A Semiautomatic Process Model Verification Method based on Process Modeling Guidelines

Valter Helmuth Goldberg Júnior¹, Lucineia Heloisa Thom¹, José Palazzo Moreira de Oliveira¹, Marcelo Fantinato² and Diego Toralles Avila¹

¹Department of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
²School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, Brazil

Keywords: Business Process, BPMN, Ontology, Process Model Quality, Modeling Guidelines.

Abstract: Designing comprehensible process models is a complex task. Process analysts must rely on the experience of expert systems managers to achieve process models with high comprehensibility, also known as pragmatic quality. In the literature, this is portrayed as process modeling guidelines that help modelers to avoid common issues which hinder the comprehension of the process model. In this paper, we propose a method for the semiautomatic verification of business process models according to process modeling guidelines. This method uses the BPMN Ontology and the ontology editor Protégé to assist the modeler with validation of the process model’s syntax before certifying its pragmatic quality. The validation of the developed method was applied to a collection of 31 process models and the results show that 23 process models of the collection contain at least one guideline violation.

1 INTRODUCTION

Business Process Management (BPM) is a discipline that provides a systematic approach to manage an organization’s work by modeling, analyzing, improving and controlling its business processes (hereafter called processes, for simplification). BPM contributes to the increase of productivity and reduction of costs through more effective, more efficient and more adaptable processes (van der Aalst, 2013).

With BPM, organizations continually seek to improve the quality of their processes. However, studies analyzing models of industry processes reveal that many process models contain issues that harm its quality, such as control flow errors, badly designed structures and layouts or incorrect labeling (Mendling et al., 2008; Leopold et al., 2016). With the process modeling being a key part of BPM, it is important to prevent these issues if we expect processes of better quality.

It is widely accepted that modeling a process is a difficult task (Mendling et al., 2010). This is usually due to the complexity of the modeling notation, its many different elements and their respective semantics (Leopold et al., 2016). Choosing the appropriate design to represent the real world process depends upon the expertise or the guidance of an experienced modeler, which may greatly influence the quality of the resultant process model. It is necessary to consider that, while the use of process modeling tools can help in this regard, they cannot guarantee a process model’s correctness nor its comprehensibility.

One way to solve this challenge and help the beginner modelers is to consolidate the knowledge of experienced modelers in process modeling guidelines, whose purpose is to help the user to reduce the complexity and the number of errors in a process model through the restriction of undesirable constructs. Many guidelines have been proposed by both practitioners (Silver, 2009; White and Miers, 2008; Allweyer and Allweyer, 2010) and researchers (Becker et al., 2000; Mendling et al., 2007; Vanderfeesten et al., 2008; Correia and Abreu, 2012). Once it is verified that a process model is following a set of guidelines, we can presume that it has good comprehensibility.

However, using guidelines to verify a process model does not make sense if the model is not syntactically correct. Any knowledge extracted from an incorrect process model has its validity compromised as at the same time that the model is readable some doubt may exists about the modeler intended representation (Reijers et al., 2015). Therefore, it is necessary to verify if a process model is correct before
considering its comprehensibility.

It is possible to verify the correctness of process models in different ways. One of these is by the means of ontologies. Ontology is the study of the nature of being, which pursues to represent the world as entities, categories and relations (Guizzardi, 2012; Mendling, 2008a). In a more practical description, an ontology provides an approach to define types, properties and relations. It is possible to use an ontology to represent BPM process models as a meta-model with inference capability to verify a process model's correctness.

Following this approach, the objective of this paper is to show how the use of ontologies may assist in the identification of problems that reduce a process model’s comprehensibility. To perform this validation, it is necessary to represent a process model as a process model ontology and use the ontology inference processor to verify a set of guidelines for process modeling, pointing out any problems that can be improved.

This paper is organized as follows: Section 2 presents a short review of previous works related to the verification of process models. Section 3 shortly introduces the basic concepts used in this paper. Section 4 explains the context chosen to work with and the method developed. Section 5 presents the application study and the results. Section 6 closes the paper with conclusions.

2 RELATED WORKS

The verification of process models is nothing new, in fact, numerous researches addressed this subject. The difference is that most of these researches are concerned with issues of correctness of a process model. In Mendling (2008a), for example, the author proposes two different approaches to verify the soundness of a process model draw using Event-Process-Chains.

Evaluating and reducing the complexity of a process model is harder to achieve. It is not possible to measure a process complexity directly and, as a consequence, many metrics have been proposed that try to solve this problem indirectly (Vanderfeesten et al., 2007; Mendling, 2008b; Gruhn and Laue, 2006; Sánchez-González et al., 2012). The validity of these metrics is evidenced through statistical experiments, where process models are judged both by the metrics and by people with varying levels of modeling experience (Cardoso, 2006; Sánchez-González et al., 2008).

As it was mentioned, many authors proposed guidelines for modeling processes, with repeated guidelines that had already been proposed, with only few small variations in the details. In Moreno-Montes de Oca and Snoeck (2014), these repeats were gathered from a systematic review about business process modeling quality from over a 100 proposed in the literature and turned into 27 unified guidelines.

Some of the existing BPMN tools try to provide some support for creating good process models. Based on the guidelines found in the referred article, a study (Snoeck et al., 2015) was performed to test how extensive was, at the time, the support of the popular BPMN tools in creating good process models. From this analysis, it is possible to learn that the Signavio1 modeller tool provides the best support for modeling processes using guidelines.

While the referenced works are built upon important process modeling concepts, none of them provide a complete approach to certify the comprehensibility of process models. The present work explains the development of a method and how to adapt each concept for an ontological approach that allows the verification of the process models in a semi-automatic way.

3 BACKGROUND

The modeling task of BPM is often done using the Business Process Model and Notation (BPMN). BPMN was developed by the Object Management Group (OMG), with the purpose of consolidating the many existing notations for process models in a single standard. This standard should provide an easy to understand notation to all stakeholders (OMG (Object Management Group), 2015).

Most BPMN elements are labeled, to identify the parts of the process they constitute. BPMN also has many different types of elements to represent a process, each with a different purpose. There are five main types covered in this paper:

- **Activities** represent tasks that may need to be performed (e.g., "Print daily report") or subprocesses, that contains multiple elements of the process inside a single element.
- **Events** correspond to things that happen instantaneously (e.g., "Purchase order received"). Events may be start, intermediate or end events.
- **Sequence Flows** link two elements together, forming the paths that may be taken during the execution of the process.
- **Gateways** serve to split or join the flow of actions performed in the process. There are differ-

1www.signavio.com
ent types of gateways: AND gateways for concurrency, XOR gateways for exclusive choices and OR gateways for inclusive choices.

- **Pools** group together elements that happen in a single organization (e.g., a university). Swimlanes may divide these pools to identify different resources of that organization (e.g., departments).

Although these types are useful in the modeling of processes, BPMN does not teach modelers how to use them to create simple and expressive process models. The consequence of this is that it is hard to achieve a good level of quality in BPMN process models.

This difficulty motivated the creation of many frameworks that try to define what is the quality of a process model and classify the different quality types that compose it. Examples of these efforts are the SEQUAL Framework (Krogstie, 2012), the Guidelines of Modeling (GoM) (Schuette and Rotthowe, 1998) and, more recently, the SIQ framework (Reijers et al., 2015), in which we base the present work, due to its simplicity and its widespread use in the literature surrounding the quality of process models. The SIQ framework defines process model quality as made of three basic quality types:

- **Syntactic Quality** identifies if a process model conforms to the rules defined by the notation utilized to create it. In other words, if a process model follows the syntax and the vocabulary of its modeling language, then it is possible to verify that process model and affirm it to be correct. To do so, the verification must check the static properties of a process model - how different types of elements are used and combined - and its behavioral properties - the process modeled should not reach a deadlock and must be completed properly, i.e., the process model is sound.

- **Semantic Quality** bears the connection between a process model and the real world process that it is supposed to represent. To check a process model’s semantic quality is to guarantee it is valid - all elements of the process model correctly represent the real world - and complete - there are no real world process parts that are missing from the process model. This check is called validation and, if it passes, the process model is ascertained to be true.

- **Pragmatic Quality** characterizes the comprehensibility of a process model. It is the certification that a user’s interpretation of a process model is equal to the real world process. If done, the process model is said to be understood.

Syntactic quality is the basis for the other two qualities. As mentioned before, it is not reasonable to consider the comprehensibility of a process model if it is not syntactically correct. The same concern must be expressed on its semantic quality. The verification of a process model must be done before its validation or certification. As previously explained, it is possible to do this verification using an ontology. More specifically, we can use an ontology design to serve as a meta-model for a process modeling notation. In the case of BPMN exists the BPMN Ontology (Rospocher et al., 2014), which supports the mapping of a BPMN process model into elements of the ontology, while preserving the relations and structures between the BPMN elements. Using this ontology, it is possible to apply an inference engine to verify the mapped process model, checking if the static properties of BPMN model, i.e., its structure, is correct according to the BPMN syntax.

Finally, assuming the process model is indeed correct, we can try to ensure its pragmatic quality, which is done by checking it via the use of process modeling guidelines. In Mendling et al. (2010, pag. 3), seven process modeling guidelines (7PMG) have been proposed that are “thought to be helpful in guiding users towards improving the quality of their models, in the sense that these are likely (1) to become comprehensible to various stakeholders and (2) to contain few syntactical errors”. These guidelines have been built upon empirical insights and, as such, provide a short but meaningful set of rules, which encouraged their use at an academic level to teach beginner modelers about quality of process models. They are as follows:

**G1 Use as Few Elements in the Model as Possible.**

The larger a process model is, the more difficult it is to understand and the more likely it is for syntactical errors to exist in it.
G2 Minimize the Routing Paths per Element. These paths are the sum of the incoming and the outgoing arcs for each element. A high number of paths in a single element makes the model harder to understand.

G3 Use One Start and One End Event. Models that satisfy this requirement are easier to understand and are less likely to have errors.

G4 Model as Structured as Possible. A process model is structured if for each gateway that splits the flow of the process model there is another gateway of the same type that joins the flow. Ideally, a structured model is like a math formula with balanced brackets, i.e., every opening bracket has a corresponding closing bracket of the same type. Structured models tend to be easier to understand and to have less errors.

G5 Avoid OR Routing Elements. The behavior of OR gateways are more difficult to comprehend and limiting their use reduces the likelihood of misinterpretations.

G6 Use Verb-object Activity Labels. There are many different labeling styles for process models. According to the literature, the verb-object style is less ambiguous and more useful than the others, like action-noun.

G7 Decompose a Model with More than 30 Elements. Like G1, a high number of elements makes the process model less understandable and more error-prone. After 30 elements, it is recommended to split the process model into smaller models, either by creating new models or by gathering a group of process model elements and replacing them with a sub-process.

It is important to note that these guidelines do not concern with the semantics of the process model. Whether a model of a specific process follows or not these guidelines this characteristic should not imply in a change of the behavior of the modeled process. All that the 7PMG rules change is the comprehensibility of the process model and the reduction of possibility that modeling errors exist in it (Mendling et al., 2010).

4 GUIDELINE-DRIVEN PROCESS MODEL VERIFICATION

To fulfill the objective of this paper, it is necessary to specify a series of steps that, with the assistance of an ontology, allows us to certify a process model’s pragmatic quality by checking whether it follows the seven process modeling guidelines.

<table>
<thead>
<tr>
<th>7PMG</th>
<th>Indicators and comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>Number of elements &gt; 30</td>
</tr>
<tr>
<td>G2</td>
<td>Maximum connector degree &gt; 5</td>
</tr>
<tr>
<td>G3</td>
<td>Number of start events &gt; 1</td>
</tr>
<tr>
<td>G4</td>
<td>Number of splits ≠ Number of joins</td>
</tr>
<tr>
<td>G5</td>
<td>Number of OR gateways &gt; 0</td>
</tr>
<tr>
<td>G6</td>
<td>Wordnet</td>
</tr>
<tr>
<td>G7</td>
<td>Number of elements &gt; 30</td>
</tr>
</tbody>
</table>

To start with the development of the framework it is necessary to decide about the form of how the process models will be represented. A few different notations for process models exist and each notation has different ways of how the process model is coded within a file. The notation used in this work is BPMN 2.0. Models using this notation may be exported to the interchangeable format defined by OMG, which is a XML file with a specific schema and a .bpmn extension (OMG (Object Management Group), 2015). From this file, it is possible to map elements from the process model in an ontology, via the BPMN Ontology.

Secondly, it is necessary to establish how to check whether a guideline is being followed or not by a process model. They must be expressed in a binary yes or no question. To allow this, each guideline must be associated to a process model indicator that may be measured and compared against optimal thresholds. Previous works (Mendling, 2008a; Recker, 2011; Mendling et al., 2012; Sánchez-González et al., 2012) have studied process model indicators to produce empirically validated thresholds for each guideline, allowing a modeler to be alerted when there is an issue that affects the process model’s comprehensibility. The table 1 presents the indicators and the optimal thresholds used to check if the process model violates each guideline. Based on these, two guidelines show problems: G1 and G6.

G1, the guideline for encouraging the use of fewer elements when modeling, becomes redundant with G7, which determines when a process model should be decomposed. This occurs because the indicator used for G1 is the same one employed for G7. Therefore, we need to choose which guideline is more appropriate for the proposed method. Since G1 is more suited to be applied when a process model is being developed and the modeler can refrain from introducing new elements, instead of when the modeling is finished, the guideline G7 is the most suited.

G6, which tells us to label activities in the verb-object style, presents a problem related to the complexity of checking the language of each label. This
Table 2: BPMN ⇒ Ontology Mapping.

<table>
<thead>
<tr>
<th>BPMN Ontology Example</th>
<th>Ontology Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Element type</td>
<td>OWL class</td>
</tr>
<tr>
<td>Activity, gateway</td>
<td>Activity, gateway</td>
</tr>
<tr>
<td>Element instance</td>
<td>Individual named</td>
</tr>
<tr>
<td>Task 1: Submit report</td>
<td></td>
</tr>
<tr>
<td>Attribute</td>
<td>Object property</td>
</tr>
<tr>
<td>Label=&quot;Name&quot;</td>
<td></td>
</tr>
<tr>
<td>Attribute value</td>
<td>Data property</td>
</tr>
<tr>
<td>Name:String= &quot;Task 1: Submit report&quot;</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Recommended Actions for each tested guideline.

<table>
<thead>
<tr>
<th>7PMG</th>
<th>Recommended action</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>Reduce the number of sequence flows connected to a single element</td>
</tr>
<tr>
<td>G3</td>
<td>Restructure the process model to reduce the number of Start and End events</td>
</tr>
<tr>
<td>G4</td>
<td>Restructure the process model to have the same number of Split and Joins</td>
</tr>
<tr>
<td>G5</td>
<td>Restructure the process model to remove the OR Gateways</td>
</tr>
<tr>
<td>G7</td>
<td>Decompose the process model</td>
</tr>
</tbody>
</table>

The entire series of steps are (figure 2):
1. Load the BPMN Ontology in Protége.
2. Extract each individual element from a BPMN model.
3. Instantiate each extracted element in Protége via OWL-API and using the BPMN Ontology.
4. Use Protége’s inference engine to verify the integrity of the new ontology.
5. Check if the process model’s indicators obey the limits defined by the modeling guidelines.
6. Recommend modeling alternatives to the process model for each guideline not followed (table 3).

5 APPLICATION STUDY AND RESULTS

To validate our method, it was applied to a collection of 31 BPMN models. These models represents the processes of a university, created by BPM students, verified and corrected by their adviser and semantically validated by the process stakeholders. To create these models, the students used the Bizagi Modeler tool, which does not offer support for the guidelines we test (Snoeck et al., 2015).

For each guideline used in the method, the associated indicators were extracted for a statistical analysis.

2http://protege.stanford.edu/

3http://www.bizagi.com
Figure 2: Steps to verify process models based on process modeling guidelines.

Table 4: Statistics for the indicators related to the guidelines G2, G3, G5, G7.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Maximum connector degree</th>
<th>Nº Start events</th>
<th>Nº End events</th>
<th>Nº OR gateways</th>
<th>Nº Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>3.7097</td>
<td>1.871</td>
<td>1.9355</td>
<td>0.2903</td>
<td>23.4194</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>0.9379</td>
<td>0.8848</td>
<td>0.892</td>
<td>0.8638</td>
<td>13.458</td>
</tr>
<tr>
<td>Minimum</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Maximum</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>76</td>
</tr>
<tr>
<td>Median</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 5: Statistics for the indicators related to the guideline G4.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Splits - Joins difference</th>
<th>AND</th>
<th>XOR</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>0.129</td>
<td>0.4839</td>
<td>-0.0968</td>
<td></td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>0.3408</td>
<td>1.3873</td>
<td>0.3962</td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>-2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

(as seen in tables 4 and 5). Based on these statistics, we tried to predict whether the process models of this collection follow or not the process modeling guidelines.

For example, in the statistics for maximum connector degree, associated with guideline G2, we see that the highest value found was 6, which is above the recommended value of 5 for this guideline. However, if we look at the average of 3.7097 and the standard deviation of 0.9379, it is possible to perceive that it is unlikely, assuming a normal distribution, that a random process model picked from our collection will have more than the recommended threshold for this guideline, which is five. Consequently, we can anticipate that the high majority of process models from the collection will follow the guideline. G5 is similar, since the average number of OR gateways is low, but not zero, and the standard deviation is almost 1, implying that a few process models do use OR gateways, up to the maximum of 4. For this reason, we suppose that a few process models violate this guideline, and our method shows this.

For G3, on the other hand, we see the opposite situation, as the number of both start and end events has the high average (almost 2), compared to the recommended use of 1 event of each. This implies that most process models from the collection have multiple start and end event and, therefore, they do not follow G3.

Analysis for guideline G4 is more complicated, since we define whether the process model is or not structured as the measure of the difference between the indicators of the number of splits and the number of joins. Not only that, it is also necessary to measure the difference for each type of gateway. This balance means that the closer to zero is the average difference and the lower is the standard deviation then it is more likely that the process models are structured. We recognize the opposite, however, since there is an imbalance of the number of XOR splits versus the number of XOR joins according to the average of 0.4839 for that indicator. Also, the high standard deviation of 1.3873 shows for the XOR gateway indicates that most process models in the collection are not structured. Therefore, we predict that many
process models of this collection infringe the guideline G4.

Finally, the statistics for G7 are slightly vague. The high, but not unreasonable, average for the measure of the number of elements suggest that the guideline is followed. Yet, the high standard deviation shows a hint that at least some process models do have more than 30 elements.

With all this information in mind, our expectations are that most process models violate guidelines G3 and G4, while following guidelines G2 and G5. Lastly, guideline G7 will have a few violations.

Comparing these conclusions with the results of the applied method (as show in table 6) shows that the predictions for guidelines G2, G4, G5 and G7 match the results. The results for guideline G3, however, is unexpected. After further analysis, we found a reason for this behavior. Many process models of the collection have multiple pools or sub-processes, both of which require, according to the notation, a new start and a new end event, causing a distortion in the number of events of each process model. Therefore, we must take this distortion into account for the statistical analysis.

Table 6: Number of violations per guideline.

<table>
<thead>
<tr>
<th></th>
<th>Total violations</th>
<th>Percent of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>1</td>
<td>3.23%</td>
</tr>
<tr>
<td>G3</td>
<td>1</td>
<td>3.23%</td>
</tr>
<tr>
<td>G4</td>
<td>22</td>
<td>70.97%</td>
</tr>
<tr>
<td>G5</td>
<td>4</td>
<td>12.90%</td>
</tr>
<tr>
<td>G7</td>
<td>6</td>
<td>19.35%</td>
</tr>
</tbody>
</table>

6 CONCLUSION

Process models that follow the guidelines of process modeling are more likely to be understood by the process stakeholders. With the help of process modeling tools that support those guidelines, modelers can represent and communicate the mechanisms of a process through the models, achieving higher levels of modeling quality that improve the work of an organization.

Through the application study, we perceived the difficulty that beginner modelers experience in creating process models following the modeling guidelines. Most process models in the collection have at least one aspect that is inefficient for the their comprehensibility (table 7). It is observable that beginners need help in finding these inefficiencies and that this help may only be accomplished if more support is provided by the modeling tools.

In this paper, it was shown that it is possible to semi-automatically verify process modeling guidelines with the assistance of the BPMN ontology. This procedure analyzes the process models and recommends solutions to increase their pragmatic quality. The plugin developed based on this procedure identifies modeling inefficiencies that violate the seven process modeling guidelines and we hope that it may contribute in helping beginner modelers to identify possible modeling inefficiencies in regard with the comprehensibility of process models.

One limitation that must be acknowledged is related with the use of the BPMN ontology to check the syntax of the process model. While it is possible to verify the static proprieties of a process model, the BPMN ontology is not suited to specify a process model’s dynamic behavior (Rospocher et al., 2014). Therefore, we were not able verify a process model’s entire syntactic quality.

While the method is by no means complete, as there are more guidelines in the literature beyond those applied here, we hope that this work brings attention to the necessity of accurate tool support for creating process models with high comprehensibility. We also believe that our work will provide a basis for future works in this area.

REFERENCES


