How the Structuring of Domain Knowledge Helps Casual Process Modelers

Jakob Pinggera[1], Stefan Zugal[1], Barbara Weber[1], Dirk Fahland[2],
Matthias Weidlich[3], Jan Mendling[2], and Hajo A. Reijers[4]

[1] University of Innsbruck, Austria
`jakob.pinggera|stefan.zugal|barbara.weber@uibk.ac.at`
[2] Humboldt-Universität zu Berlin, Germany
`fahland@informatik.hu-berlin.de|jan.mendling@wiwi.hu-berlin.de`
[3] Hasso-Plattner-Institute, University of Potsdam, Germany
`matthias.weidlich@hpi.uni-potsdam.de`
[4] Eindhoven University of Technology, The Netherlands
`h.a.reijers@tue.nl`

Abstract. Modeling business processes has become a common activity in industry, but it is increasingly carried out by *non-experts*. This raises a challenge: How to ensure that the resulting process models are of sufficient quality? This paper contends that a prior structuring of domain knowledge, as found in informal specifications, will positively influence the act of process modeling in various measures of performance. This idea is tested and confirmed with a controlled experiment, which involved 83 master students in business administration and industrial engineering from Humboldt-Universität zu Berlin and Eindhoven University of Technology. In line with the reported findings, our recommendation is to explore ways to bring more structure in the specifications that are used as input for process modeling endeavors.

1 Introduction

Business process modeling is the task of creating an explicit, graphical model of a business process from internalized knowledge on that process. This type of conceptual modeling has recently received considerable attention in information systems engineering due to its increasing importance in practice [1]. Business process modeling typically involves two specific, associated roles. A *domain expert* concretizes domain knowledge into an informal description, which is abstracted into a formal model by a *system analyst*. This works well if the domain expert and the process modeler closely interact with each other, and if both the domain expert and the process modeler attain a high level of expertise. However, these two conditions are often not met. Increasingly, *casual* modelers are involved in process modeling initiatives, who are neither domain nor process modeling experts. Many organizations do not reserve the time or resources for iterative and consensus-seeking approaches. To illustrate, we are in contact with a financial services provider that employs over 400 business professionals of which only *two*

are skilled process modelers. Process modeling activities in this organization are often carried out by IT specialists with a low process modeling expertise. As a consequence, process modeling is driven by informal requirement specifications as provided by domain experts, and models are generated in a small number of cycles with little opportunity for feedback and interaction.

Requirements engineering points to the importance of structure in requirements specifications [2, 3]. Along these lines, our central idea is that *the quality of process models may be improved through providing inexperienced, casual process modelers with well-structured domain descriptions.* To investigate this contention we designed an experiment in which we *observed* and *measured* how a process modeler creates a formal process model from an informal requirements specification. We varied the level of content organization of an informal requirements specification (serving as proxy for differently skilled domain experts) and traced its impact on process model quality. The subjects in our experiment were 83 students who received only limited prior process modeling training.

The paper is structured as follows. Section 2 discusses the background of our research. Section 3 describes our experimental framework. Section 4 covers the execution and results of our experiment. Section 5 concludes the paper.

2 Background

For discussing factors that influence the creation of a formal process model, it is necessary to first reflect on *conceptual modeling* in general. Conceptual models are developed and used during the requirements analysis phase of information systems development [4]. At that stage, conceptual modeling is an exchange process between a *domain expert* on the one hand and a *system analyst* on the other hand [3, 5]. Typically, a domain expert can be characterized as someone with (1) superior, detailed knowledge of the object under consideration but often (2) minor powers of abstraction beyond that knowledge. The strengths of the system analyst are exactly the opposite. In this sense, the domain expert is mainly concerned with *concretization*, which refers to the act of developing an informal description of the situation under consideration. The system analyst, in contrast, is concerned with *abstraction*, i.e., using the informal description to derive a formalized model of the object.

The interaction between domain expert and analyst comprises elicitation, mapping, verification, and validation. In the *elicitation* step the domain expert produces an initial problem specification, also referred to as the *dialogue document*. As natural language is human's essential vehicle to convey ideas, this is normally written in natural language [3]. The primary task of the system analyst is to *map* the sentences of this informal dialogue document onto concepts of a modeling technique. The resulting *formal model* can be *verified* using the syntax rules of the technique. The formal model, in turn, can be translated into a comprehensible format for *validation* purposes. DeMarco states that a dialogue document is not the problem in the analysis if it is a *"suitably partitioned spec with narrative text used at the bottom level"* [6]. This statement is in line

with more general insights from cognitive psychology that the presentation of a problem has a significant impact on the solution strategy [7]. This need for a good *organization* of the requirements is reflected in various works that suggest guidelines and automated techniques for increasing the quality of informal descriptions, particularly of use cases [8, 9]. The quality of the dialogue document can be improved using a multitude of requirements elicitation techniques [10].

In the situation of *casual* modeling the steps of mapping, verification, and validation are conducted by a system analyst with limited abstraction capabilities. Currently, we lack a detailed understanding of what the process of mapping an informal dialogue document to a process model looks like and what exactly the abstraction capabilities entail that we expect from a good system analyst or process modeler. In this paper, we focus on the organization of domain knowledge as it has to be done during mapping to a formal model. To investigate its impact on the creation of process models, we provide different dialogue documents in an experiment that have different degrees of internal organization. For subjects like graduate students without established expertise in modeling, we should be able to observe the consequences of a lack of content organization. Insights from this investigation might improve guidelines on organizing a dialogue document and the effectiveness of approaches supporting the modeling process.

3   Research Setup

The main goal of our experiment is to investigate the impact of content organization of the dialogue document on the *modeling outcome* and the *modeling process*. To this end, we designed a task of creating a formal process model in BPMN syntax from an informal dialogue document under varying levels of content organization. To investigate the very *process* of process modeling, we recorded every modeling step in the experiment in a *log*. In this section the setup of our experiment is described in conformance with the guidelines of [11].
Subjects: In our experiment, subjects are 66 students of a graduate course on Business Process Management at Eindhoven University of Technology and 17 students of a similar course at the Humboldt-Universität zu Berlin. Participation in the study was voluntary. The participants conducted the modeling in the CHEETAH BPMN Modeler [12] which is a graphical process editor, specifically designed for conducting controlled experiments.
Objects: The object to be modeled is an actual process run by the "Task Force Earthquakes" of the German Research Center for Geosciences (GFZ), who coordinates the allocation of an expert team after catastrophic earthquakes. The task force runs in-field missions for collecting and analyzing data, including seismic data of aftershocks, post-seismic deformation, hydrogeological data, damage distribution, and structural conditions of buildings at major disaster sites [13]. In particular, subjects were asked to model the "Transport of Equipment" process of the task force. The task force needs scientific equipment in the disaster area to complete its mission. We provided a description of how the task force transports its equipment from Germany to the disaster area.

Factor and Factor Levels: The considered factor in our experiment is the organization of the dialogue document. We provided our subjects with dialogue documents with varying degrees of content organization simulating the structuring capabilities of domain specialists. The documents differ in the order in which the process is described (Factor Levels: breadth-first, depth-first and random order description). For all three dialogue document variants, we created a natural language description of the process from a set of elementary text blocks, each block describing one activity of the process. Depending on the factor level, the text blocks were ordered differently. The breadth-first description begins with the start activity and then explains the entire process by taking all branches into account. The depth-first description, in turn, begins with the start activity and then describes the entire first branch of the process before moving on with other branches. Finally, the random description yields a dialogue document for which the order of activity text blocks does not correlate with the structure of the process model.

Response Variables: As response variable we considered *accuracy* of the resulting model, estimated by comparing each model to a reference model of the process. Here, we relied on the graph-edit distance, which defines the minimal number of atomic graph operations needed to transform one graph into another. The graph-edit distance, in turn, can be leveraged to define a similarity metric [14]. For our setting, we weighted insertion and deletion operations of edges and nodes equally, whereas node substitutions are not taken into account as they have been established manually for corresponding pairs of activities. The corresponding hypothesis is:

Null Hypothesis H_0: There is no significant difference in the accuracy of the created process models between the three groups.

4 Performing the Experiment

This section describes the preparation and execution of the experiment, as well as the analysis and discussion of the results.

Preparation: As part of the set-up of the intended experiment, we provided the task to model the "Transport of Equipment" process from a natural language description. Three variants of the task were created: A depth first description (Variant D), a breadth first description (Variant B) and a random description (Variant R). To ensure that each description is understandable and can be modeled in the available amount of time, we conducted a pre-test with 14 graduate students at the University of Innsbruck. Based on their feedback, the modeling task descriptions were refined in several iterations.

Execution: The experiment was conducted at two distinct, subsequent events. The first event took place early November 2009 in Berlin, the second was performed a few days later in Eindhoven. The modeling session started with a demographic survey and was followed by a modeling tool tutorial in which the basic functionality of the BPMN Modeler was explained to our subjects. This was followed by the actual modeling task in which the students had to model

the "Transport of Equipment" process. Roughly a third of the students were randomly assigned to the D variant of the modeling task, another third to the B variant and the remaining third to the R variant. After completing the modeling task, the students received a questionnaire on cognitive load.

Data Validation: Once the experiment was carried out, logged data was analyzed. We discarded data of 8 students because respective data was incomplete. Finally, data provided by 66 Eindhoven students and 17 Berlin students was used in our data analysis.

Data Analysis: In total 83 students participated in our experiment. Out of the 83 students 27 worked on the breadth-first description, 25 on the depth-first description and 31 on the random description. To assess how accurately the 83 models reflect the "Transport of Equipment" process, we compared each model to a reference model of the process using the graph-edit distance. A statistical analysis revealed a significant difference in accuracy between the three groups in terms of this similarity metric (p=0.0026), see Fig. 1. Pairwise Mann-Whitney tests showed a significant difference between breadth-first and random (p=0.0021 < 0.05/3) and between depth-first and random (p=0.0073 < 0.05/3). No difference can be observed between breadth-first and depth-first (p=0.4150).
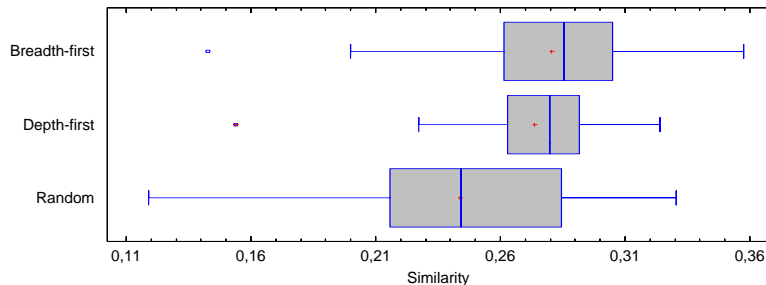
Fig. 1. Accuracy of Models

Discussion of Results. Our concern in this paper is the impact of the organization level of an informal specification on the outcome of a modeling process. This contention seems to be confirmed: An explicit ordering of the specification is positively related to the *accuracy* of the process model that is derived from it. The models created from a breadth-first and depth-first description are significantly more similar to the reference model than those created on the basis of the randomized description. The group dealing with the latter had to re-organize the dialogue document quite extensively. This suggests that casual modelers would perform better when presented with well structured specifications.

How can the insights from this study be exploited? First of all, it seems reasonable to be selective with respect to the domain experts that will be involved in drawing up the informal specifications. After all, some may be more apt than others to bring structure to such a document. Secondly, it may be feasible to instruct domain experts on how to bring structure to their specifications. In a research stream that is concerned with structuring *use cases* [9, 15], various

measures can be distinguished to ease the sense-making of these. For example, one proposal is to create pictures along with use cases that sketch the hierarchical relations between these or simply use numbering to both identify and distinguish between logical fragments of the use cases.

5 Summary and Outlook

This paper presented findings from an experiment investigating the impact of content organization of the informal dialogue document to both the modeling outcome and the modeling process. Apparently, a breadth-first organization was best suited to yield good results, indicating that industrial practice of process modeling can be improved when selecting domain specialists with good content organization skills. Our future work aims at further investigating the process of creating process models.

References

1. Indulska, M., Recker, J., Rosemann, M., Green, P.: Business process modeling: Current issues and future challenges. In: Proc. CAiSE '09. (2009) 501–514
2. Davis, A. et al.: Identifying and measuring quality in a software requirements specification. In: Proc. METRICS. (1993) 141–152
3. Frederiks, P.J.M., van der Weide, T.P.: Information Modeling: The Process and the Required Competencies of Its Participants. DKE **58** (2006) 4–20
4. Wand, Y., Weber, R.: Research Commentary: Information Systems and Conceptual Modeling - A Research Agenda. ISR **13** (2002) 363–376
5. Hoppenbrouwers, S., Proper, H., Weide, T.: A fundamental view on the process of conceptual modeling. In: Proc. ER '05. (2005) 128–143
6. DeMarco, T.: Software Pioneers. Contributions to Software Engineering (2002)
7. Lindsay, P., Norman, D.: Human Information Processing: An introduction to psychology. Academic Press (1977)
8. Cockburn, A.: Writing Effective Use Cases. Addison-Wesley (2000)
9. Rolland, C., Achour, C.B.: Guiding the Construction of Textual Use Case Specifications. DKE **25** (1998) 125–160
10. Davis, A.M. et al.: Effectiveness of requirements elicitation techniques: Empirical results derived from a systematic review. In: Proc. RE (2006) 176–185
11. Wohlin, C., et al.: Experimentation in Software Engineering: an Introduction. Kluwer (2000)
12. J. Pinggera and S. Zugal and B. Weber.: Investigating the Process of Process Modeling with Cheetah Experimental Platform. accepted for ER-POIS '10 (2010)
13. Fahland, D., Woith, H.: Towards Process Models for Disaster Response. In: Proc. PM4HDPS '08. (2008) 254–265
14. Dijkman, R., Dumas, M., García-Bañuelos, L.: Graph Matching Algorithms for Business Process Model Similarity Search. In: Proc. BPM '09. (2009) 48–63
15. Constantine, L., Lockwood, L.: Structure and style in use cases for user interface design. Object Modeling and User Interface Design (2001) 245–280