The ICoP Framework for Business Process Model Matching

Matthias Weidlich

The William Davidson Faculty of Industrial Engineering and Management
Technion – Isreal Institute of Technology
weidlich@tx.technion.ac.il

Keywords: Process Modelling, Model Analysis, Change Management

1 Introduction

In order to identify operational commonalities and differences, organisations compare business processes. Such comparison is driven, for instance, by a merger of organisational units or by standardization efforts that require an assessment to which extent operations conform to company-wide or industry-wide standards. Once business processes are documented by process models, *process model matching* evolved as a means to support comparative analysis of operations. Here, matching refers to the identification of correspondences between activities of two process models. Tools supporting the matching step are called *matchers*. In this abstract, we present the ICoP framework, which supports the development of such matchers.

2 Challenges in Business Process Model Matching

It is a major challenge for matching business process models that those models rarely use the same level of detail. Also, variability with respect to the textual description of functionality is commonly observed. For illustration consider a business process model that comprises an activity \`Check Invoice'. Another related model may split up this activity into several separate steps, e.g., \`Verify Customer Data', \`Decide on Correctness of Data', and \`Approve Invoice'. Those differences relate to refinement and meronymy. Further differences stems from synonymy and homonymy. That is, one process model may describe an activity as \`Check Invoice', whereas another model may relate to this activity with the label \`Verify Bill'. Organisations rarely align their documentation of business processes. Consequently, heterogeneity in the representation and description of operations are likely to be observed in practice and hinder any comparative analysis [1].

Differences related to refinement and meronymy impose particular challenges for the identification of correspondences. In contrast to a setting that shows only $1:1$ correspondences, it is not feasible to analyse the whole set of potential correspondences due to the implied combinatorial problem. Consider the possibility of $1:n$ correspondences between two process models with n and m activities. Then, the number of 1:x matches is given by $n * \binom{m}{x} + m * \binom{n}{x}$ with $\binom{n}{x}$ as the binomial coefficient that defines the number of x-element subsets of an n-element set. As such, the number of possible combinations makes exploration of the whole set of potential correspondences infeasible and calls for heuristics that find suitable candidate correspondences.

The problem of matching business process models is close to matching of data schemas or ontologies [3,4]. For several decades, work in this research area has combined structural analysis with natural language processing to identify correspondences between entities of data schemas or ontologies. However, those works show a predominant focus on 1:1 matches, such that *`1:n and n:m mappings [..] are currently hardly treated at all'* [3]. Further, the few existing exceptions cannot be applied for process model matching since they have been tailored for data models and, e.g., exploit data instances for matching.

3 The ICoP Framework

To solve the problem of matching process models and cope with complex correspondences in particular, we presented in the ICoP framework [5]. The framework proposes an architecture and a set of re-usable components for assembling matchers.

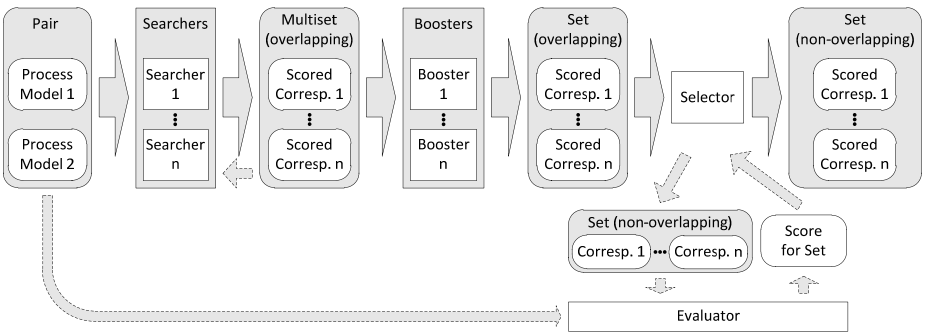

Figure 1: The ICoP architecture.

The overall ICoP architecture is illustrated in Figure 1. It defines a multi-step approach to cope with the combinatorial challenges outlined earlier. Given two process models, *searchers* extract potential correspondences. Search heuristics implemented within this framework analyse the labels of activities (e.g., by a vector space scoring) and apply different heuristics to group activities (e.g., based on fragments obtained by applying graph decomposition to the process model). Second, scored potential correspondences are conveyed to *boosters* that aggregate correspondences and adapt their score. For instance, 1:n correspondences that subsume other identified 1:1 correspondences will receive an increased matching score. A *selector* builds up the actual set of correspondences from the set of candidates. It selects the best candidates that are non-overlapping. This selection is guided in two ways: the scores of the candidates and an evaluation score. The latter is derived from an *evaluator*, which assigns a single score to a set of selected correspondences. This score may be derived with knowledge about the original process models. The selection of correspondences is done in an iterative manner. In each step, the selector selects a set of correspondences and the evaluator assigns a score to this set, which is then used within the selector to modify the set of correspondences. Upon completion of the selection procedure, the final set of correspondences between activities of the process models is produced.

4 Conclusion

In this abstract, we gave an overview of the ICoP framework to address business process model matching. It provides a flexible and adaptable architecture for the implementation of matchers along with a set of predefined components for assembling matchers. Experimental results on the evaluation of the framework can be found in [5]. Those indicate that the framework is able to identify a significant amount of complex correspondences between business process models. Still, those experiments also show that a certain similarity of labels is needed to achieve good results. Also, in [5], we highlighted how existing matching techniques, e.g., based on the graph edit distance as proposed in [2], may be incorporated in the framework.

In future work, we aim at extending the set of matching components available within the ICoP framework. Part of speech tagging has recently successfully be applied to labels of process models, cf., [6], which suggests to leverage this information also for process model matching. The precision of generating candidate correspondences may benefit from an object or action aware handling of activity labels. In addition, external knowledge such as WordNet shall be incorporate in the matching procedure. The benefits of taking external knowledge into account have been demonstrated for matching data schemas [7]. Note that external knowledge for the domain of business processes is available in terms of reference models like the MIT Process Handbook.

1. Dijkman, R.: Diagnosing differences between business process models. In: Proceedings of BPM. (2008) 261-277
2. Dijkman, R., Dumas, M., García-Bañuelos, M.: Graph Matching Algorithms for Business Process Model Similarity Search. BPM 2009:48-63
3. Rahm, E., Bernstein, P.A.: A survey of approaches to automatic schema matching. VLDB Journal 10(4) (2001) 334-350
4. Euzenat, J., Shvaiko, P.: Ontology matching. Springer-Verlag (2007)
5. Weidlich, M., Dijkman, R., Mendling, J.: The ICoP Framework: Identification of Correspondences between Process Models. CAiSE 2010:483-498
6. Leopold, H., Smirnov, S., Mendling, J.: Recognising Activity Labeling Styles in Business Process Models. EMISAIJ 6(1):16-29 (2011)
7. Dhamankar, R., Lee, Y., Doana, A., Halevy, A., Domingos, P.: IMAP: Discovering complex semantic matches between database schemas. In: SIGMOD (2004) 383-394