Model Matching Challenge: Renaming Elements

Pit Pietsch∗
Software Engineering Group
University of Siegen
pietsch@informatik.uni-siegen.de

Abstract: The following example illustrate a low-quality correspondences produced by a generic model differencing tool. The example is originally used in [1] to discuss limitations of generic matching approaches where the heuristics cannot be configured. It consists of two Business Process Model and Notation Version 2.0 (BPMN2) diagrams which model a simple shipment process. The diagrams were created with the BPMN2 Visual Editor for Eclipse under a Windows 7 OS. This contribution is considered as a challenge for model matching algorithms.

1 Description of the Example

Figure 1: Variants of a simple shipment process

There are no structural changes Between Version A and Version B in Figure 1, i.e. no edit

∗P. Pietsch is supported by the DFG under grant KE 499/5-1.
steps that add or delete model elements were applied; instead all tasks in Version B have been renamed.

2 Actual and Expected Results

Why this example can be considered as a challenge for state-of-the-art model matching algorithms becomes clear when the models are compared with EMF Compare [1], a well-known generic model differencing tool integrated in the Eclipse IDE.

Figure 2 shows the computed difference between Version A and Version B. Only one renaming, i.e. the renaming of task \textit{Check Order} to \textit{Receive Order}, is recognized correctly. The other four renamings were not recognized and instead deletions and creations of the tasks are reported. Furthermore all sequence flows were matched, even if their source and target elements are reported as not corresponding.

![Figure 2: Difference between version A and B reported by EMF Compare](image)

Arguably, the computed difference is of low quality and does not represent the actual change. The expected result is that all tasks and sequence flows are matched. The example and the expected result can be downloaded at [2].

\footnote{\textsuperscript{1}It is assumed for the following example that EMF Compare is configured to compare models only on the basis of similarities.}
3 Cause of Low Quality Results

The reason for the low quality of the difference is that the EMF Compare matching algorithm puts a strong emphasis on the names of elements. This assumption works well for class diagrams and similar model types, where almost all model elements have names and can often even be uniquely identified based on them. Conversely, this heuristic fails for model types where elements often have no or only generic names. BPMN2 diagrams are even more complicated in this regard, because they often contain unnamed elements that also lack additional local properties which can be used in the similarity computation, e.g. Sequence Flows. Hence, model differencing tools for process models have to take the environment of such elements into account to perform a reliable matching.

References
