

# Automatic Generation of Questionnaires for Managing Configurable BP Models

A. Jiménez-Ramírez<sup>1</sup>, B. Weber<sup>2</sup>, I. Barba<sup>1</sup> and C. Del Valle<sup>1</sup>

<sup>1</sup>University of Seville, Dpto. Lenguajes y Sistemas Informáticos, Seville, Spain

<sup>2</sup>University of Innsbruck, Department of Computer Science, Innsbruck, Austria

**Keywords:** Configurable Business Process Models, Classification Trees, Questionnaires.

**Abstract:** Managing large collections of business process (BP) models is increasingly being necessary for organizations. For this, configurable BP models can be used for managing these BPs while allowing analysts to understand what these BPs share and what their differences are. Before the execution of the configurable BP model, a BP model has to be selected from it. This selection is typically performed by an analyst who manually individualizes the model in order to address the business requirements. Unlike existing approaches, we propose a totally automated method to create a questionnaire-based application for guiding a business expert on individualizing a model.

## 1 INTRODUCTION

A Business Process (BP) can be defined as a set of activities which are performed in coordination in an organization to achieve a business goal (Weske, 2007). These activities can be manual activities, other BPs, or even pieces of software. Nowadays, in order to support BPs, BP Management (BPM) embraces methods, techniques, and software to design, enact, control, and analyze operational processes involving humans, organizations, applications, and other sources of information (van der Aalst et al., 2003). Such management generally follows a strict methodology to ensure the quality of the information systems which are created. Typically, the traditional BPM life cycle (Weske, 2007) includes four phases, i.e., process design & analysis (i.e., a design of the BP is created following the requirements), system configuration (i.e., the software defined in the BP design is implemented), process enactment (i.e., the software is executed following the BP design) and evaluation (i.e., monitoring information or logs are analyzed to look for design improvements) (Weske, 2007).

It becomes increasingly common for organizations to deal with large collections of BP models (e.g., due to company mergers (Rosa et al., 2012), BP models extracted from declarative specifications (Jimenez-Ramirez et al., 2013), etc.). Therefore, as shown in (Dijkman et al., 2012), more and more research is done on BP model collections. In literature,

different techniques are proposed to manage such collections (Dijkman et al., 2012). Among others, variant management techniques (Rosemann and van der Aalst, 2007; Gottschalk et al., 2008; Hallerbach et al., 2010) keep track of BP variants of a similar BP that co-exist within a collection. Such collections are typically represented as a configurable BP model (Rosemann and van der Aalst, 2007; Rosa et al., 2012), i.e., a modeling artifact that captures a family of BP models in an integrated manner and that allows analysts to understand what these BP models share, what their differences are, and why and how these differences occur. Configurable BP models include specific nodes called configurable nodes which represent the variation points of the model.

Before the organizations can execute a configurable BP model, it needs to be *individualized*, i.e., a single BP model has to be selected from it. Therefore, a new phase, namely *configuration & individualization*, is defined in the BPM life cycle between the process design & analysis and the system configuration phases (La Rosa et al., 2008).

### 1.1 Problem Statement

Generally, in order to individualize a configurable BP model (cf. Fig. 1 (1)), the analyst considers the business requirements (i.e., which are typically specified by the business expert, cf. Fig. 1 (2)) and manually performs the individualization (i.e., she selects

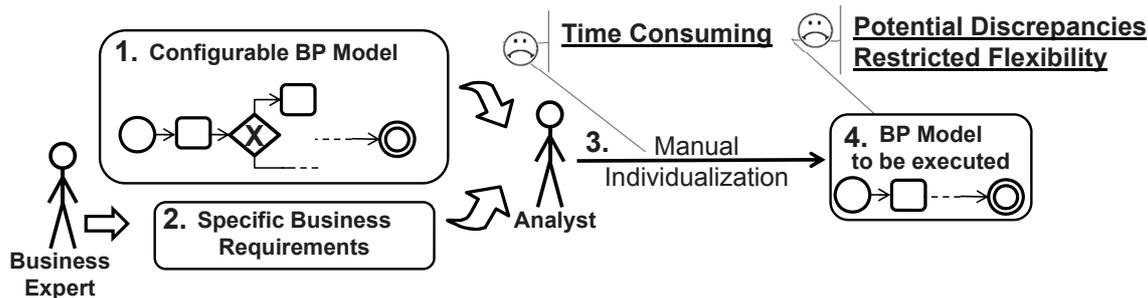


Figure 1: Motivation.

one of the variants which exist in the configurable BP model, cf. Fig. 1 (3)). Therefore, this manual task can be time-consuming and, in addition, it can lead to discrepancies between the business expert and the selected BP model. Furthermore, since the different BP models of a configurable BP model share many commonalities (Rosemann and van der Aalst, 2007) (even in the initial parts of them), selecting a single BP model at configuration-time unnecessarily restricts the flexibility especially when the context might change (cf. Fig. 1 (4)).

## 1.2 Contribution

The current work proposes a method for supporting the business expert in individualizing the model through a questionnaire-based approach, i.e., a sequence of questions each one created for individualizing a part of the model (La Rosa et al., 2008). Taking a configurable BP model as starting point (cf. Fig. 2 (1)), the questionnaire which is automatically generated consists of different questions written in the business language (i.e., using properties that can be measured in the BP models of the configurable BP model and that have enough semantic to be understandable by the business expert, cf. Fig. 2 (2)). Therefore, the business expert can individualize the models herself by answering questions (cf. Fig. 2 (3)) without the intermediation of the analyst. Furthermore, the generated questionnaires are intended to individualize the model in an incremental way, i.e., guiding the execution of the configurable BP model. Therefore, the proposed method starts individualizing initial parts of the model and iteratively individualizes further succeeding parts during run-time, i.e., when more information is available to take decisions (cf. Fig. 2 (4)).

Unlike existing approaches (La Rosa et al., 2008; Rosa et al., 2009), the current work proposes a method for *automatically* generating questionnaires for managing configurable BP models.

In this work, a preliminary study of the aforementioned process (cf. Fig. 2) is conducted. The main

contributions of this paper are: (1) the automatic generation of questionnaires for a configurable BP model and (2) the method for using such questionnaires at run-time for guiding the execution of a configurable BP model.

This paper is organized as follows: Sect. 2 introduces backgrounds on related areas, Sect. 3 shows how the questionnaires are created from a configurable BP model and how they can be used at run-time, and Sect. 4 includes some conclusions and future work.

## 2 BACKGROUND

In this paper, BP models (cf. Sect. 2.1) are considered as parts of configurable BP models (cf. Sect. 2.2). Such configurable BP models are used as starting point for automatically generating questionnaires (cf. Sect. 2.3) which will individualize them.

### 2.1 Business Process Models

BPs are commonly used to coordinate activities between organizations. To deal with BPs, in this work the BP graph definition (cf. Def. 1) introduced in (Rosa et al., 2012) is used.

**Definition 1.** A *BP Graph*  $G = (gid, N, Pairs)$  is identified by  $gid$  and consists of a set of pairs of nodes  $n \in N$ , i.e.,  $Pairs$ . Each pair denotes a direct edge between two nodes in the graph. A node  $n \in N$  is a tuple  $\langle nid, l, t \rangle$  where  $nid$  is a unique identifier of a node in the graph,  $l$  is its label, and  $t$  is its type.

Such definition of graph allows to represent a BP model in many different BP languages (BPMN, 2011), e.g., BPMN or EPC (cf. Example 1). As an example, the types of nodes (i.e.,  $t$ ) in BPMN language (BPMN, 2011) are 'activity', 'event', or 'gateway'. A node of type 'gateway' allows labels (i.e.,  $l$ ) 'AND', 'OR', 'XOR', etc., while 'event' nodes allow 'start' and 'end' labels.

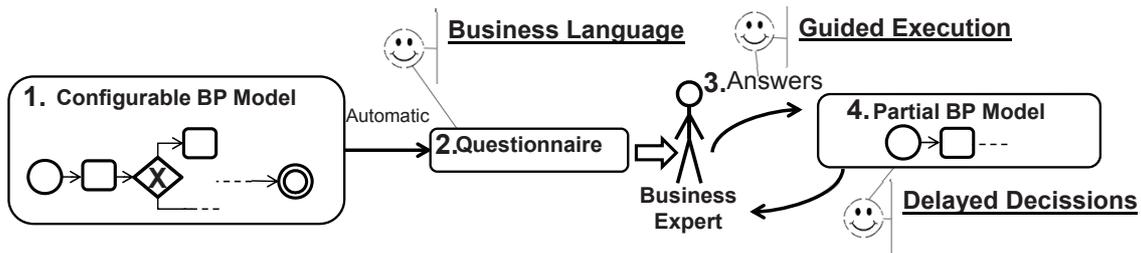


Figure 2: Overview of our contribution.

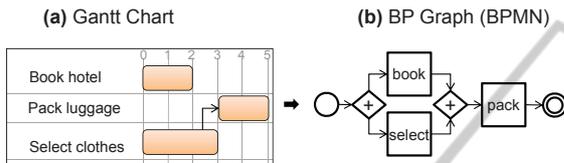


Figure 3: A schedule (a) as a BP Graph (b).

**Example 1.** Figure 3 (a) shows three activities which are scheduled to prepare a holiday and they are depicted as a Gantt chart (Gantt, 1913). The activities 'book a hotel', 'select the clothes' and 'prepare the luggage' are considered. In addition, Fig. 3 (b) shows the related BP graph using BPMN. This graph consists of the following 7 nodes (cf. Def. 1):  $\langle 1, \text{start, event} \rangle$ ,  $\langle 2, \text{AND, gateway} \rangle$ ,  $\langle 3, \text{book, activity} \rangle$ ,  $\langle 4, \text{select, activity} \rangle$ ,  $\langle 5, \text{AND, gateway} \rangle$ ,  $\langle 6, \text{pack, activity} \rangle$  and  $\langle 7, \text{end, event} \rangle$ ; which are paired (cf. Def. 1) as follows: (1, 2), (2, 3), (2, 4), (3, 5), (4, 5), (5, 6), and (6, 7).

## 2.2 Configurable BP Model

Typically, different BPs (cf. Def. 1), also called *variants*, can be performed in scenarios which entail high variability. In most cases these plans share many commonalities. Hence, these variations can be combined in a configurable BP model (i.e., a modelling artifact that capture a family of BP models in an integrated manner) leading to a compact representation (Rosa et al., 2012; Rosemann and van der Aalst, 2007; La Rosa et al., 2008; van der Aalst et al., 2006). Generally, configurable BP models allow analysts to understand what these variations share, what their differences are, and why and how these differences occur (Rosemann and van der Aalst, 2007).

Configurable BP models are typically created by hand (1) from scratch, (2) from an existing BP model by including possible adaptations (Gottschalk et al., 2008), or (3) by merging some BP models related to the same or similar goals which already exist (Rosa et al., 2012; Jimenez-Ramirez et al., 2013). In the last case, the source BP models need to be compared and merged, which might result in a tedious, time-

consuming and error-prone process if it is performed by hand (Rosa et al., 2012). To overcome these problems, there exist approaches focused on automatically merging different BP models in a configurable BP model (Rosa et al., 2010; Rosa et al., 2012).

Configurable BP models can be represented by configurable BP graphs, which are defined (cf. Def. 2) based on (Rosa et al., 2012).

**Definition 2.** A *Configurable BP Graph*  $CG = (G, E2I, N2LI)$  consists of: (1) a graph,  $G = (gid, N, Pairs)$  (cf. Def. 1), (2) a function  $E2I$  that maps each edge  $e \in Pairs$  to a set of BP graph identifiers (i.e.,  $E2I$  identifies which branches of  $CG$  belong to each source BP graph which is merged in  $CG$ ), (3) a function,  $N2LI$  that maps each node  $n \in N$  to a set of pairs  $\langle gpid, l \rangle$  where  $gpid$  is a BP graph identifier and  $l$  is the label of node  $n$  in graph  $gpid$  (i.e.,  $N2LI$  identifies which nodes, with the corresponding label, belong to each graph which is merged in  $CG$ ).

A configurable BP graph includes configuration nodes for those points where the BP graphs which are included differ (cf. Example 2). Therefore, each branch and node of the configurable BP graph can be related either to one or more BP graphs. To store these relations, each branch/node of the configurable BP graph includes identifiers related to the corresponding BP graph (i.e.,  $E2I$  function). In addition, nodes also store the associated label related to each identifier (i.e.,  $N2LI$  function).

**Example 2.** Figure 4 shows 2 graphs which are merged into a configurable BP model<sup>1</sup>. The first gateway in Fig. 4(b) is a configurable node which corresponds to an 'OR' gateway in the process 1 (it does not explicitly appear) and an 'AND' gateway in the process 2.

## 2.3 Questionnaires

Questionnaire models (Rosa et al., 2009) are generally created to support the user during the individual-

<sup>1</sup>As there is not ambiguity, some labels are not shown (i.e., they are the same as in the branch).

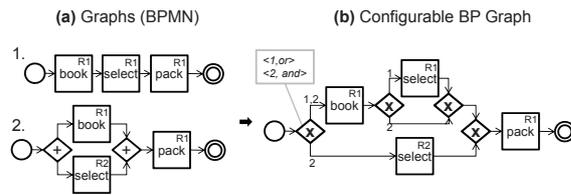


Figure 4: Two BP graphs (a) are merged into a single configurable BP graph (b).

ization of the configurable BP models. The main benefits of using them are: (1) they guide the user in such a way that choices are presented in a proper order and (2) they avoid invalid configurations which may lead to errors.

Typically, questions which are within a questionnaire are manually created. In addition, such questions are related to boolean *facts* which are associated to configuration *actions* (the reader is referred to (Rosa et al., 2009) for a review on interactive questionnaires). Therefore, each time a question is answered, an action is fired which individualizes a part of the model. The sequence of answers to different questions will individualize the configurable BP model in such a way a single BP model is selected.

Unlike the work presented in this paper, questionnaires generally individualize configurable BP models before starting the execution and thus, unnecessarily restricting the flexibility.

### 3 THE PROPOSED METHOD

In this section, the proposed method (cf. Fig. 5) is described.

#### 3.1 Configurable BP Model

As an initial step, to track the BP models during the method, all of them are labeled (cf. Example 3). Then, the configurable BP model (cf. Fig. 5 (2)) is executed until a configurable node appears (cf. Fig. 5 (2)).

**Example 3.** For the sake of simplicity, the running example of Fig. 6(a) comprises four BP models which represent different ways of executing four activities (i.e., A, B, C and D). Each BP model is label with an integer. Furthermore, a group of properties of the application domain for each BP model is provided (cf. Fig. 6 (b) where time ( $T$ ), benefit ( $B$ ) and risk ( $R$ ) properties are provided for each model).<sup>2</sup> Such

<sup>2</sup>This properties are manually provided and must be well-defined to be understandable by the domain expert.

properties are related to the business language, e.g.,  $T$  is related to how long the business is opened and  $R$  refers to the maximum risk that the business can afford.

The configurable BP model associated to the aforementioned four BP models are depicted in Fig. 7(a). In this model, four different configurable nodes are depicted with a bold diamond. In the first configurable node, labeled as 1, two alternatives are possible. The left branch comprises variant 4 (i.e., where activity A is not executed), and the right branch comprises variants 1 to 3 (where activity A is executed).

#### 3.2 Generating Classification Trees

When a configurable node is encountered we apply a method for generating a set of questions related to this node. Among other techniques as *discrimination* or *cluster* analysis, the current method uses classification trees (i.e., models that predict the value of a target variable based on several inputs variables) (Breiman, 1984) to predict which outgoing branch would correspond to a given assignment of property value.<sup>3</sup> Specifically, for each configurable node which is encountered, a classification tree is created (cf. Fig. 5 (3)) using the property values of the BP models as input variables (cf. Example 4).

**Example 4.** Fig. 7(b) shows the classification tree which results from using the CART algorithm (Breiman, 1984)<sup>4</sup> by providing the table of Fig. 1(b) as input variables and the strings left and right as target variables. As can be seen, in the resulting classification tree, the BP models for which  $T > 5$  correspond to the right branch. In contrast, the BP models for which  $T \leq 5$  correspond to (1) the right branch if  $R \leq 10$ , (2) to the left branch otherwise.

#### 3.3 Generating Questions

A set of questions is then created for each decision tree (cf. Fig. 5 (4)). For this, one question is automatically generated for each intermediate node of the decision tree, and the possible answers for the question are the different labels which are written on the outgoing branches of this node. These questionnaires are in charge of narrowing down the variants of the configurable BP model.

<sup>3</sup>Though other techniques can be use for classifying, we select the decision trees because of its hierarchical feature.

<sup>4</sup>Different methods can be used to create a classification tree. The suitability and quality of each classification method depend on the characteristics of the data. This analysis is out of the scope of the current approach.

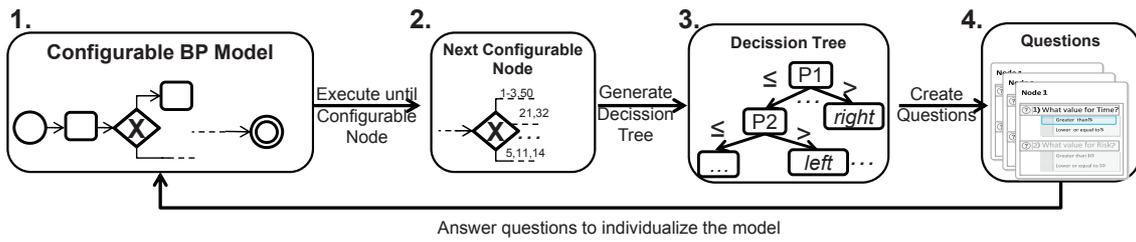


Figure 5: The proposed method.

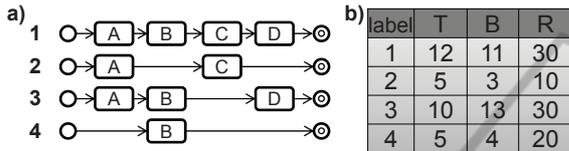


Figure 6: (a) Different BP models. (b) Table of BP model properties.

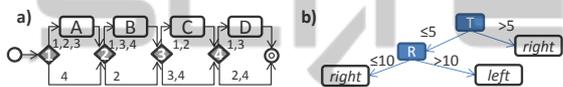


Figure 7: (a) Configurable BP model related to the BP models of Fig. 6(a). (b) Classification tree for node 1.

The text of the questions are automatically generated from the information of the properties (cf. Example 5).

**Example 5.** A simple questionnaire related to the decision tree of Fig. 7(b) is shown in Fig. 8(a). Since this decision tree has two intermediate nodes (i.e., T and R), two questions are created. Moreover, since each node has two branches, each question has two options. Initially, only the question related to T is enabled. Considering that the well-defined business properties stated that T is related to the closing time of the office, the generated question would look like What time would you close the office?. The second question has to be answered only if the user selects the second option of the first question (i.e., In 5h. or less) which is related to the branch  $T \leq 5$  of the decision tree.f

### 3.4 Incremental Configuration

Once a questionnaire is resolved, the configurable BP model is individualized by removing the BP models that do not belong to the edge which result selected in the questionnaire. Thereafter, the model is executed until a new configurable node is reached (cf. Example 6).

**Example 6.** Supposing that the user selects the first answer of the first question of the questionnaire of Fig. 8(a) (i.e., In more than 5 h.), BP models 2 and

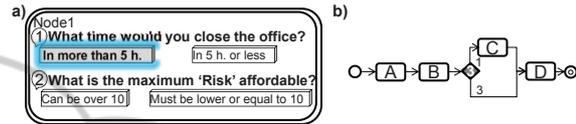


Figure 8: (a) Questionnaire for node 1. (b) The resulting configurable model after removing variants 2 and 4.

4 are removed from the configurable model since they have a time property " $\leq 5$ ". This results in the configurable BP model of Fig. 8(b). The second and fourth configurable nodes of Fig. 7(a) are not depicted in Fig. 8(b) since BP models 1 and 3 belong to the same outgoing branches in these nodes, i.e., the right branch. However, the third configurable node requires to select one of the two branches, and then a new questionnaire is generated. The configuration process continues until only one BP model remains (i.e., representing one specific variant) in the configurable BP model.

## 4 CONCLUSIONS AND FUTURE WORK

The manual individualization of configurable BP models is time consuming and typically requires support by an analyst. Questionnaire-based approaches are suitable methods to support the user while individualizing these models. However, to the best of our knowledge, there is not an automatic method for generating such questionnaires (Rosa et al., 2009).

In this paper an automatic method for generating questionnaires is proposed based on the domain variables of the configurable BP model. The generated questionnaires are proposed to be used to individualize the model during its execution. The initial experimental results over a case of study are promising. As future work we plan to (1) improve the semantics of the questions which are created since they seem too artificial and (2) conduct several case studies to illustrate the feasibility of the proposed method at run-time.

## REFERENCES

- BPMN (2011). Business Process Model and Notation (BPMN), Version 2.0. url <http://www.omg.org/spec/BPMN/2.0/>. [Online; accessed 1-June-2011].
- Breiman, L. (1984). *Classification and regression trees*. The Wadsworth and Brooks-Cole statistics-probability series. Chapman & Hall.
- Dijkman, R., Rosa, M. L., and Reijers, H. A. (2012). Managing large collections of business process models-current techniques and challenges. *Comput. Ind.*, 63(2):91–97.
- Gantt, H. (1913). *Work, wages, and profits*. Engineering Magazine Co.
- Gottschalk, F., van der Aalst, W. M., Jansen-Vullers, M. H., and La Rosa, M. (2008). Configurable workflow models. *International Journal of Cooperative Information Systems (IJCIS)*, 17(2):177–221.
- Hallerbach, A., Bauer, T., and Reichert, M. (2010). Capturing variability in business process models: The provop approach. *Journal of Software Maintenance and Evolution: Research and Practice*, 22(6-7):519–546.
- Jimenez-Ramirez, A., Barba, I., Del Valle, C., and Weber, B. (2013). Generating multi-objective optimized business process enactment plans. In Salinesi, C., Norrie, M., and Pastor, s., editors, *Advanced Information Systems Engineering*, volume 7908 of *Lecture Notes in Computer Science*, pages 99–115. Springer Berlin Heidelberg.
- La Rosa, M., Dumas, M., and ter Hofstede, A. H. (2008). Modelling business process variability for design-time configuration. *Handbook of Research on Business Process Modeling*.
- Rosa, M., Aalst, W., Dumas, M., and ter Hofstede, A. (2009). Questionnaire-based variability modeling for system configuration. *Software & Systems Modeling*, 8(2):251–274.
- Rosa, M. L., Dumas, M., Kaarik, R., and Dijkman, R. M. (2010). Merging business process models. In *18th International Conference on Cooperative Information Systems (CoopIS '10) 2010*, pages 96–113, Crete, Greece. Springer-Verlag Heidelberg.
- Rosa, M. L., Dumas, M., Uba, R., and Dijkman, R. M. (2012). Business process model merging : An approach to business process consolidation. *ACM Transactions on Software Engineering and Methodology (TOSEM)*.
- Rosemann, M. and van der Aalst, W. (2007). A configurable reference modelling language. *Inform Syst*, 32:1–23.
- van der Aalst, W., Dreiling, A., Gottschalk, F., Rosemann, M., and Jansen-Vullers, M. (2006). Configurable process models as a basis for reference modeling. In Bussler, C. and Haller, A., editors, *Business Process Management Workshops*, Lecture Notes in Computer Science, pages 512–518. Springer Berlin Heidelberg.
- van der Aalst, W., ter Hofstede, A., and Weske, M. (2003). Business Process Management: A Survey. In *Proc. BPM*, pages 1–12.
- Weske, M. (2007). *Business Process Management: Concepts, Languages, Architectures*. Springer.