as technology and task). A socio-technical approach in ontology engineering has been proposed partially in several other methodologies. The ways of approaching the ontology engineering socio-technically are such as involving domain experts and/or end-users as the central actor, such as in DILIGENT [14], HCOME [6], and HOLSAPPLE [4]. These methodologies proposes to begin the development process not from scratch (using a seed model), as well as proposing to build an ontology by reusing existing ontologies. The context of decision-making in collaborative work is found in [5] which proposed a process through a consensus mechanism. Ref. [9] uses a strategy in building a consensus in ontology engineering in the context of collaborative networks. Ref [13] gives a detailed description of how to build this consensus in NEON methodology as part of the specification and conceptualization process. However, their approach is more a technological support rather than a social approach as it focused on using wiki technology in the negotiation on a higher level. In our study a socio-technical approach to ontology engineering (STOE) methodology is developed. The social aspect is considered equally with the technical aspects in all phases of the ontology development starting from a planning phase until the evaluation phase. Some of the advantages gained by this approach is the knowledge sharing that can be encouraged as early as possible in the development phases, a best practice proposed in [4].

3. Methodology

In our study the initial methodological model was first proposed based on a comparative analysis and by conducting a critical review on several relevant methodologies in ontology engineering. Next the proposed model is triangulated through a meta-ethnography study. The initial model development of the methodology was performed by synthesizing qualitative research data from literatures as identified, searched and selected. The method used in the synthesis of qualitative research publication data in ontology engineering is meta-ethnography. Meta-ethnography as proposed by [8] has seven steps as follow [2]: (1) getting started, (2) deciding what is relevant to the initial interest, defining the focus of the synthesis; locating relevant studies; making decisions on inclusion; and quality assessment, (3)

reading the studies, (4) determining how the studies are related, (5) translating the studies into one another, (6) synthesizing translations, and (7) expressing the synthesis.

We investigate the selected literature by doing translation, comparing, analysing and synthesizing. Inductively identification of the key concepts and phrases of words is carried out through thinking deeply and rereading of the text carefully. The emergence frequency of translated words to concept in the dataset is a proxy to the conceptual model. The result from each translation and the synthesis is validated internally using Kappa method through member check which is aimed to avoid subjectivity in the study. The inter-rater reliability (IRR) is used as a measure for the validity of the interpretation. To understand whether observers should have the same opinion, Kappa coefficient K is used as inter-observer reliability estimation which is defined as:

$$K = \frac{P_a - P_e}{1 - P_e} \quad (1)$$

whereby P_a denotes the observed percentage of agreement, and P_e denotes the probability of expected agreement due to chance. The factor $(1 - P_e)$ gives the degree of agreement that is achievable above chance, and $(P_a - P_e)$ gives the degree of agreement actually achieved above chance. If the raters are in complete agreement then $K = 1$. Testing using Kappa method is carried out carefully with respect to prevalence and bias that may occur from the observations. IRR were obtained using three observers, which is the minimum number allowed in Fleiss Kappa method (multi-rater).

4. Results

Deductively and structurally, concepts (categories) were acquired in socio-technical system based on the model proposed by [7], which consists of four variables: people, technology, tasks and structure. Using these four variables then we do the synthesis (inductive) from the literature to get the concepts or themes (metaphors). Table 1 shows the results from the synthesis process.

Table 1. An excerpt from the results of the synthesis processes

Findings / Evidences	Themes / Metaphors	Concept	Category	Direct Affected Category	Phase
<i>"The users emphasized the need for local control over their ontologies"</i>	Local control	Work procedures	Structure	People	Analysis, Design
<i>"They asked explicitly for a system without a central server, where knowledge sharing was integrated into the normal work, but where different kinds of information, like files, emails, bookmarks and addresses could be shared with others"</i>	Collaboration integrated into normal work	Distributed Collaborative Work	Structure	Technology	Planning
<i>"they commented, that they would share more files at a later stage, when they would feel more confident with the system"</i>	Need confidence in sharing information	Self-efficacy	People	Task	Analysis, Design
<i>"The new users behaved in a similar way as the users in the first stage and did not share many folders, as they wanted to gain confidence in the system first"</i>	New user first of all, gain confidences	Self-efficacy	People	Task	Analysis, Design

The findings showed that none of the socio-technical variables (concepts) show up directly in the frequency count, but most show up indirectly which we called themes or metaphors (e.g., support tools). However, when we explore the keyword frequency using the concepts from our classification model we find that all of the socio-technical variables in our model are represented in the literatures to varying degrees, with the technology variable being the most represented, followed by people then structure. From the results of the synthesis, a cross tabulation is carried out into category along with its themes of the phases identified. The cross-tabulation data analyses shows that there exist socio-technical aspects in all phases of the ontology engineering. The results showed the interdependency between the technical aspects (tasks and technology) and the social aspects (people and structure) in all phases of the ontology

engineering process, in accordance with the socio-technical model proposed by [7]. These findings are used as evidence to confirm Leavitt's theory in the context of socio-technical system. These findings also validate the initial model proposed in ontology engineering methodology.

Finally, we took a completely inductive approach, to determine other concepts in the context of life-cycle model which were synthesized from the dataset to confirm the life-cycle conceptual model of the initial socio-technical model. In the phase frequency distribution, we confirm the emergence of phase such as “planning,” “analysis,” “design,” “implementation,” and “evaluation” which confirms the proposed life-cycle model. The life-cycle model proposed is slightly different to the Waterfall model by [10] and [12].

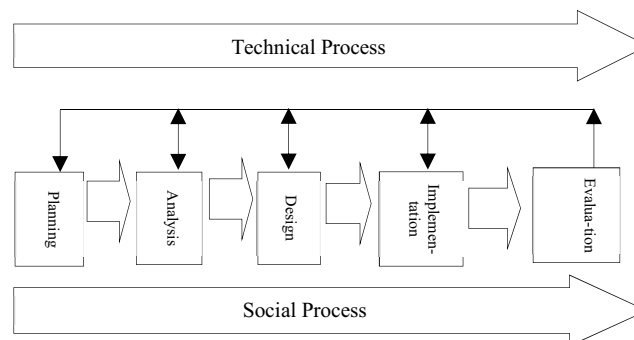


Figure 2. A Life-cycle Model in a Socio-technical Ontology Engineering

Validating the Model

Using a Cohen Kappa we initiated the calculation and found that the normal proportion of the total rating is 95% so that in this case there is a prevalence in the distribution according to the rating category. Table 2. below show an excerpt from the calculation.

Table 2. Kappa coefficient calculation uses two rater data samples

		Rater 1		Row Margin	
		normal	abnormal		
Rater 2	normal	199	24	223	rm1
	abnormal	10	1	11	rm2
Column Margin		209	25	234	
		cm1	cm2		n

Therefore, there should some correction in the calculation of kappa to this prevalence. According [3] for the case of two rater the correction in Cohen Kappa, namely $PABAK = 2Pa - 1$, wherein: raw percent of agreement (Pa)= $200/232=86.2\%$, expected agreement (Pe)= 87.1 , prevalence index= $(199-1)/234=0.85$, $PABAK=0.72$. From the results of the calculation we obtained $PABAK$ 0.72 which means that the resulting synthesis is reliable and the data is substantially valid. In the cases of observations made by more than two people, Kappa formula to calculate Pa can be used according to Fleiss. While a correction to the value of Kappa is done by using a calculation in [11], Pe is calculated using the correction $1/k$, where k is the number of category. The results using multi-rater (three) are as follows: the validation is performed against every step and sequentially carried out starting from the translation in the first order, the translation into second order, then translation into category and the directly affected category, and lastly the synthesis into phase. Validation data in first-order translation, $Pa=0.876$, $Pe=0.889$, $Kappa=-0.121$, adjusted Kappa against prevalence= 0.752 . Within second order translation we found, $Pa=0.962$, $Pe=0.963$, $Kappa=-0.037$, adjusted Kappa against prevalence = 0.924 . Validation

of data in the synthesis into the first category: $P_a=0.876$, $P_e=0.889$, $Kappa=-0.12$, adjusted $Kappa=0.752$. Validation of data in the synthesis into second category: $P_a=0.880$, $P_e=0.886$, $Kappa=-0.058$, adjusted $Kappa=0.760$. Validation of data in the synthesis into phase: $P_a=0.963$, $P_e=0.963$, $Kappa=-0.019$, adjusted $Kappa$ against prevalence= 0.926 .

5. Conclusion and Future Researches

This study highlights the complex interdependent relationship between technology factors and social factors and shows the need to address the socio-technical aspect in the ontology engineering methodology in order to have a balanced treatment between the technical process and the social process along the whole ontology engineering phases. The results of the analysis and synthesis are initial model which will be validated in a case study of an ontology development. The results of the case study will provide another validation of the model. Findings from our qualitative research synthesis will be refined through a survey on participants' perception on the operationalization of the methodology through case studies to empirically test the methodological model of the socio-technical ontology engineering. Overall the model validation is carried out by triangulation. The triangulation across various methods of data collection is very useful because it reinforces the validity of the model as well as providing multiple perspectives and to provide data and information that is complementary. In this paper the given description in the process is still in outline form (high-level). Improvements to this methodology will be done iteratively through case studies in ontology development which are undertaken to get a detailed picture of sub-processes and sub-activities.

Acknowledgement

The author would like to thank to Head of E-Government and E-Business laboratory in University of Indonesia and University of Indonesia as a funder of this research.

References

- [1] Al-Debei M M, and Fitzgerald G 2009 OntoEng: a design method for ontology engineering in information systems *Proc. of the ACM OOPSLA'09 (ODiSE 2009)* pp 1-25.
- [2] Britten N, Campbell R, Pope C, Donovan J, Morgan M, and Pill R 2002 Using meta ethnography to synthesise qualitative research: a worked example *J. of Health Services Research and Policy* **7(4)** 209-15.
- [3] Byrt T, Bishop J, and Carlin J B 1993 Bias, prevalence and kappa *J. of Clinical Epidemiology* **46(5)** pp 423-9.
- [4] Herrmann T, Loser K U, and Jahnke I 2007 Sociotechnical walkthrough: a means for knowledge integration *The Learning Organization* **14(5)** pp 450-64..
- [5] Holsapple C W, and Joshi K D 2004 A formal knowledge management ontology: Conduct, activities, resources, and influences *J. of the Assoc. for Information Sci. and Technol.* **55(7)** pp 593-612.
- [6] Kotis K, and Vouros G A 2006 Human-centered ontology engineering: The HCOME methodology *Knowledge and Information Syst.* **10(1)** pp 109-31.
- [7] Leavitt H J 1965 Applying Organizational Change in Industry: Structural, Technological and Humanistic Approaches *J. G. March (Ed.) Handbook of Organizations* (Chicago, IL, USA: Rand McNally) pp 1144-70.
- [8] Noblit, George W, and Hare R Dwight 1988 *Meta-Ethnography: Synthesizing Qualitative Studies* Vol. 11 (Sage).
- [9] Pereira C, Sousa C, and Soares A L 2009 Short-term semantic consensus: towards agile ontology specification for collaborative networks *Working Conf. on Virtual Enterprises* (Berlin, Heidelberg: Springer) pp 301-10.
- [10] Pressman R S 2010 *Software Engineering: A Practitioner's Approach* Ed. 7 (McGraw-Hill Higher Education).
- [11] Randolph J J 2005 *Free-Marginal Multirater Kappa (multirater K [free]): An Alternative to*

- Fleiss' Fixed-Marginal Multirater Kappa* [Online Submission].
- [12] Sommerville I 2011 *Software engineering* Ed. 9 (USA: Addison-Wesley).
 - [13] Suárez-Figueroa M C, Gómez-Pérez A, and Fernández-López M 2012 The NeOn methodology for ontology engineering *Ontology Eng. in a Networked World* (Berlin,Heidelberg: Springer) pp 9-34.
 - [14] Tempich C, Pinto S, Staab S, and Sure Y 2004 A case study in supporting DIstributed, Loosely-controlled and evolvInG Engineering of oNTologies (DILIGENT) *Proc. of the 4th Int. Conf. on Knowledge Management (I-KNOW'04)* pp 225-32.
 - [15] De Nicola A, Missikoff M, and Navigli R 2005 A proposal for a unified process for ontology building: UPON *Database and Expert Systems Applications* (Berlin/Heidelberg: Springer) pp 655-64.
 - [16] Pinto H S, Staab S, and Tempich C 2004 DILIGENT: Towards a fine-grained methodology for DIstributed, Loosely-controlled and evolvInG Engineering of oNTologies *Proc. of the 16th European Conf. on Artificial Intelligence* (IOS Pres) pp 393-7.
 - [17] Fernández-López M, Gómez-Pérez A, and Juristo N 1997 *Methontology: from ontological art towards ontological engineering*.
 - [18] Kotis K, and Vouros G A 2006 Human-centered ontology engineering: The HCOME methodology *Knowledge and Information Syst.* **10(1)** pp 109-31.
 - [19] Sure Y, Staab S, and Studer R 2002 Methodology for development and employment of ontology based knowledge management applications *ACM Sigmod Record* **31(4)** pp 18-23.