Evaluating the Evaluation of Process Matching Techniques

M. Sc. Tom Thaler, B. Sc. Philip Hake, Privatdozent Dr. Peter Fettke, Prof. Dr. Peter Loos

Institute for Information Systems (IWi) at the German Research Center for Artificial Intelligence (DFKI) and Saarland University, 66123 Saarbrücken, E-Mail: (tom.thaler | philip.hake | peter.fettke | peter.loos@iwi.dfki.de)

Abstract

The matching of different process models and their nodes plays an important role, as shown in the manifold matching techniques developed during the last years. A well-established approach for the determination of those techniques' quality is the consideration of precision and recall values related to a reference matching. Nevertheless, it is remarkable that it is not clear, what such a reference matching should be and how to reach a general acceptance. As developing a reference matching requires the decision and a consensus on what a matching should represent, we conceptualize that task as a decision problem, which can be well-structured or even ill-structured. The paper evaluates the evaluation of process matching techniques by theoretical argumentations and in terms of inquiring the process of reference matching development. Based on the results, the authors propose some guidelines containing the three phases *idea*, *definition* and *criteria* supporting that task.

1 Motivation

Business process models are core artifacts of the current information systems research and practice. Organizations have large model repositories containing hundreds or even thousands of models [4, 7], which serve as knowledge base for process execution and further business process management activities. Managing such a great many of process models leads to the need of effective and efficient methods for comprehensive process analyses.

Some applications of these analyses are conformance checking, reusability of models or model fragments, company merges and much more. For these applications it is important to have knowledge on the analogy between process models or their nodes and thus, to have knowledge on their correspondence. Against that background more and more matching techniques were developed during the last years (e. g. [1, 12]). The high relevance of those techniques is also shown by the existence of a Process Matching Contest in context of the BPM Conference 2013 [3], where authors were invited to develop a process matching technique and apply it to two given model repositories. As it is an established approach, evaluating the quality of matching techniques with precision, recall and f-measure values, the contest provided an abstract of a reference matching, which was then used for rating the submitted algorithms.

Evaluating the capacities of particular algorithms by precision and recall and adequate reference solutions is well probed in the area of information retrieval [2]. Like done in the mentioned process matching contest, also other papers in the field of business process model matching try to assign these methods to the context at hand, but an evaluation as well as a discussion on the adequacy of their application is still missing.

Against that background and while analyzing the given reference matching from the contest, it is no surprise that it is sometimes not clear, why a node was matched to another node in the way it is, while oneself would prefer another matching. These different *human matching behaviors* show, that it is not clear what a reference matching should be and how to reach a general acceptance of a reference matching. A missing consensus denotes an unclear definition of a reference matching and thus, leads to misinterpretations of the quality measurement results. Concerning that unclear definition we consider the development of a reference matching as a decision problem, which is sometime well-structured and sometimes ill-structured. As evaluating matching techniques by precision and recall is based on the assumption of an existing and unique reference matching, a major issue is the fact, that ill-structured decision problems are not solvable as is and well-structured decision problems may have several solution.

Against that background, this paper aims at evaluating the evaluation of process matching techniques in general and the development of reference matches in particular. As we do not agree with the assumption of an existing and unique reference matching, we discuss both aspects, namely the existence and the uniqueness. Finally we propose some guidelines containing the phases *idea*, *definition* and *criteria* to address the located defects.

The fundamentals of the topic (especially process matching, decision problems and measuring the quality of process matching techniques) are being clarified in section 2. Section 3 discusses the existence of reference matches based on the fundamentals of decision theory, while section 4 discusses the uniqueness of them based on evaluating the challenges of establishing a reference matching in a real matching scenario. In section 5, the authors propose some guidelines supporting the task of reference matching development, while section 6 demonstrates these guidelines by applying them to the use case of section 4. The paper closes with a conclusion in section 7.

2 Fundamentals

2.1 Structuredness of Decision Problems

Well-structured decision problems have distinct solutions and can be solved easily. However ill-structured decision problems contain defective components within the decision model, which are attributed to a lack of information [9]. An ill-structured decision problem contains at least one defective component and cannot be solved easily. Defects are (1) *perception*: The decision problem is defective in the way that the perception of the decision problem is defective. The problem is not considered as a decision problem and therefore further steps solving the decision problem are not taking into account. (2) *Differentiation:* A decision problem is ill-structured if not all possible solutions are known or the known alternatives cannot be distinguished since they overlap. (3) *Effect:* If the different solutions are known but the effect of an alternative cannot be predicted the decision problem can be considered ill-structured. (4) *Evaluation:* A defective component of evaluation exists if the result of the decision taken cannot be measured according to its influence on reaching the predefined objective of the decision problem. (5) *Objective:* If some objectives of a decision problem are unknown the decision problem can be considered ill-structured. (6) *Solution:* A decision problem suffers from a defective solution if the known solutions are not considered to be efficient.

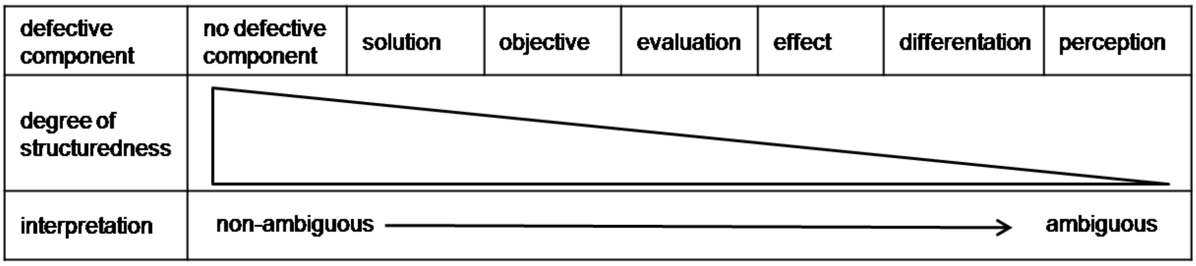

Figure 1: ill-structured decision problems [9]

To obtain a solution the ill-structured decision problem must be firstly converted to a structured decision problem. Figure 2 depicts the different types of defective components ordered descending according to the influence on the structuredness of a decision problem. The figure shows that a decreasing degree of structuredness leads to an increasing ambiguous interpretation of how the defective component can be resolved. Therefore structuring ill-structured decision problems can result in different feasible structured decision problems.

2.2 Process Matching

Matching is described as the process that takes two schemas as input, referred to as the source and the target, and produces a number of matches between the elements of these two schemas based on an particular correspondence [10]. Thereby the term schema has a broad interpretation and can comprise database schemas (e. g. [6]) as well as arbitrary other model schemas.

The term process matching can be divided into two different fields – matching process models (1) and matching nodes of process models (2). Matching process models means, that models are matched to other models based on criteria like similarity, equality or analogy. Mostly different process model repositories serve as a base for that task. An application scenario in that field is the merging of companies, where it is necessary to synchronize different processes, e. g. in context of administration or acquisition.

Matching nodes of process models, which is the focus of the paper at hand, describes the matching of single nodes, set of nodes or node blocks of one model to the corresponding elements of another model. Such matches are used for harmonizing business process models or inductively derive reference models from different individual models. The determination of matches between process models (1) is often based on a node matching as described in [1, 12]. While [1] presents 19 different similarity measures for business process models with their underlying, mostly 1:1 node matching techniques, [12] develops a similarity measure for process models based on M:N node matches. The cardinality describes the cardinal number of node sets, which are being matched to each other. A sample of a node matching with both, 1:1 and M:N matches is visualized in Figure 2.

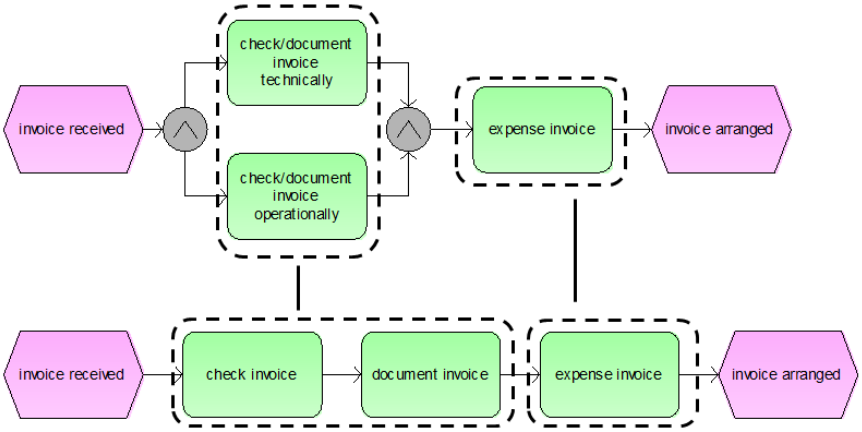

Figure 2: Node matching example

Generally a matching can be observed by the *human matching behavior*. The term describes all matches generated by individuals, without a consideration of their matching intention.

Formally a node matching is formalized as follows, whereby we allude to prescind from a concrete schema or modeling language, as that aspect is of minor importance in the context at hand. We further argue, that it is possible to adapt this generic formalization to an arbitrary modeling language.

Definition 1 (Elementary/Complex Node Matching) [12]. Let $G = (N, E)$ be a business process model with: N is a non-empty set of Nodes and E is a non-empty set of edges. Let A_1 be a subset of nodes of N_1 and A_2 be a subset of nodes of N_2. A match $m \in M$ is denoted by a tupel (A_1, A_2) of two sets of activities. A match (A_1, A_2) is called *elementary match*, iff $|A_1| = |A_2| = 1$ and *complex match*, iff $|A_1| > 1 \vee |A_2| > 1$.

The given definition 1 is not a definition in the strong sense but a formalization. This is founded in the fact, that formal criteria for a match are not given. Therefore, the definition only covers the formal border of a match, while neither necessary nor sufficient criteria are specified. In fact, the instantiation of a match is considered as a decision problem, which can be well-structured or ill-structured. The analysis of real matching problems with regard to the decision theory allows us both to explicate the main defects on the decision how to match and to develop proposals handling these defects.

2.3 Quality of Process Matching Techniques

According to prior work [2], a fundamental requirement for the evaluation of matchtes is the existence of a benchmark. Thereby a benchmark is based on the idea of evaluation scenarios in terms of expected solutions as a basis for comparison. In the following, these expected solutions are called *reference matches*.

Four metrics that have been used extensively in the area of matching technique evaluation are precision, recall, f-measure and fall-out, which are all intended to quantify the proximity of the results generated by a matching technique to those expected [5]. Thereby, the precision, recall and f-measure are of major importance in the domain of information retrieval [11] and, thus, also established in the context of evaluating process matching techniques [8].

Precision is the fraction of found node matches, that is correct in terms of the reference matching. Recall is the fraction of the correct node matches, that are found [12]. The f-measures is the harmonic mean between precision and recall and fall-out is the rate of incorrectly discovered matches out of the number of those non-expected. We classify each match either true-positive (TP), true-negative (TN), false-positive (FP) or false-negative (FN) as visualized in the following Figure 3 [3].

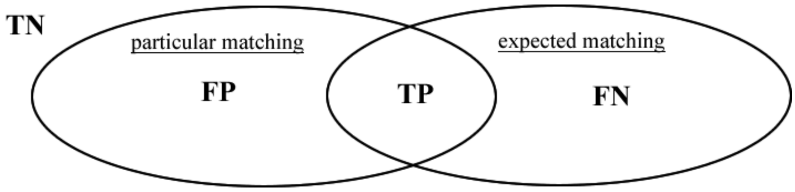

Figure 3: Relevant sets of matches for precision, recall and f-measure calculation

Definition 2 (Precision, Recall, F-Measure). A matching M is a set of matches m (as described in definition 1). Let M_p be a particular matching and M_r be a reference matching, so that:

- TP (true positive) is the intersection of M_p and M_r: $TP = M_p \cap M_r$
- TN (true negative) is the complement of M_p union M_r: $TN = \overline{M_p \cup M_r}$
- FP (false positive) is the difference of M_p and M_r: $FP = M_p \backslash M_r$
- FN (false negative) is the difference of M_r and M_p: $FP = M_r \backslash M_p$

Then $precision = \frac{|TP|}{|TP|+|FP|}$, $recall = \frac{|TP|}{|TP|+|FN|}$ and $fmeasure = \frac{2\, x\, precision * recall}{precision + recall}$ is the harmonic mean of precision and recall.

Note, that these evaluation criteria are based on the assumption that the set M_r is given.

3 The Non-Existence of Reference Matches

How mentioned before, matching nodes of process models is considered as a decision problem, which can be well-structured or ill-structured. Ill-structuredness in that area is often founded in a missing objective, which leads to the problem, that it is not possible to rate different alternatives (defect in effect) and thus, it is not possible to decide on a particular match. According to the decision theory, these cases are considered as ill-structured decision problems, which have no solution by definition. This leads to the finding, that a reference matching does not exist in all cases. Therefore, the concept of evaluating process matching techniques by precision and recall related to a reference matching collapses in case of ill-structured decision problems, as it postulates the existence of such a matching.

Articles in the field of business process matching mostly postulate the existence of a unique reference matching, which is then often developed in an “ad-hoc” manner [3]. This perception disagrees with the achievement of established literature [2]. In fact, reference matches (TP+FN) result from human matching behavior. Thus, two sets of nodes are matched, if there is at least one human, who matches these sets. In many cases different people match nodes in different ways and sometimes they are not able to decide how to match two particular sets of nodes (ill-structured decision problem). Therefore, those matches are ambiguous. One should not only distinguish between *match* and *not match*, but also take *unsure* into account. As recent works postulate the existence of a unique reference matching, it is acted on the assumptions, that a matching is (1) independent from the matching person, (2) independent from time and (3) independent from the matching objective. As we argue, that different people match nodes in different ways, and as the decision for or against a match may vary over time (e. g. dependent on the knowledge) as well as the dependency of a match on a concrete evaluation or application scenario, these assumptions are not supportable.

4 The Non-Uniqueness of Reference Matches

4.1 Approach

In order to discuss the uniqueness of reference matches we conduct the following matching scenario. Two different model repositories [3], respectively containing nine models were delivered to three experts in modeling. The experts conducted a discourse on how a reference matching could be instantiated for one model of a repository to the other models of the same repository. Thereby, the first repository contains process models describing the admission process of different German universities; the second repository contains process models describing the birth registration process of different countries. Thus, all considered model pairs describe a consistent domain. While discussing the possibilities of instantiating particular matches, the authors faced different challenges. Although we postulate that every single expert has the same knowledge of the given process models, the argumentations of all experts were quite different, nevertheless all arguments were reasonable and comprehensible. Therefore, in many cases a consensus has not been reached, because the different argumentations could not be falsified. These cases are being illustrated, whereby the major issues of matching nodes are noted, subsequently analyzed, abstracted and categorized.

4.2 Illustration of Specific Issues

Extent of a match. Regarding the admission processes of Berlin and Erlangen (Figure 4), we recognized different understandings of correspondence. Matching the functions *send application* turned out to be clearly when assuming equality as correspondence. Then again the matching of the antecedent functions proved to be rather difficult. Considering a strict (exhaustive) equality as correspondence only *add certificate of bachelor degree* is matched. Although only these two functions are considered to be equal, all of the functions seem to overlap in the way that documents are added to the application process. Thus they can be matched if we assume a certain overlapping in meanings as correspondence.

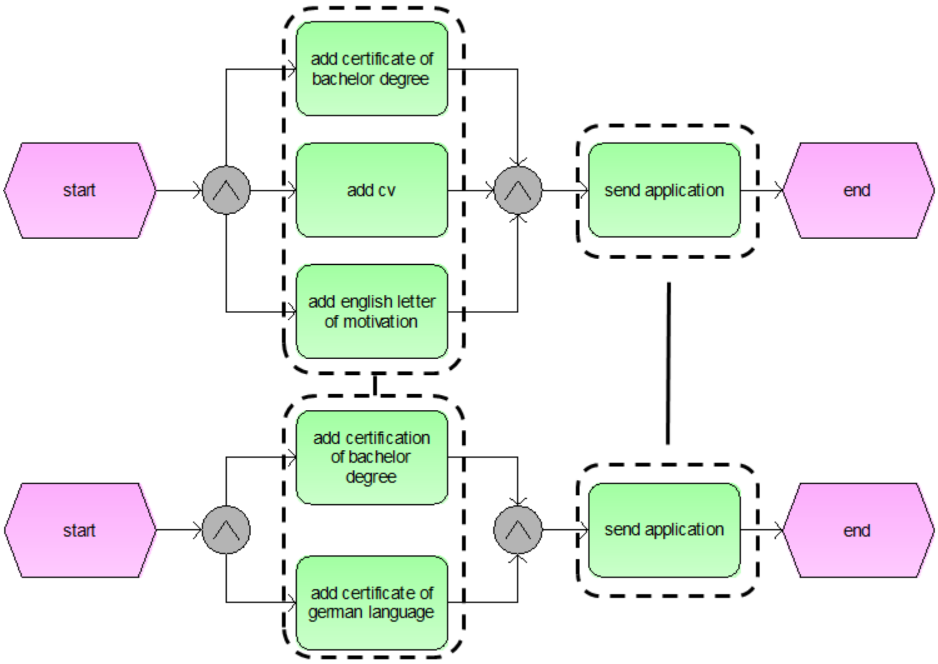

Figure 4: Exhaustive equal and overlapping (Admission processes: Berlin - Erlangen)

Figure 4 depicts a transitive matching of overlapping nodes. The functions described cannot be considered as equal but they share common meanings, hence, they can be described as overlapping matches. The function *check complete* and *complete and in time* can be matched because both perform a completeness check. Furthermore *complete and in time* can be matched to *check valid* according to an overlapping since only a document received in time is a valid document. Other validation steps of *check valid* overlap with the evaluation of *evaluate and decide about acceptance*. Finally *decide about acceptance* can be matched to *evaluate and decide about acceptance*.

On the one hand one could aggregate the discovered overlapping matches by matching the function *evaluate and decide about acceptance* to *check complete* as well as *decide about acceptance* to *complete and in time*. This leads to the fact that some nodes of the match do not overlap at all. On the other hand the matching could avoid a matching of functions that do not overlap. In this way several matches of the process models given are possible. For example the matching could be ({*check complete, check valid}, {complete and in time*}) and ({*decide about acceptance}, {evaluate and decide about acceptance*}) or ({*check complete}, {complete and in time*}) and ({*check valid, decide about acceptance}, {evaluate and decide about acceptance*}).

The defect of the described issue related to the decision theory ranges from differentiation to effect. It could be argued, that there are overlapping solutions, which lead to an indistinguishability and makes the decision problem ill-structured. At the same time there is another defect, as the effect of choosing one of the available solutions is not predictable.

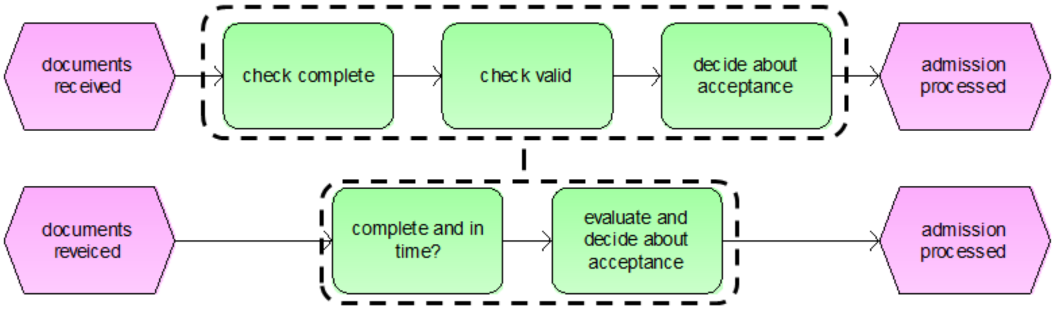

Figure 5: Concatenation overlapping matches

Interpretation of equivalence. Interpreting a match allows the argumentation due to an equivalence of actions or an equivalence of objectives. Figure 5 depicts an application process holding functions that share the same objective but not the same action – there are different processes for choosing adequate applicants. While one university arranges assessment centers, others prefer taking interviews. Thus, they do essentially different things for achieving the same objective, namely the choice of adequate applicants. One could now argue against a match, as the functions do not represent the same tasks. Otherwise, one could argue for a match, as the objective – choosing an adequate applicant – is equal in both processes.

The problem here is the unclarity of objective. As described in section 2.1, an unknown objective makes a decision problem ill-structured, as it not clear how to act in particular cases like this.

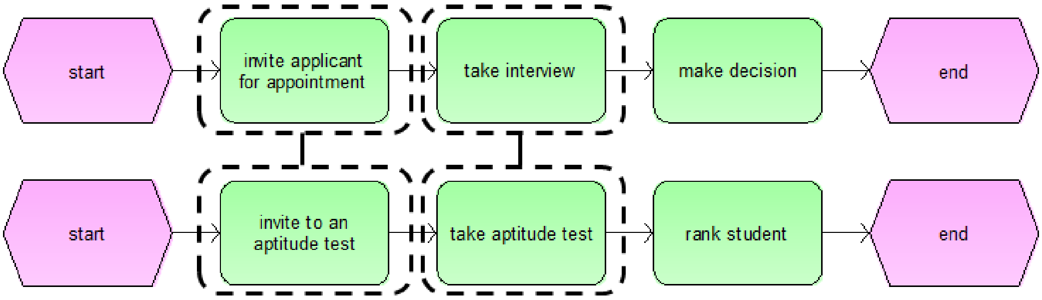

Figure 6: Matching equivalent objectives (Admission processes: Erlangen - Cologne)

Interpretation of meaning. Considering the admission processes of a German and an English university, we identified the German university testing the German language skills of applicants, while the English university checks English skills. Thus, as both universities test different languages, the corresponding nodes in the process model are not equal. This argumentations leads to the decision, that there will no matching established for these two activities. Nevertheless, they are analog since both test the national language skills, what again militates for the matching of both activities. The defect of that issue can also be seen in terms of a missing objective.

4.3 Intermediate result

The illustration of specific issues showed, that in many cases it is not possible to reach a consensus on particular matches, moreover the human matching behavior is diverging and sometimes an individual is not sure how to match. If the human matching behavior is divergent, there are good arguments to match nodes in a specific way and there are also good arguments to match nodes in a different way. In most cases the defect is founded in a missing objective for the matching, which leads to a consideration as an ill-structured decision problem. As ill-structured decision problems have no solution, also a generally accepted reference matching does not exist without further manifold and context-sensitive hypotheses (e. g. in terms of an evaluation or application scenario). Acting on those hypotheses may transform an ill-structured decision problem to a more well-structured one. In fact, well-structured decision problems are solvable, but they may have several solutions. Thus, if a reference matching exists, it is not unique in all cases. Therefore the assumption of a unique reference matching in the field of reference matching development is not supportable.

We further argue, that a reference matching and thereby also the set M (Definition 2) does not exist a-priori (before developing), but ex-post. Since a reference matching obviously depends on further assumptions and the consensus of a set of experts, a reference matching cannot exist till it is developed. E. g. if a word is being translated to another language for the first time, it is not possible to act on the assumption, that the translation was available before.

5 Reference Matching Development Guidelines

5.1 Overview

To address the found defects, we will now propose some guidelines supporting the process of reference matching development. We already identified different aspects with important roles in that area. These aspects include the definition of a context in terms of an evaluation or application scenario as well as a methodology to reach the acceptance of a reference matching.

The guidelines (Figure 7) consist of the three phases, *idea* (1), *definition* (2) and *criteria* (3). Idea describes several different possibilities of interpreting a matching and thus, allows a transformation of ill-structured decision problems of process matching to more well-structured ones. Definition (2) instantiates these ideas, while criteria (3) describe different possible requirements establishing a concrete match.

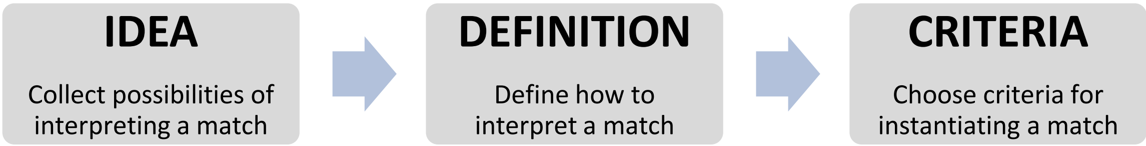

Figure 7: Reference Matching Development Guidelines

This concept of idea, definition and criteria allows us both to discuss different essential aspects of reference matching development in a closed manner and varying these aspects in context of different evaluation and application scenarios.

5.2 Idea

In the following we derive some ideas specifying a reference matching from the issues presented in section 4. These ideas describe a set of definition aspects which should be selected in terms of defining a reference matching and is the basis for transforming ill-structured decision problems to more well-structured ones. In the following no algorithm is presented but some general principles to cope with the challenges of different matching contexts.

Model constructs. Independent from the process modeling language the question arises what elements of a model should be matched and what matching strategy needs to be applied on what element type. Generally one could distinguish between the matching of nodes and edges. Considering e. g. EPCs, related elements like functions, events and others which hold a label can intuitionally be matched regarding their labels. But when it comes to connectors the label based matching technique cannot be applied. Furthermore in different contexts it can be meaningful to match only particular node types, while others are being ignored.

Extent. The extent of a matching can be described by the following characteristics. (1) *Exhaustive:* A simple way to match nodes is based on the equality of two nodes. This means that only nodes that hold the same meaning are allowed to be matched. Given two functions F_1 and F_2, a match of F_1 and F_2 is only valid if the activities represented by the functions are considered as exhaustively equal, so that one function does not contain more content than the other. This might even include that the activities are carried out by the equivalent organizational unit. (2) *Overlapping:* Another approach to match nodes covers the matching of nodes that are similar in the way that they share at least a common meaning. This approach is a relaxation of the strict equality postulated in the antecedent approach. (3) *Concatenation:* A further relaxation leads to an approach which allows transitive matches of nodes. Given the nodes N_1, N_2 and $N_{3,}$ $\{N_1,N_3\}$ is a transitive match, iff $\{N_1,N_2\}$ and $\{N_2,N_3\}$ represent overlapping matches. Matches which contain transitive matches are referred to as thematic concatenation.

Equivalence. Due to the interpretation of a match it can be differentiated between an equivalence of action and an equivalence of objective. Equivalence of actions means nodes or (sub)paths representing the same action(s) for achieving an objective, while equivalence of objective means, that only the objectives are equal.

Meaning. One should also differentiate between the analogy and the equality of meanings. Considering the example, mentioned in section 4.1, it is often easy to check the analogy by abstracting. Thus, “check German language skills” as well as “check English language skills” can be abstracted to “check national language skills” in that context.

5.3 Definition

In the definition phase, the characteristics of a reference matching in context of a particular evaluation or application scenario are specified. This specification is based on the ideas introduced in the prior section and thus, explains what a matching should represent. The definition phase leads to more clarity concerning ill-structured decision problems, which therefore become well-structured. A concrete definition in the context of a sample evaluation scenario is given in section 6.

5.4 Criteria

Generally, the meaning of a statement is considered as its sense. There are several different possibilities to reach a general acceptance on the intended sense of a statement, which can be distinguished. Accepting a statement as true by consensus means the possibility to reach an unconstrained consensus solely based on arguments. In contrast to that, a statement can be accepted as true, if it agrees with facts of the real world, thus, it can be observed. Moreover one could regard truth as coherence within a system of statements. All these possibilities are considered as criteria for a particular match and may be meaningful in different contexts. It is also possible to combine these criteria. E. g. a match is considered as correct, if a consensus is reached on it, but it must also be coherent in the system of the whole reference matching. It depends on the scenario, which solution is being selected. However, we present some thoughts, which can be used for selecting the most suitable criteria.

As the systems of statements changes with different models, which can lead to different interpretations of truth, it can be hard keeping coherence within a reference matching. Nevertheless coherence is an important criterion in general, especially in the context of process model matching. In case of a homogenous system of statements, this criterion might be very helpful.

An examination of statements may often be unpractical in context of high level business process models, as these are only restricted representations of reality. If the process model is detailed enough for observing the process in real world, the assignment of that criterion may be meaningful.

In many contexts, the consensus seems being suitable, whereby the accepted matches should be coherent in the whole. This criterion can be applied in terms of reference matching development, if the involved experts have the knowledge to assess models or schemas and their elements adequately. Thus, experts must have detailed knowledge on the processes in general and on the meaning of the containing node (labels) to decide about the correctness of a match related to a matching definition.

Nevertheless we presented criteria for establishing a reference matching in a particular context, we argue, that adequate and generally accepted criteria are still missing.

6 Application Scenario

To illustrate the presented guidelines, we prepared a sample scenario based on the models analyzed in section 4 with a focus on the different university admission processes. The objective is to generate a reference model in an inductive manner, based on the given individual models.

Idea. In terms of collecting the possibilities on how to match, we generally use the proposed ideas in section 5.2. As the available business process models are EPCs, we adapt the constructs to our data material. Thus, we consider the constructs *functions*, *events*, *connectors* and *edges*. Further adaptions are not necessary.

Definition. In order to inductively generate a reference model for university admissions based on a set of given individual models, we are interested in both, the needed activities and the structure of the different processes. Therefore functions and connectors are focused, the other EPCs constructs will not be matched. One characteristic of an inductively developed reference model is an abstraction from details. Thus, it is important to know the objective of particular activities, while the detailed action is of minor importance. This means, that facts are being matched, if they are equivalent in objective respectively have an analogue meaning. Against that background, concatenations are allowed as well, since it is possible to match model fragments (node blocks) which have the same objective in the whole but are different in detail.

Criteria. Reference models are high level in general, which is especially founded in their generalization. Therefore it is hard to observe particular aspects in real world. This leads to the decision accepting a match as correct, if a set of domain experts reach a consensus on it. Moreover, a particular match must be coherent in terms of the whole reference matching.

Considering the illustration of specific issues in section 4.1 in context of our matching scenario, it is much easier to decide for or against the presented alternatives. Following the guidelines leads to a definition of the context, which makes ill-structured decision problems on how to match process models more well-structured. The mentioned scenario shows the high dependency of a reference matching on the context in terms of a concrete scenario, whereby the characteristics of a definition can differ in an extensive manner. Against that background the given example should not be understood as a final solution for this specific matching problem, but as one possibility.

7 Conclusion

The paper at hand brought several findings to the light. In contrast to the recent literature, we recognized, that the only generally accepted reference matching does not exist for two process models. If single matches lead to ill-structured decision problems, there is no solution for a reference matching, as it is not possible to reach a consensus about a matching. In case of well-structured decision problems there is a reference solution, but not only one solution. In fact, a reference matching always depends on further assumptions, derived from concrete evaluation or application scenarios. Thus, depending on the scenario definition and the criteria of accepting a match as correct, there may be several reference matches.

That is also grounded in the finding, that a reference matching does not exist until its establishment. As neither necessary nor sufficient criteria are specified in the matching definition and since we detected, that the unproblematic deduction of a generally accepted reference matching it not possible in most cases, it must be given by an external resource, e. g. a set of experts. This leads to the result, that a reference matching does not exist a-priori, but ex-post to its establishment.

We also brought to the light, that there are many challenges establishing an accepted reference matching. We analyzed these challenges and proposed some guidelines containing the three phases *idea*, *definition* and *criteria* to support the development of reference matches in context of individual evaluation scenarios. We explicitly point out, that these guidelines, especially the presented *ideas* are only an assistance supporting that task and do neither constitute a final solution for the problem, nor impose the demand on completeness. Therefore, the guidelines should be discussed in the community and further developed in future work.

8 Literatur

[1] Becker, M., Laue, R. (2012): A comparative survey of business process similarity measures. Computers in Industry 63, 148-167

[2] Bellahsene, Z., Bonifati, A., Duchateau, F., Velegrakis, Y. (2011): On Evaluating Schema Matching and Mapping. In: Bellahsene, Z., Bonifati, A., Rahm, E. (eds.) Schema Matching and Mapping, pp. 253-291. Springer, Berlin, Heidelberg

[3] Cayoglu, U., Dijkman, R., Dumas, M., Fettke, P., Garcia-Banuelos, L., Hake, P., Klinkmüller, C., Leopold, H., Ludwig, A., Loos, P., Mendling, J., Oberweis, A., Schoknecht, A., Sheetritt, E., Thaler, T., Ullrich, M., Weber, I., Weidlich, M. (2013): The Process Model Matching Contest 2013. 4th International Workshop on Process Model Collections: Management and Reuse (PMC-MR'13), Beijing

[4] Dijkman, R., Dumas, M., van Dongen, B., Käärik, R., Mendling, J. (2011): Similarity of business process models: Metrics and evaluation. Information Systems 36, 498-516

[5] Euzenat, J., Shvaiko, P. (2007): Ontology Matching. Springer-Verlag New York, Inc.

[6] Evermann, J. (2009): Theories of meaning in schema matching: An exploratory study. Inf. Syst. 34, 28-44

[7] Houy, C., Fettke, P., Loos, P., van der Aalst, W., Krogstie, J. (2011): Business Process Management in the Large. Business & Information Systems Engineering 3, 385-388

[8] Kunze, M., Weske, M. (2013): Methods for Evaluating Process Model Search. In: 4th International Workshop on Process Model Collections: Management and Reuse (PMC-MR'13), pp. 1-12.

[9] Piazza, F. (2010): Data Mining im Personalmanagement: Eine Analyse des Einsatzpotenzials zur Entscheidungsunterstützung. Gabler

[10] Rahm, E., Bernstein, P.A. (2001): A survey of approaches to automatic schema matching. International Journal on Very Large Data Base 10, 334-350

[11] Van Rijsbergen, C.J. (1979): Information Retrieval. Butterworth-Heinemann

[12] Weidlich, M., Dijkman, R., Mendling, J. (2010): The ICoP Framework: Identification of Correspondences between Process Models. In: Pernici, B. (ed.) Advanced Information Systems Engineering: 22nd International Conference, CAiSE 2010, LNCS 6051, pp. 483-498. Springer, Berlin, Heidelberg