Proceedings of the
First International Workshop on
Visual Formalisms for Patterns
(VFfP 2009)

Visualization of Business Process Modeling Anti Patterns

Ralf Laue and Ahmed Awad

12 pages
Visualization of Business Process Modeling Anti Patterns

Ralf Laue\textsuperscript{1} and Ahmed Awad\textsuperscript{2}

\textsuperscript{1}Chair of Applied Telematics / e-Business
Computer Science Faculty, University of Leipzig, Germany
laue@ebus.informatik.uni-leipzig.de

\textsuperscript{2}Business Process Technology Group
Hasso Plattner Institut, University of Potsdam, Germany
ahmed.awad@hpi.uni-potsdam.de

Abstract:
The most common way to model business processes is to use a graphical modeling language. The most widespread notation are business process diagrams modeled in the language BPMN. In this paper, we formalize structural patterns that can lead to control flow errors in such graphical models. For expressing such error patterns, we use the visual query language BPMN-Q. By using a query processor, a business process modeler is able to identify possible errors in business process diagrams. Moreover, the erroneous parts of the business process diagram can be highlighted when an instance of an error pattern is found. This way, the modeler gets an easy-to-understand feedback in the visual modeling language he or she is familiar with.

Keywords: business process model, business process diagram, BPMN-Q, visualization

1 Introduction

Patterns are used in software engineering to describe reusable solutions for common problems. A prominent example are design patterns [GHJV95], reusable solutions in the field of software design. Patterns have also been used to describe commonly occurring bad practices. These patterns are also known as anti-patterns. In this article, we will formalize structural anti-patterns in business process models (BPM).

BPM are advanced variants of flow charts. Different business process modeling languages share a core set of modeling constructs. Within a BPM, activities can be arranged in sequential order, and routing constructs can be used for modeling alternative and parallel threads.

In the last years, several approaches for detecting errors in the control flow of BPM (for example deadlocks) have been published. Those approaches varied from structural analysis of BPM to the examination of behavioral state space.

There are already several tools that detect problems in BPM. Examples can be found in [Esh02, CK04, Wyn06] – this list is far from being complete. While the problem of detecting errors can be regarded as being solved, many of the tools still fail to give a readable feedback on how to correct the error.
In this paper, we address the presentation of errors in a visual manner. We formalize several error patterns as BPMN-Q queries [Awa07]. When a query is structurally matched by a BPM, the matching part of the model is the part containing a problem.

2 Preliminaries

2.1 Business Process Modelling Notation

There are several visual languages for modeling business processes. The Business Process Modelling Notation (BPMN) is the most widespread language, for this reason we use it for the examples in this paper. However, the techniques described in this paper can also be applied for other languages, because most business process modeling languages share certain basic constructs. In this section, we will shortly describe those basic constructs of the BPMN language. For more details about BPMN, the reader is referred to www.bpmi.org.

Events (something that happens during the lifetime of a business process) are represented by a circle. Although not formally required by the standard, every BPMN model should have at least a start event (depicting the fact that the process is instantiated) and an end event (depicting the fact that the process has been completed).

Activities (tasks that have to be performed) are represented by a rectangle with rounded corners. The flow of control (called sequence flow in BPMN terminology) between the activities is depicted by arcs. The direction of such an arc shows in which order the activities have to be performed.

Gateways can be used for forking and joining paths that have to be performed in parallel or (based on certain conditions) alternatively. There are two kinds of gateways: Splits have more than one outgoing arc, and joins have more than one incoming arc. Gateways are represented by a diamond shape.

When used as a split, an exclusive gateway splits the sequence flow to exactly one of its outgoing branches. When used as a join, it awaits one incoming branch being completed before triggering the outgoing flow. This kind of gateways (which we call XOR-gateways) is depicted by a $\Diamond$ symbol.

When splitting, a parallel gateway activates all outgoing branches; the activities on these branches are executed in parallel. When used as a join, a parallel gateway waits for all incoming branches to complete before triggering the outgoing flow. Parallel gateways are depicted by a $\bigoplus$ symbol. We will refer to this kind of gateways as AND-gateways.

The inclusive gateway is something in-between the exclusive and the parallel gateway. When used as a split, some of the outgoing branches (but at least one) are activated. When merging, an inclusive gateway waits until all active incoming branches have been completed before triggering the outgoing flow. Inclusive gateways (or OR-gateways) are depicted by a $\bigodot$ symbol.

Fig. 1 contains all mentioned kinds of gateways. The model shows a simplified business process in a bank. When a customer applies for a real-estate credit, the customer’s credit rating, the real estate construction documents and the land register record are checked. All these activities are done in parallel, therefore an AND-gateway is used in the model. As the result of those assessments, the application either will be rejected or the contract is to be prepared. The XOR-gateway means that only one of the activities "Reject Application" and "Prepare Contract" can take place. After the contract has been prepared, the process either can end or the bank might offer additional products: a loan protection insurance and a residence insurance. Whether a loan
protection insurance, a residence insurance or both are offered, has to be decided case-by-case. The OR-gateway shows that only one of the activities or both of them (in parallel) can take place.

2.2 BPMN-Q: A Visual Language for Querying Business Processes

BPMN-Q [Awa07, ADW08] is a visual language based on BPMN. It is used to query BPM by matching a process model graph to a query graph.

A BPMN-Q query is represented as a business process diagram that might contain the following additional elements (whose graphical representation is shown in shown in Fig. 2):
(a) **Variable Node**: refers to (unknown) activities in a query.
(b) **Generic Node**: refers to an unknown node in a process. It could evaluate to any node type.
(c) **Generic Split / Generic Join**: refers to any type of split / join gateways.
(d) **Negative Sequence Flow**: states that there is no arc from a node A to a node B.
(e) **Path**: states that there must be a path from a node A to a node B.
(f) **Negative Path**: states that there is no path from a node A to a node B.

The result of a graphical query is given by a sub-graph of the original BPM. An exemplary query and its match are shown in Fig. 3. When matching the process graph in Fig. 3(a) to the query in Fig. 3(b), the result of the query is the sub-graph that contains the nodes B and D and all nodes on the path from B to D. (see Fig. 3(c)).

The query shown in Fig. 3 looks for all paths between B and D. It is also possible to exclude some graph elements from a path search by assigning names to elements and adding an exclude property to a Path edge: For instance, in the query in Fig. 5(a), the XOR-split is named ?s. By adding the exclude property to the path search from ?nd1 to ?j, the search will be limited to paths from ?nd1 to ?j which do not pass the XOR-split ?s.

To process a BPMN-Q query, a query graph is matched to the structure of the business process. A BPMN business process diagram can be defined as a directed typed graph as follows:

**Definition 1** A business process diagram (or process graph) is a tuple \( P_G = (N,A,E,G,F) \) where
- \( N \) is a finite set of nodes that is partitioned into the set of activities \( A \), the set of events \( E \), and the set of gateways \( G \). An event \( e \in E \) is called a start event if it does not have incoming edges. It is called an end event if it has no outgoing edges. All nodes in a process graph are attributed by unique IDs.
- \( F \subseteq N \times N \) is the sequence flow relation between nodes.
The query language BPMN-Q provides additional types of edges between nodes:

**Definition 2** A query graph is a tuple $Q_G = (N_Q, A_Q, E_Q, G_Q, S_Q, P_Q, X_Q)$ where

- $N_Q$ is a finite set of nodes that is partitioned into the set of activities $A_Q$, the set of events $E_Q$, and the set of gateways $G_Q$.
- $S_Q \subseteq N_Q \times N_Q$ is the set of sequence flow edges.
- $P_Q \subseteq N_Q \times N_Q$ is the set of path edges.
- $X_Q \subseteq N_Q \times N_Q$ is the set of negative path edges.

Nodes of the query graph can be identified by assigning labels. Labels can be activity names or special names either starting with "@" for variable activities or with "?" for other nodes and path edges. Thus, a function $l : N_Q \cup P_Q \rightarrow \Sigma^*$ is a labeling function assigning labels (identifiers) to nodes and path edges in the query, where $\Sigma$ is an alphabet.

The labels of nodes can be used within the query in the exclude property of path edges. Thus, $exc : P_Q \rightarrow 2^{\{l(n) : n \in N_Q \cup P_Q \text{ where } l(n) \text{ is defined}\}}$ is a function to evaluate the exclude property of a given path edge.

With the start of query processing, the query processor tries to bind the nodes in the query graph to the nodes in the process graph. For each node in the query graph, the set of nodes in the process graph having the same type are identified. A bind is considered matching if it satisfies all sequence flow edges, path edges, and negative path edges in the query graph. Otherwise, the binding is dropped. For the query to find a match, each node in the query graph must have at least one matching binding. More details about the processing of queries can be found in [Awa07].

### 3 BPMN Soundness Patterns

#### 3.1 Soundness

The most important correctness criterion for BPM is the soundness property, originally introduced by van der Aalst for workflow nets [van97].
For a business process model to be sound, three properties are required:

1. In every state that is reachable from a start state, there must be the possibility to reach a final state (option to complete).
2. If a state has no subsequent state (according to the transition relation that defines the precise semantics), then only events without outgoing arcs (end events) must be marked as being "active" in this state (proper completion).
3. There is no element of the model that is never processed in any execution of the model (no needless elements).

Violations of the soundness criterion usually indicate an error in the model.

3.2 Pattern Catalogs

To our knowledge, the first categorization of error patterns based on the structure of a BPM has been compiled at the University of Osaka. In [OIKK99], five so called deadlock-patterns are discussed. The basic concepts used in [OIKK99] are reachability (in a graph) and transferability (the fact that the control flow will always reach some node in a model if another node has been reached before). The authors of [OIKK99] claim that a model always has a deadlock if one of the patterns can be detected. Unfortunately, this claim is wrong: Fig. 4 shows a sound model for which [OIKK99] would report a deadlock between split node $x$ AND-join node $a$. The reason behind the wrong error report is that by using the concept of transferability, it is not possible to realize that both incoming control flows at join $a_1$ will always synchronize.

In [LK05], Liu und Kumar analyzed how unstructured BPM can be mapped into structured ones with the same behavior. For this purpose, they categorized entries into and exits from a control structure between a split and a corresponding join. In particular, they named the combinations that will lead to control-flow errors.

Koehler and Vanhatalo discuss "typical modeling errors extracted from hundreds of actual process models created in different tools" [KV07]. The anti-patterns discussed in [KV07] are well-known cases for an incongruity between the type of a split and the type of a join. Vanhatalo et al. [VVL07] first decompose a BPM with exactly one start node and exactly one end node into single-entry-single-exit regions. Afterwards, they use a simple heuristic for finding errors in such regions (for example a model is unsound if there is an AND-split in a region but no AND-join).

Mendling [Men07] uses reduction rules for finding errors in Event Driven Process Chains. These reduction rules include information about possible error cases in a model.

All mentioned pattern systems have in common that they include the typical errors that result from a mismatch between the type of a split and the type of a join and from choosing a non-XOR gateway as a loop entry or loop exit. In the next chapter, we will use BPMN-Q for expressing...
Visualization of Business Process Modeling Anti Patterns

these patterns.

4 Anti Patterns Expressed in BPMN-Q

In this section, we express the patterns mentioned in the previous section as BPMN-Q queries. When a query finds a match, the result is the matching sub-graph of the process. This way, the localization of the erroneous part of the BPM is given for free without a need to translate between verification tools and the visual representation of the model.

4.1 (X)OR-split/AND-join Combination

A deadlock can occur when an (X)OR-split opens alternative paths that are later joined by an AND-join. The query in Fig. 5(a) captures this situation between an XOR-split ?s and the AND-join ?j. In order to find two paths from ?s to ?j whose only common nodes are ?s and ?j, we try to find a path from ?nd1 (the successor of ?s) to ?j. This path is given the name ?p1. Similarly, we try to find another path from s to ?nd2. The exclude property of the latter path is set to ?s, ?j, ?p1. Exclusion of ?p1 instructs the BPMN-Q query processor to evaluate path ?p1 first and then to search for other paths which do not share any node with ?p1. The exclusion of ?s, ?j on both paths is necessary to prevent false alarms that can result from loops. Finally, the negative path from ?nd1 to ?nd2 prevents false alarms in cases like the one shown in Fig. 5(c).

4.2 Entry Into a Parallel Control Block

Usually, in a block that starts and ends with an AND-gateway, there is no chance for a deadlock, because all incoming branches of the AND-join have been activated before. However, if on a path from the split to the join there is an (X)OR-join that can receive activations not originating from the AND-split, a deadlock can occur. We call this (X)OR-join an entry into the AND block. This situation is captured declaratively in Fig. 6. The use of two different start events, ?ev1, ?ev2, forces the query processor to find matches in BPMs having more than one start event.
4.3 AND-Join as an Entry Into a Loop

The query in Fig. 7 describes another situation where a BPM could suffer from a deadlock. Whenever an AND-join is part of a loop where only a subset of its input points are activated, a deadlock occurs. This is declaratively represented by a path edge from \(?j\) to \(?in1\) and a negative path edge from \(?j\) to \(?in2\).

4.4 AND-Join After (X)OR-Split Does Not Synchronize

A deadlock can occur if an AND-join awaits to be activated on all its incoming arcs, but an
(X)OR-split before this AND-join can lead the flow of control away in another direction. The query in Fig. 8 describes this situation.

4.5 AND-split/XOR-join Combination

Lack of synchronization is a modeling error that occurs whenever an AND-split is combined with an XOR-join. All outgoing branches from the AND-split will be activated. However, the semantics of the XOR-join is to wait for the completion of exactly one of its incoming branches. Due to space limitations, we do not provide a separate query for this type of errors. Rather, we can reuse the query in Fig. 5(a) with modifications: We switch the roles of the XOR and the AND in that query. Also, we drop the negative path between \(nd1\) and \(nd2\). The resulting query captures the "lack of synchronization errors".

4.6 Infinite Loop

Another modeling error occurs when a part of the process is activated infinitely often. This can happen when a modeler mistakenly represent a loop exit with an AND-split rather than an XOR. The query in Fig. 9(a) describes this situation: Path edge \( p \) from the AND Split \( s \) back to itself represents the looping case. To prevent reporting false alarms, e.g., the case of a parallel block nested in a loop, the query requires to have a totally distinct path from \( s \) to an end event.

5 Using the Patterns

As common BPM languages share the most basic modeling elements, the application of our patterns is not restricted to the language BPMN. Previously, the patterns have been successfully used for the language Event-Driven Process Chains [van99]. For this purpose, they have been implemented into the modeling tool bflow\(^1\), using the openArchitectureWare Check language\(^2\) [KKGL08]. Another implementation, using a larger repository of 23 patterns, has been made using logical reasoning with Prolog. In [GL09], it has been shown that a pattern-based heuristic reasoning detects control-flow errors almost as accurate as model checkers which explore the whole state space of a model. However, other than approaches using model checkers, it does not suffer from the state-space explosion problem.

\(^1\) http://www.bflow.org
\(^2\) http://www.eclipse.org/gmt/oaw/
To test our approach based on graphical BPMN-Q queries, we searched for the patterns discussed in the previous section in 109 models taken from the public repository of the modeling tool Oryx (www.oryx-editor.org). Table 1 shows how many models were detected that contained instances of the patterns. It also shows the false alarms, i.e. the cases for which the model does not suffer from the discussed problem despite matching the query. The query processor ran on a PC with 2GB RAM and an Intel dual core processor at 1.83 GHz under the operating system Windows XP with Service Pack3.

An observation that is worth mentioning is that all false alarms occurred because of other errors that are located elsewhere in a model. An example is shown in Fig. 10. It matches the BPMN-Q query AND-split/XOR-join Combination, because there are two distinct paths from the AND-split to the XOR-join. However, at execution, this error would never occur because the process would either terminate without problems or deadlock before reaching the XOR-join. Of course, this possible deadlock will be detected by the pattern AND-Join After (X)OR-Split Does Not Synchronize. In situations like this, our pattern-based approach delivers too many warnings (which is better than missing a legitimate warning).

<table>
<thead>
<tr>
<th>Anti-pattern</th>
<th>Erroneous Models</th>
<th>False Alarms</th>
<th>Processing Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X)OR-split/AND-join Combination</td>
<td>4</td>
<td>0</td>
<td>208.735 sec</td>
</tr>
<tr>
<td>Entry Into a Parallel Control Block</td>
<td>0</td>
<td>0</td>
<td>237.515 sec</td>
</tr>
<tr>
<td>AND-Join as an Entry Into a Loop</td>
<td>1</td>
<td>0</td>
<td>219.453 sec</td>
</tr>
<tr>
<td>AND-Join After (X)OR-Split does not Synchronize</td>
<td>7</td>
<td>0</td>
<td>221.250 sec</td>
</tr>
<tr>
<td>AND-split/XOR-join Combination</td>
<td>2</td>
<td>5</td>
<td>212.391 sec</td>
</tr>
<tr>
<td>Infinite Loop</td>
<td>0</td>
<td>0</td>
<td>229.578 sec</td>
</tr>
</tbody>
</table>

Table 1: Results of applying anti-pattern queries
6 Related Work

While soundness is necessary for the correctness of a BPM, it does not yet guarantee that the model really conforms to the business rules. For this reason, several researchers applied model-checking in order to verify statements like "Each delivery must always be preceded by a payment". These approaches make it necessary to specify the properties to validate as temporal formulas. Usually, this is too difficult for the business people who work with the models.

As an alternative, several authors developed notations based on a process modeling language to express the allowed executions of a BPM. Such approaches have been presented for Event-Driven Process Chains [Rum99, SM06, FF08], BPML [Bra05], UML Activity Diagrams [FESS07] and for BPEL specifications [WKH08, LMX07]. Quartel et al. [QDS05] use the Interaction Systems Design Language for expressing dependencies among (distributed) business processes. Van der Aalst and Pesic [vP06] suggested a new language DecSerFlow for specifying the properties of a single service or service compositions. As a different stream of research, Barros et al. [BDG07] and Rommelspacher [Rom08] suggested graphical languages for expressing complex events in business processes. The main difference between those graphical languages and BPMN-Q is that BPMN-Q is used to formulate queries about the business process model itself (i.e. its graphical structure), not about the state space of its executions. This makes it possible to use BPMN-Q for searching for modeling problems without having to compute the state space of all possible executions. Another query language that is working on the graphical structure of a model is BPMN VQL [FT08]. Its main purpose is to find crosscutting concerns in BPM. Our language BPMN-Q is more expressive than BPMN VQL. For instance, generic nodes, negative paths, and variable names in BPMN-Q have no equivalent constructs in BPMN VQL.

7 Conclusion and Directions for Future Research

The approach presented in this paper can be used for detecting control flow errors in business process models. By using BPMN-Q queries, it is possible to give a business process modeler a feedback not only about the presence of errors but also about the part of the model that causes the error. This information is given in the visual formalism the modeler is familiar with.

So far, more advanced BPMN constructs like exception handling have not yet been considered in the graphical language. Inclusion of such constructs would enable us to express even more complex patterns like the ones published in [GL07] and [RPH08].

Another direction of future research can be to reduce false warnings in cases like the one shown in Fig. 10 where a model contains more than one instance of one of our error patterns.

Bibliography


