International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC)

IJISEBC, 01, 1, 2014

25 Challenges of Semantic Process Modeling

25 Desafíos de la Modelación de Procesos Semánticos

Jan Mendling[1], Henrik Leopold[2], Fabian Pittke[1]

[1] Institute for Information Business, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

[2] Department of Computer Sciences, Vrije Universiteit Amsterdam, Faculty of Sciences, De Boelelaan 1081, 1081HV Amsterdam, The Netherlands

jan.mendling@wu.ac.at, h.leopold@vu.nl, fabian.pittke@wu.ac.at

ABSTRACT. Process modeling has become an essential part of many organizations for documenting, analyzing and redesigning their business operations and to support them with suitable information systems. In order to serve this purpose, it is important for process models to be well grounded in formal and precise semantics. While behavioural semantics of process models are well understood, there is a considerable gap of research into the semantic aspects of their text labels and natural language descriptions. The aim of this paper is to make this research gap more transparent. To this end, we clarify the role of textual content in process models and the challenges that are associated with the interpretation, analysis, and improvement of their natural language parts. More specifically, we discuss particular use cases of semantic process modeling to identify 25 challenges. For each challenge, we identify prior research and discuss directions for addressing them.

KEYWORDS: Process modeling, Formal semantics, Natural language processing, System analysis and design.

IJISEBC, 01, 1, 2014

1. Introduction

Process models play an important role in various application scenarios that relate to system analysis and design [Wes12, DRMR13]. They often serve as a specification to bridge between business requirements and workow implementation. Process models have been intensively studied in terms of their behavioural properties, for instance on the basis of formalisms such as Petri nets, automata, labeled transition systems or temporal logic, to name but a few [van00].

Compared to the extensive stream of research into behavioural semantics, it is surprising to observe that the textual content of process models has received by far less attention. This fact reects a painful gap in current research since the domain understanding of process models builds the more on its textual content the less the persons creating and reading the models have a formal education in computer science. On the one hand, it is often casual modelers from the line of business that work with models [Ros06], and these tend to pay little attention to behavioural semantics. On the other hand, their model understanding strongly depends on the appropriate formulation of text labels in the process model and their accurate interpretation [MRR10].

The aim of this paper is to make this identified research gap more transparent. To this end, we define a modeling language with an explicit reference to its textual content and describe the interpretation of text on the three levels of the single label, the model fragment and the whole model collection. We use these three levels to organize 25 challenges of semantic process modeling. These 25 challenges relate to the various tasks that involve the interpretation, analysis and improvement of text labels in a process model. In this way, it complements prior research on tasks and use cases as identified for business process modeling and process mining [vdA13], change patterns [WRR08] and refactorings [WRMR11].

The paper is structured as follows. Section 2 describes the setting in which process models are created and their different components. Section 3 identifies challenges of working with label. Section 4 describes challenges of working with textual labels on the level of a model fragment or a whole model. Section 5 describes challenges in the relation to the management of an overall model collection and its textual content. Section 6 discusses the challenges before Section 7 concludes the paper.

2. Background

Process modeling plays an important role in various areas of system analysis and design. Specifically business process modeling was identified as one of the most prominent applications of conceptual modeling altogether [DGR+06]. Modeling techniques are typically used for creating models of good quality. The different components of a modeling technique are illustrated in Figure 1.

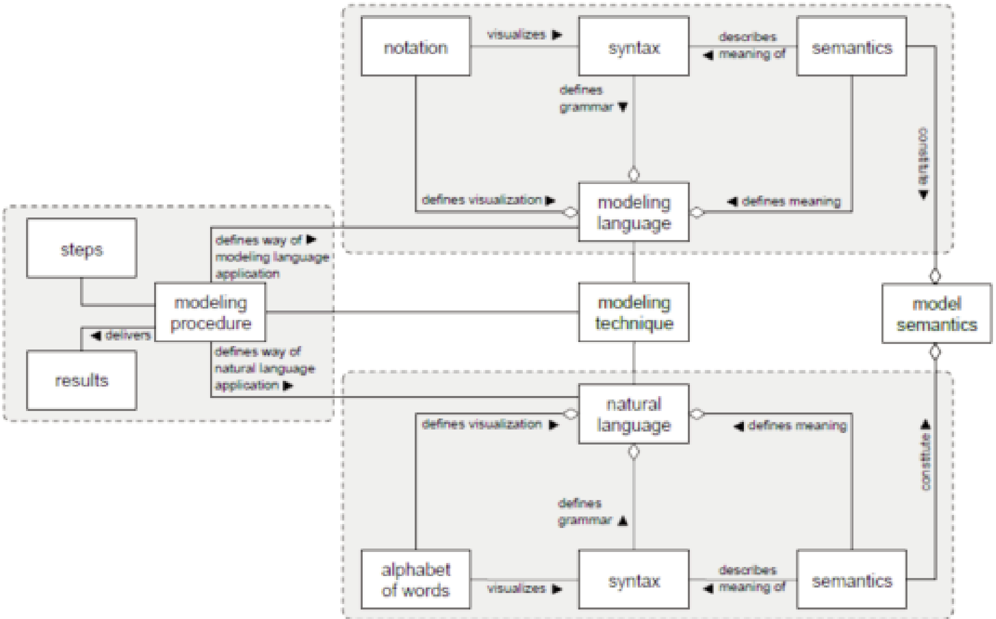

Figure 1. Syntax and Semantics of Process Models [Leo13].

Classically, a modeling technique has been considered to consist of two interrelated parts: a modeling language and a modeling procedure [Men08]. The modeling language consists of three parts: syntax, semantics and a notation. The syntax defines a set of elements and a set of rules how these elements can be combined. A synonym is modeling grammar [WW90, WW95, WW02]. Semantics bind these elements to a precise meaning. For process model, behavioural semantics are often defined using Petri net concepts [LVD09]. The notation defines a set of graphical symbols that are utilized for the visualization of models [Moo09]. The modeling procedure defines steps by which a modeling language can be used [WW02, DRMR13]. The result of applying the modeling procedure is a model that complies with a specific modeling language.

Recent research has extended this classical conceptualization with a more explicit specification of the textual parts of models. Therefore, Figure 1 shows the natural language part as a separate component. The terminology used in the models is defined by the alphabet of words while the syntax is defining the rules of building text fragments that are permissible for the specific type of model [Leo13]. For instance, the activity label of a process model is typically assumed to contain a verb and a business object [LESM$^+$13]. The semantics in this context refer to the precise interpretation of the words used in the label.

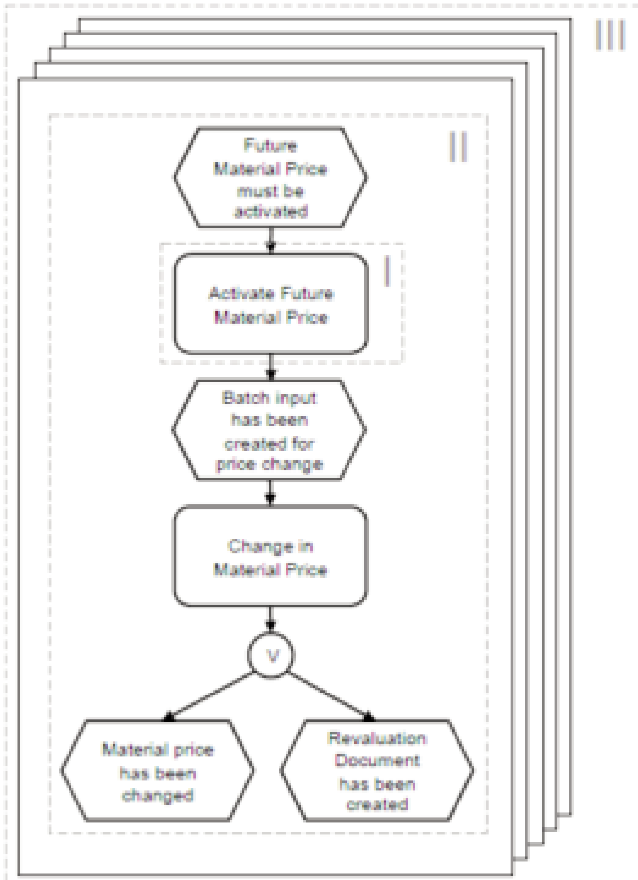

Figure 2. Three Levels of Semantic Process Modeling.

Extending the perspective of process modeling towards the explicit discussion of natural language components is promising specifically for applications that require to analyze both behavioural and textual semantics, such as process model matching [WDM10], process model reuse [KFSO14], service identification [LM12], or model translation [BESL$^+$13]. On the other hand, this more integral perspective on conceptual modeling reveals various challenges.

In the following sections, we aim to describe tasks and corresponding challenges. We organize them into three categories that are based on the extent of their textual content (see Figure 2).

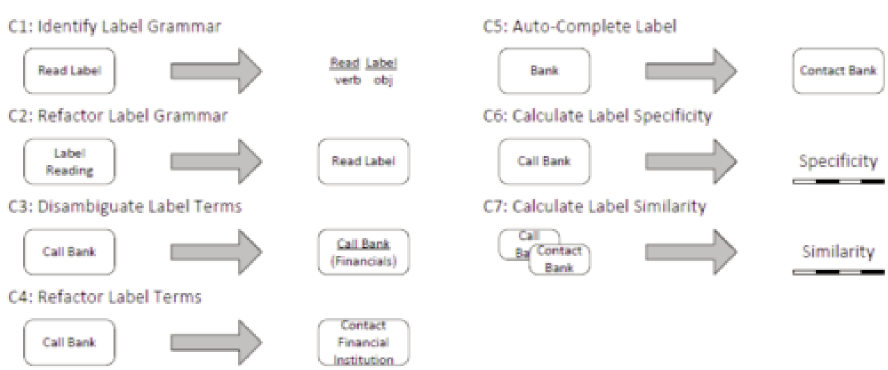

Figure 3. Challenges in Relation to Labels.

IJISEBC, 01, 1, 2014

The first category relates to labels and their analysis. The second category describes analysis onvthe level of whole models or model fragments. Finally, the third category discusses challenges on the level of whole model collection. Each challenge is structured accordingly. We discuss each challenge by clarifying the goals and the necessary input information of the associated task. Based on that, we further specify the challenges linked to a particular task and illustrate them with the help of small examples. Finally, we conclude with a short summary of prior research and explain how the respective challenge has been addressed with conceptual or technical solutions.

3. Label Challenges

In this section, we describe various challenges on analyzing and reworking labels of elements that appear in a process model. Figure 3 gives an overview.

C1: Identify Label Grammar. The goal of this task is the automatic identification of the semantic components of a process model element label. The input for this task is an element label and, if applicable, the process model and the process model collection the label is part of.

The challenge of this task is the proper recognition of the various and potentially ambiguous grammatical label structures. It is further complicated by the shortness of element labels and the fact that they often do not represent proper sentences. As a result, it is dificult to always identify the correct part of speech of label terms. As an example, consider the label "plan data transfer", which may refer to the "planning" of a "data transfer" or the "transfer" of "plan data". Prior research has approached this challenge by describing grammatical styles of labels and defining corresponding parsers [LSM10]. Ambiguity can be resolved based on the inclusion of further contextual and external knowledge [LESM$^+$13]. Besides the recognition of the label grammar, the resulting techniques can also be used for checking the compliance with a grammatical guideline [LESM$^+$13, BDH$^+$09, DHLS09].

C2: Refactor Label Grammar. The goal of this task is to refactor the existing grammar of a particular label to a more desirable grammatical style. The input for this task is the label and its previously identified semantic components.

The challenges in the context of this task include lemmatization, i.e. deriving the base form from an inected word, as well as the proper recognition of compound words. As an example, consider the label "new user registration". For refactoring this label into the widely requested verb-object style [Sil11, MRR10, MRvdA10], we first need to transform the nominalized action "registration" into to the verb "register". Second, we have to recognize that the adjective "new" refers to the "user" and not to the entire "user registration". As a result, we obtain the refactored verb-object label "register new user". Prior research has approached these challenge by building on WordNet and a number of structural heuristics [LSM12].

C3: Disambiguate Label Terms. The goal of this task is to recognize the meaning of a term from a process model element label. The input for this task is a label term including its context, i.e., the label it belongs to and, if applicable, the model and the process model collection the label is part of.

The challenge of this task is to identify the correct meaning of a word despite the limited context that is provided by process model element labels. As an example, consider the label "check application". Depending on the context, the word "application" could refer to a "job application" as well as a "computer application". Prior research has approached this challenge by selecting the most probable meaning from lexical databases such as WordNet [PLM13] or BabelNet [PLM15] based on the label context.

C4: Refactor Label Terms. The goal of this task is to replace syntactically identical words with different meanings (homonyms) and syntactically differing words with the same meaning (synonyms) with unambiguous alternatives. The input for this task is a label term including its context and the previously identified meaning of that label term. The challenge of this task is to identify un-ambiguous and suitable alternatives for the considered homonymous or synonymous term. As an example, consider the homonym “application". Depending on the context, the word “application" may be, for instance, replaced with “job application". In case synonyms such as “invoice" and “bill", a choice for the most suitable word must be made. Prior research has approached this challenge by building on the meanings and the context information from the lexical database BabelNet [PLM15].

C5: Auto-Complete Label. The goal of this task is to automatically provide useful suggestions for completing an incomplete label. The input for this task is an incomplete label, for instance, only consisting of a business object combined with further context information, such as the process model or the process model collection.

The challenge of this task is to recognize the context of a label, to generate suitable completion candidates, and to rank them according to their relevance. As an example, consider the label "bank", which only consists of a business object. An automated technique would be required to analyze the context and to propose a suitable action such as “contact" or “call". Prior research has approached this problem by building on existing process knowledge [CHSB13].

C6: Calculate Label Similarity. The goal of this task is to obtain a (realistic) similarity value between 0 and 1 for two given process model element labels. The input for this task are two process model element labels. If required, additional information such as the previously derived semantic components may complement the labels.

The challenge of this task is to identify means that facilitate the realistic measurement of the semantic similarity of two labels. The task is complicated by the specificity of many terms that are used in process models as well as different levels of granularity. As an example, consider the two labels “check application documents" and “evaluate CV". Apparently, the second label is a sub task of the first. However, it represents already a challenge to properly quantify the similarity between “document" and “CV". Prior research has approached this challenge by computing and aggregating the Lin similarity among the words or the semantic components of the two labels [CDD+13]. Non-semantic approaches based on the Levenshtein distance have been, for example, proposed in [EKO07, DDvD+11].

C7: Calculate Label Specificity. The goal of this task is to quantify the specificity of a given process model element label. The input for this task is a process model element label and, if required, its semantic components.

The challenge of this task is to identify suitable means for measuring the specificity of the label terms as well as the label as a whole. Particularly challenging are labels which contain words that cannot be found in lexical databases such as WordNet. As an example, consider the label “call customer service hotline". The specificity of the term “hotline" can be, for instance, determined based on the position of the word in the WordNet taxonomy. However, this is not possible for the term “customer service hotline" as this term is not part of the WordNet database. Prior research has approached this challenge by using on WordNet [Fri09, KB07] and other heuristics such as label length and the number of semantic components [LPM13].

4. Model Challenges

In this section, we describe various challenges on analyzing and reworking semantic fragments of process models. Figure 4 gives an overview.

C8: Discover Label Mapping. The goal of this task is to map a phrase to a text label. The input for this task is a process model with its text labels and a piece of text containing several phrases.

The challenge of this task is to identify the activity in the process model, which is semantically the closest to a text sentence. As an example, consider a natural language text containing the sentence "he ordered the book" and a process model containing activities with the labels "send invoice" and "order book". Prior research has approached this task as a classifier problem. The tool SemTalk helps to maintain consistency between labels and separate concepts [FWAW03].

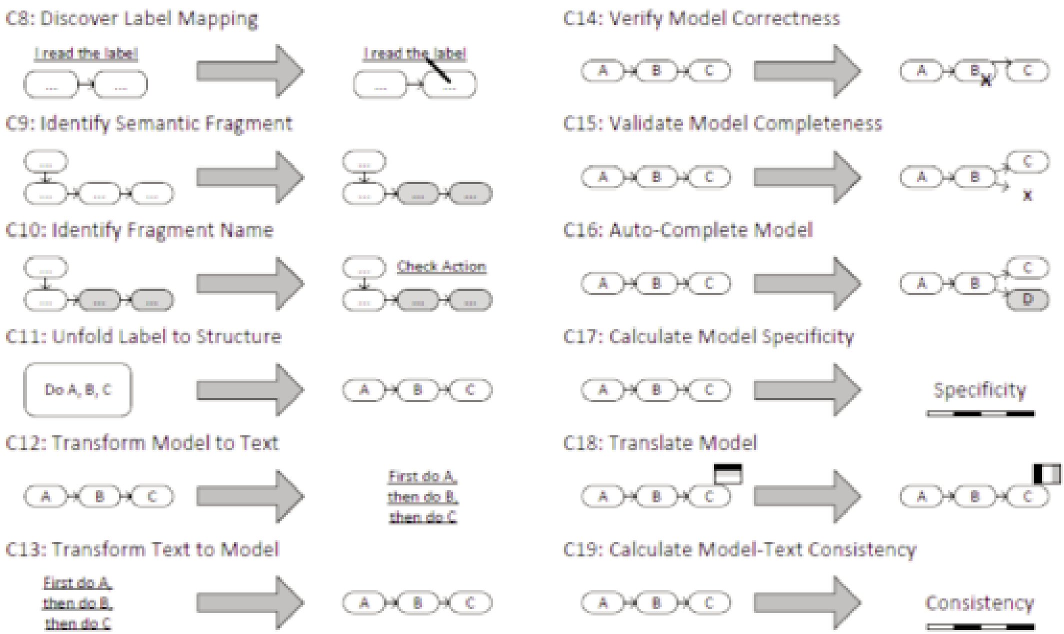

Figure 4. Challenges in Relation to Models.

C9: Identify Semantic Fragment. The goal of this task is to identify a fragment of a process model that is semantically closely related. The input for this task is the model and the labeled activities.

The challenge of this task is to determine those activities that are related and can be described as a whole on a more abstract level. As an example, consider the activities "receive order" and "check order". Together, these are activities that both relate to the handling of orders. Prior research has approached this challenge by different approaches on process model abstraction. One approach uses semantic relations such as meronymy [SDMW10] and different notions of distance [RMD11, SRW12]. Various abstraction scenarios are summarized in [SRWN12].

C10: Identify Fragment Name. The goal of this task is to identify the name of a set of activities that describe them at a more abstract level. The input for this task is a process fragment containing the set of activities.

The challenge of this task is to find a name for this fragment that captures its content in a semantically meaningful way. Also, the name of activities can be defined from different perspectives, e.g. what is being done or what is supposed to be achieved. As an example, consider again the "activities "receive order" and "check order". A technique for naming this fragment should propose a label like "handle order". Prior research has approached this challenge by describing different strategies for defining a name of a fragment or a whole process based on theories of meaning such that different proposals can be derived automatically [LMRR14].

C11: Unfold Label to Structure. The goal of this task is to decompose a label into different activities and to transform this into a corresponding fragment of a process model. The input for this task is an activity label that describes more than just a single activity.

The challenge of this task is to identify that several activities are described and which structure can best capture their semantics. As an example, consider a single activity label "receive and check order". Apparently, the single label refers to two activities which might be executed in parallel or sequential order. Prior research has approached this challenge by identifying commonalities in process model collections and deducting regular anti patterns that incorporate several activities in one activity label [PLM14].

C12: Transform Model to Text. The goal of this task is to transform a process model into a natural language process description. The input for this task is a process model along with the semantic component annotations of its elements.

The challenge of this task is to present the non-sequential structure of a process model in a sequential fashion. In addition, the text should be as natural as possible. As an example, consider the sequence of the activities "receive order", "check order", and "send products" in a process model. A technique to transform the model fragment into text should create a text fragment like "The process begins with the receipt of an order. After the order is checked, the products are sent to the customer". Prior research has approached this challenge by proposing a technique that automatically generates a textual representation of a given process model based on the refined process structure tree and the meaning text theory [LMP14]. Template-based approaches have been proposed in [Cos10, MB13].

C13: Transform Text to Model. The goal of this task is to elicit a process model from a natural language process description. The input for this task is a piece of text.

The challenge of this task is to properly discover the activities as well as the order of activities including decisions, concurrency, and loops. As an example, consider the text "The kitchen prepares the meal. In the meantime, the waiter takes care of the beverages". An automated technique would have to recognize the roles "kitchen" and "waiter", the activities "Prepare meal" and "Take care of beverages" as well as the fact that the two activities are conducted in parallel. Prior research has approached this challenge by applying standard natural language processing techniques and a number of signal words and phrases [FMP11, dAGSB09, GKC07, SP10].

C14: Verify Model Correctness. The goal of this task is to check whether a process model is correct according to the semantics defined by its activity labels. The input for this task is the process model together with semantic annotations for the activity labels.

The challenge of this task is to identify those activities that significantly inuence the control-flow of the process model and validate if the control-flow matches the semantics of the label. As an example, consider the activity label "assess application" that requires an application has been checked for completeness. Therefore, there has to be a prior activity that guarantees this requirement to be fulfilled. Prior research has approached this challenge by propagating preconditions and effects over the process model for semantic verification [WHM10] or by building on linguistic knowledge [vdVGvdR97, GL11]. Correctness by design is provided by approaches using automatic planning of business processes [HLD+05].

C15: Validate Model Completeness. The goal of this task is to check whether a process model is correct according to the semantics defined by its activity labels. The input for this task is the process model together with semantic annotations for the activity labels.

The challenge of this task is to identify those activities that significantly inuence the control-flow of the process model and validate if the control-flow matches the semantics of the label. As an example, consider the activity label "assess application" that may either result in an approval or a rejection. For the sake of semantic model consistency, the application cannot be accepted and rejected at the same time and thus demands an exclusive decision after the activity. Prior research has approached this challenge by using semantic error patterns, for instance based on antonyms [GL11]. The approach in [TF07] discusses the opportunities of using semantic web technologies to reason about process models. Validation in the context of customization of process variants is discussed in [DB13, GBPG13, AMGG14].

C16: Auto-Complete Model. The goal of this task is to provide user-assistance during process modeling and to avoid typographical and syntactical errors in process models. The input for this task is a set of process models from which suggestions to complete the process are created.

The challenge of this task is the definition of learning recommendation system that suggests a list of meaningful process fragments that may be entered at this current position within the process model. As an example, consider again the sequence of the activities "receive order", "check order". A recommendation system might suggest a XOR-split with the activity "send products" if the check is successful "inform customer" if the check fails. Prior research has approached this challenge by using business rules and structural constraints to propose appropriate process fragments [HKO07].

C17: Calculate Model Specificity. The goal of this task is to identify and adjust element labels according to their level of detail within a hierarchy of process models. The input for this task is a process model as well as its position in a process hierarchy or a process architecture.

The challenge of this task is to measure the concept of specificity and to recommend actions to adjust element labels that do not comply to the level of detail within the process hierarchy. As an example, consider the a sequence of activities "receive order", "check purchase order", and "send products" which describes the handling of an incoming order on a general level. Apparently, the second activity is too specific as it entails a particular type order that needs to be checked. Prior research has approached this challenge by providing a set of syntactical and semantic metrics that measure the granularity of element labels [LPM14].

C18: Translate Model. The goal of this task is to overcome the language barrier for re-using of process models in multi-national companies. The input for this task is a process model in a particular language.

The challenge of this task is dealing with the short texts in labels, recognizing the context of the process model, and appropriately translating the process model into the target language. As an example, consider the activity "receive order". If we consider a translation of this activity, the translation system should be capable to recognize that the word order is used in the sense of a commercial document and not in the sense of a military command and thus chose the appropriate translation. Prior research has approached this challenge by developing a technique for the automated translation of business process models that builds upon statistical machine translation and word sense disambiguation [BESL+13].

C19: Calculate Model-Text Consistency. The goal of this task is to measure the consistency between a process description as a process model and as a natural language text and to identify notable differences between these descriptions. The input for this task is the process model together with a textual process description.

The challenge of this task is again defining abstract representation to map the content of both text and model and identifying deviations of both types. As an example, consider a sequence of the activities "receive order", "check order", and "send products" as well as the text fragment "After the order is received, the respective products are send to the customer". Apparently, the textual description is not consistent to the activity sequence because one activity is missing in the textual description. Prior research has approached this challenge by translating a textual description into process models resolving arising inconsistencies either in an automated or mediated manner [GKC07].

5. Collection Challenges

In this section, we describe various challenges on analyzing and reworking semantic fragments of process models. Figure 5 gives an overview.

IJISEBC, 01, 1, 2014

C20: Discover Model Mapping. The goal of this task is to discover a mapping between the sets of activities of two process models. The input for this task is a pair of process models and a similarity matrix over the pairs of activities.

The challenge of this task is that activities are potentially described on different levels of granularity such that not only 1:1, but also 1:n and n:m matches are possible. As an example, consider the coarse-granular activity "build car" in one model and the sequence of "purchase parts", "assemble parts", and "check car" in a second model. Prior research has approached this challenge by using concepts from ontology matching [WDM10]. These have been extended towards using constraints to reduce the search space [LNW$^+$12] and including feedback [KLW$^+$]. A comparison of different techniques is reported in [CDD$^+$13].

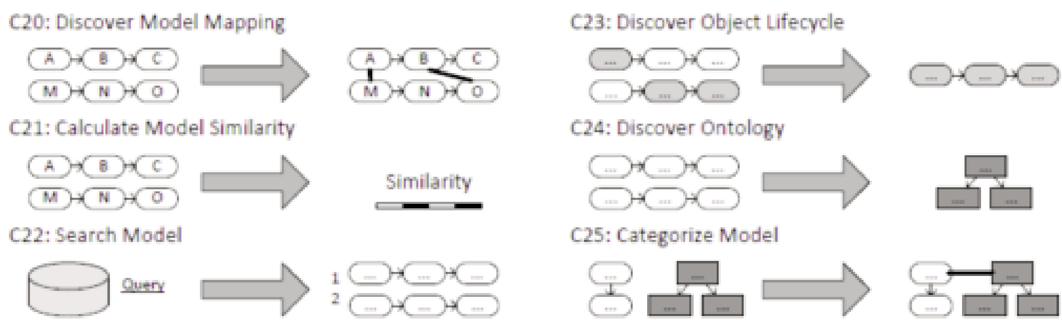

Figure 5. Challenges in Relation to Collections.

C21: Calculate Model Similarity. The goal of this task is to determine how similar process models are. The input for this task is a pair of process models and a mapping between their activities.

The challenge of this task is to consider adequately different aspects of representational heterogeneity including labels, structure and behaviour. For example, there are different ways to model the fact that both activities A and B are executed or just one of them. Models can be trace equivalent, but have different structure. Prior research has approached this challenge by defining behavioural abstractions. The behavioural profile [WMW11], transition adjacency [ZWW$^+$10] and matrix relations [ABDG14] define behavioural relations over the cartesian product of activities. The matrices of two models can then be compared cell-wise [DvDD$^+$13]. As an alternative, graph edit distance can be used [DGD09]. Similar approaches are defined in [EKO07, EG07, CGB06]. A comparison of approaches is reported in [DDvD$^+$11].

C22: Search Model. The goal of this task is to rank process models of a collection according to how similar they are to a given search query. The input for this task is a search query and a collection of process models.

The challenge of this task is identify those features that are supposedly relevant for calculating the semantic distance between the query and each of the process models. As an example, consider a query containing the term "Human Resources". A suitable technique would be able to identify also models that do not contain this term, but also those that contain related terms such as "employee" or "contract". Prior research has approached this challenge by building on WordNet [APW08] and language modeling [QAR11]. Alternatively, query languages such as PQL [KB04] and BPMN-Q [ADW08], as well as indexing [YDG12, JWW$^+$10] or clustering techniques [QAR11, RMKL12]. Also, behavioral profiles are used to search for models [KWW11].

C23: Discover Object Lifecycle. The goal of this task is to discover the lifecycle of objects from the activities described in a collection of process models. The input for this task is a collection of process models and the semantic annotation of the activity labels.

The challenge of this task is to integrate the parts of the lifecycle, which might be scattered over several models. For example, consider one model including the activities “receive order" and “check order" and a second model with the activities “check order" and “confirm order". Prior research has approached this challenge by identifying action patterns between activity pairs [SWMW12], which can be synthesized to lifecycle models of the respective business objects [SWM12]. Based on these lifecycle models, compliance between process models and object lifecycle can be discussed [KRG07].

C24: Discover Ontology. The goal of this task is to discover a formal ontology from a collection of process models. The input for this task is a collection of process models and the semantic annotation of the activity labels.

The challenge of this task is to extract pieces of information that can be used for identifying formal concepts and relationships. As an example, consider decomposition relationships between process models and semantic groupings that are not explicitly defined. Prior research has approached this challenge for building taxonomies [PW11].

C25: Categorize Model. The goal of this task is to identify a category in which a particular model fits best. The input for this task is a process model and a taxonomy, which is in the simplest case a set of categories.

The challenge of this task is that category descriptions might contain only a few terms and that process models might include tasks that relate to several categories. As an example, consider PCF Taxonomy, which contains 1131 hierarchically organized concepts. Prior research has not addressed this challenge in detail. Promising directions include the extension of existing approaches for semantic annotation of process models [FT09, LD05, BSPW08, BDW07?]. There is also work that identifies categories inductively from the models [MDM13].

6. Discussion

This section discusses the state of current research on semantic business process modeling based on the challenges identified above. At this stage, it has to be noted that the merits of these challenges should not seen in terms of a claim for completeness - indeed, it is unclear whether it is feasible to provide a complete list of challenges at all. The benefits of this compilation has to be seen much more in its capability of separating well-researched areas from topics that have received little attention so far. Therefore, we want to structure this discussion along the following lines: tasks that we observe to be well-researched, tasks that call for more research, and base techniques that could help to advance semantic process modeling.

Among the well-researched tasks, we regard the identification and refactoring of label grammar (C1 and C2), the calculation of similarity (C7 and C21), the identification of a semantic fragment (C9), and the search for particular models (C22) as mature tasks. Approaches addressing tasks of C1 and C2 perform well with real-world data and have a high accuracy in processing the labels. A similar observation can be made for approaches of label similarity (C7) and model similarity (C21). In particular for the latter, research approaches have incorporated the element labels, the model structure, and the model behavior as relevant aspects of model similarity and proposed several metrics for its calculation. With regard to the identification of semantic fragments (C9) and to the search of process models (C22), we identify a considerable number of approaches covering several requirements with regard to these tasks. Thus, we also conclude that these tasks are well understood and supported by recent approaches.

Turning to the tasks that require more research, we want to highlight the tasks that relate to the specificity of labels (C6), the alignment of text and model (C12, C13, C19), and the ontology- related tasks (C24, C25). As outlined before, specificity-related tasks try to adjust the label components depending on their level of detail within a process model landscape. In such as setting, finding the appropriate level of granularity is still an open challenge [DVR11] and despite prior efforts not addressed in sufficient detail. Regarding the alignment of models and text, we observe that non-analysts increasingly work with process models and require a solid understanding of the underlying process. In order to support non-analysts in understanding and problem-solving tasks with reference to the process at hand, textual descriptions of the processes are maintained as a complement to the process models process models. However, we observe a notable gap of approaches that provide an alignment of process models and textual descriptions. Similarly, we find approaches for integrating ontologies with process models [HLD$^+$05, HR07]. While the creation of ontologies is work-intensive, difficult and often domain-specific [PSFGP10], it would be desirable to support these task in an automatic fashion. We identified only a very small number of approaches that address this challenge. Thus, we call for more approaches to learn ontologies from process models and to link process models to existing ontologies or taxonomies.

The challenges also revealed several base techniques from which existing solutions of semantic process modeling would potentially benefit. Among them, we identify the integration of text corpora, such as Wikipedia or related repositories, as well as the and extending the set of semantic relationships as most promising. The integration of large text corpora and corpus-based techniques might be a suitable direction for working around the limitations of general purpose databases like WordNet in terms of its vocabulary. The rich spectrum of semantic relationships might support the discovery of an ontology as well as the search and categorization of process models. So far, only a limited amount of semantic relations have been used. Specifically, homonym and synonym relations have been used to correct ambiguous terminology in process models, while meronym relations have been proven as useful to find semantic fragments. However, there are still semantic relations left might support specific tasks. Future research should consider the usage of a broader range of semantic relationships, including hypernyms, hyponyms, meronyms, holonyms, antonyms, or troponyms.

7. Conclusion

In this paper, we shed light onto the challenges that relate to the analysis of the textual content in process models. We identified a number of 25 challenges that arise when dealing with the textual content on the level of a single label, on the level of a process model and on the level of a process model collection. For each challenge, we identified necessary input information, further specified the challenge with the help of examples, and explain how related work has addressed the challenge so far. In light of these challenges we hope to increase the interest and the awareness of future research streams towards the textual content of process models. We expect our list of challenges to help in positioning current research activities and in fostering innovative ideas to address the identified gaps.

Cómo citar este artículo / How to cite this paper

Mendling, J., Leopold, H., & Pittke, F. (2014). 25 Challenges of Semantic Process Modeling. International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC), Vol. 1, Num. 1, pp. 78-94. Consultado el [dd/mm/aaaa] en www.ijisebc.com

References

[ABDG14] Abel Armas-Cervantes, Paolo Baldan, Marlon Dumas, and Luciano García-Bañuelos. Behavioral comparison of process models based on canonically reduced event structures. In Shazia Wasim Sadiq, Pnina Soffer, and Hagen Völzer, editors, Business Process Management - 12th International Conference, BPM 2014, Haifa, Israel, September 7-11, 2014. Proceedings, volume 8659 of Lecture Notes in Computer Science, pages 267-282. Springer, 2014.

[ADW08] Ahmed Awad, Gero Decker, and Mathias Weske. Efficient compliance checking using BPMN-Q and temporal logic. In Marlon Dumas, Manfred Reichert, and Ming-Chien Shan, editors, Business Process Management, 6th International Conference, BPM 2008, Milan, Italy, September 2-4, 2008. Proceedings, volume 5240 of Lecture Notes in Computer Science, pages 326-341. Springer, 2008.

[AMGG14] Mohsen Asadi, Bardia Mohabbati, Gerd Gröner, and Dragan Gasevic. Development and validation of customized process models. Journal of Systems and Software, 96:73-92, 2014.

[APW08] Ahmed Awad, Artem Polyvyanyy, and Mathias Weske. Semantic querying of business process models. In Enterprise Distributed Object Computing Conference, 2008. EDOC'08. 12th International IEEE, pages 85-94. IEEE, 2008.

[BDH+09] J. Becker, P. Delfmann, S. Herwig, L. Lis, and A. Stein. Towards Increased Comparability of Conceptual Models - Enforcing Naming Conventions through Domain Thesauri and Linguistic Grammars. In ECIS 2009, June 2009.

[BDW07] M. Born, F. Dörr, and I. Weber. User-Friendly Semantic Annotation in Business Process Modeling. In WISE 2007 Workshops, volume 4832 of LNCS, pages 260-271. Springer, 2007.

[BESL+13] Kimon Batoulis, Rami-Habib Eid-Sabbagh, Henrik Leopold, Mathias Weske, and Jan Mendling. Automatic business process model translation with bpmt. In Advanced Information Systems Engineering Workshops, pages 217-228. Springer, 2013.

[BSPW08] Andreas Bögl, Michael Schre, Gustav Pomberger, and Norbert Weber. Semantic annotation of epc models in engineering domains to facilitate an automated identification of common modelling practices. In ICEIS, pages 155-171, 2008.

[CDD+13] Ugur Cayoglu, Remco M. Dijkman, Marlon Dumas, Peter Fettke, Luciano García-Banñuelos, Philip Hake, Christopher Klinkmüller, Henrik Leopold, André Ludwig, Peter Loos, Jan Mendling, Andreas Oberweis, Andreas Schoknecht, Eitam Sheetrit, Tom Thaler, Meike Ullrich, IngoWeber, and Matthias Weidlich. Report: The process model matching contest 2013. In Lohmann et al. [LSW14], pages 442-463.

[CGB06] Juan Carlos Corrales, Daniela Grigori, and Mokrane Bouzeghoub. BPEL processes matchmaking for service discovery. In Robert Meersman and Zahir Tari, editors, On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE, OTM Confederated International Conferences, CoopIS, DOA, GADA, and ODBASE 2006, Montpellier, France, October 29 - November 3, 2006. Proceedings, Part I, volume 4275 of Lecture Notes in Computer Science, pages 237-254. Springer, 2006.

[CHSB13] Nico Clever, Justus Holler, Maria Shitkova, and Jörg Becker. Towards auto-suggested process modeling-prototypical development of an auto-suggest component for process modeling tools. In EMISA, pages 133-145, 2013.

[Cos10] Ahmet Coskuncay. An Approach for Generating Natural Language Specifications by Utilizing Business Process Models. Master's thesis, Middle East Technical University, 2010.

[dAGSB09] Joao Carlos de A.R. Goncalves, Flavia Maria Santoro, and Fernanda Araujo Baiao. Business process mining from group stories. International Conference on Computer Supported Cooperative Work in Design, pages 161-166, 2009.

[DB13] Wassim Derguech and Sami Bhiri. Business process model overview: Determining the capability of a process model using ontologies. In Witold Abramowicz, editor, Business Information Systems - 16th International Conference, BIS 2013, Poznan, Poland, June 19-21, 2013. Proceedings, volume 157 of Lecture Notes in Business Information Processing, pages 62-74. Springer, 2013.

[DDvD+11] R. M. Dijkman, M. Dumas, B. F. van Dongen, R. Käärik, and J. Mendling. Similarity of Business Process Models: Metrics and Evaluation. Information Systems, 36(2):498-516, 2011.

[DGD09] Marlon Dumas, Luciano García-Bañuelos, and Remco M. Dijkman. Similarity search of business process models. IEEE Data Eng. Bull., 32(3):23-28, 2009.

[DGR+06] Islay Davies, Peter Green, Michael Rosemann, Marta Indulska, and Stan Gallo. How do practitioners use conceptual modeling in practice? Data & Knowledge Engineering, 58(3):358-380, 2006.

[DHLS09] A. Delfmann, S. Herwig, L. Lis, and A. Stein. Supporting Distributed Conceptual Modelling through Naming Conventions - A Tool-based Linguistic Approach. Enterprise Modelling and Information Systems Architectures, 4(2):3-19, 2009.

[DRMR13] M. Dumas, M.L. Rosa, J. Mendling, and H. Reijers. Fundamentals of Business Process Management. Springer, 2013.

[DvDD+13] Remco M. Dijkman, Boudewijn F. van Dongen, Marlon Dumas, Luciano García-Bañuelos, Matthias Kunze, Henrik Leopold, Jan Mendling, Reina Uba, Matthias Weidlich, Mathias Weske, and Zhiqiang Yan. A short survey on process model similarity. In Janis A. Bubenko Jr., John Krogstie, Oscar Pastor, Barbara Pernici, Colette Rolland, and Arne SØlvberg, editors, Seminal Contributions to Information Systems Engineering, 25 Years of CAiSE, pages 421-427. Springer, 2013.

[DVR11] Remco Dijkman, Irene Vanderfeesten, and Hajo A Reijers. The road to a business process architecture: an overview of approaches and their use. The Nederlands: Einhoven University of Technology, 2011.

[EG07] Rik Eshuis and Paul W. P. J. Grefen. Structural matching of bpel processes. In Fifth IEEE European Conference on Web Services

(ECOWS 2007), 26-28 November 2007, Halle (Saale), Germany, pages 171-180. IEEE Computer Society, 2007.

[EKO07] Marc Ehrig, Agnes Koschmider, and Andreas Oberweis. Measuring similarity between semantic business process models. In John F. Roddick and Annika Hinze, editors, Conceptual Modelling 2007, Proceedings of the Fourth Asia-Pacific Conference on Conceptual Modelling (APCCM2007), Ballarat, Victoria, Australia, January 30 - February 2, 2007, Proceedings, volume 67 of CRPIT, pages 71-80. Australian Computer Society, 2007.

[FMP11] Fabian Friedrich, Jan Mendling, and Frank Puhlmann. Process model generation from natural language text. In Advanced Information Systems Engineering, pages 482-496. Springer, 2011.

[Fri09] Fabian Friedrich. Measuring semantic label quality using wordnet. In Proceedings of the 8th Workshop Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten (EPK), pages 42-57, 2009.

[FT09] Chiara Francescomarino and Paolo Tonella. Supporting ontology-based semantic annotation of business processes with automated suggestions. In Enterprise, Business-Process and Information Systems Modeling, volume 29 of LNBIP, pages 211-223. Springer Berlin Heidelberg, 2009.

[FWAW03] C. Fillies, G. Wood-Albrecht, and F. Weichhardt. Pragmatic applications of the Semantic Web using SemTalk. Computer Networks, 42(5):599-615, 2003.

[GBPG13] Gerd Gröner, Marko Boskovic, Fernando Silva Parreiras, and Dragan Gasevic. Modeling and validation of business process families. Inf. Syst., 38(5):709-726, 2013.

[GKC07] A.K. Ghose, G. Koliadis, and A. Chueng. Process Discovery from Model and Text Artefacts. In 2007 IEEE Congress on Services, pages 167-174. IEEE Computer Society, 2007.

[GL11] V. Gruhn and R. Laue. Detecting Common Errors in Event-Driven Process Chains by Label Analysis. Enterprise Modelling and Information Systems Architectures, 6(1):3-15, 2011.

[HKO07] Thomas Hornung, Agnes Koschmider, and Andreas Oberweis. Rule-based autocompletion of business process models. In CAiSE Forum, volume 247, 2007.

[HLD+05] Martin Hepp, Frank Leymann, John Domingue, Alexander Wahler, and Dieter Fensel. Semantic business process management: A vision towards using semantic web services for business process management. In e-Business Engineering, 2005. ICEBE 2005. IEEE International Conference on, pages 535-540. IEEE, 2005.

[HR07] Martin Hepp and Dumitru Roman. An ontology framework for semantic business process management. Wirtschaftinformatik Proceedings 2007, page 27, 2007.

[JWW+10] T. Jin, J. Wang, N. Wu, M. La Rosa, and A. ter Hofstede. Eficient and accurate retrieval of business process models through indexing. In Proceedings of the 18th CoopIS, Part I, volume 6426 of Lecture Notes in Computer Science, pages 402-409. Springer, 2010.

[KB04] M. Klein and A. Bernstein. Towards high-precision service retrieval. IEEE Internet Computing, 8(1):30-36, 2004.

[KB07] Agnes Koschmider and Emmanuel Blanchard. User assistance for business process model decomposition. In Proceedings of the 1st IEEE International Conference on Research Challenges in Information Science, pages 445-454, 2007.

[KFSO14] Agnes Koschmider, Michael Fellmann, Andreas Schoknecht, and Andreas Oberweis. Analysis of process model reuse: Where are we now, where should we go from here? Decision Support Systems, 66:9-19, 2014.

[KLW+] Christopher Klinkmüller, Henrik Leopold, Ingo Weber, Jan Mendling, and André Ludwig. Listen to me: Improving process model matching through user feedback. In Shazia Wasim Sadiq, Pnina Soffer, and Hagen Völzer, editors, Business Process Management - 12th International Conference, BPM 2014, Haifa, Israel, September 7-11, 2014. Proceedings, volume 8659 of Lecture Notes in Computer Science, pages 84-100. Springer.

[KRG07] Jochen Malte Ku ster, Ksenia Ryndina, and Harald Gall. Generation of business process models for object life cycle compliance. In Gustavo Alonso, Peter Dadam, and Michael Rosemann, editors, Business Process Management, 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007, Proceedings, volume 4714 of Lecture Notes in Computer Science, pages 165-181. Springer, 2007.

[KWW11] Matthias Kunze, Matthias Weidlich, and Mathias Weske. Behavioral similarity - a proper metric. In Proceedings of the 9th

International Conference on Business Process Management, volume 7481 of Lecture Notes in Computer Science, pages 166-181. Springer, 2011.

[LD05] Y. Lin and H. Ding. Ontology-based Semantic Annotation for Semantic Interoperability of Process Models. In CIMCA 2005, pages 162-167, Los Alamitos, CA, USA, 2005. IEEE Computer Society.

[Leo13] Henrik Leopold. Natural Language in Business Process Models: Theoretical Foundations, Techniques, and Applications, volume 168 of LNBIP. Springer, 2013.

[LESM+13] Henrik Leopold, Rami-Habib Eid-Sabbagh, Jan Mendling, Leonardo Guerreiro Azevedo, and Fernanda Araujo Baiao. Detection of naming convention violations in process models for different languages. Decision Support Systems, 56:310-325, 2013.

[LM12] Henrik Leopold and Jan Mendling. Automatic derivation of service candidates from business process model repositories. In Business Information Systems, pages 84-95, 2012.

[LMP14] H. Leopold, J. Mendling, and A. Polyvyanyy. Supporting process model validation through natural language generation. Software Engineering, IEEE Transactions on, 40(8):818-840, Aug 2014.

[LMRR14] Henrik Leopold, Jan Mendling, Hajo A. Reijers, and Marcello La Rosa. Simplifying process model abstraction: Techniques for generating model names. Inf. Syst., 39:134-151, 2014.

[LNW+12] Henrik Leopold, Mathias Niepert, Matthias Weidlich, Jan Mendling, Remco M. Dijkman, and Heiner Stuckenschmidt. Probabilistic optimization of semantic process model matching. In Proceedings of the 10th international conference on Business process management, pages 319-334, 2012.

[LPM13] Henrik Leopold, Fabian Pittke, and Jan Mendling. Towards measuring process model granularity via natural language analysis. In Business Process Management Workshops - BPM 2013 International Workshops, Beijing, China, August 26, 2013, Revised Papers, pages 417-429, 2013.

[LPM14] Henrik Leopold, Fabian Pittke, and Jan Mendling. Towards measuring process model granularity via natural language analysis. In Business Process Management Workshops, pages 417-429. Springer, 2014.

[LSM10] H. Leopold, S. Smirnov, and J. Mendling. Refactoring of Process Model Activity Labels. In NLDB 2010, volume 6177 of LNCS, pages 268-276. Springer, 2010.

[LSM12] Henrik Leopold, Sergey Smirnov, and Jan Mendling. On the refactoring of activity labels in business process models. Information Systems, 37(5):443-459, 2012.

[LSW14] Niels Lohmann, Minseok Song, and Petia Wohed, editors. Business Process Management Workshops - BPM 2013 International Workshops, Beijing, China, August 26, 2013, Revised Papers, volume 171 of Lecture Notes in Business Information Processing. Springer, 2014.

[LVD09] Niels Lohmann, Eric Verbeek, and Remco M. Dijkman. Petri net transformations for business processes - A survey. T. Petri Nets and Other Models of Concurrency, 2:46-63, 2009.

[MB13] Saleem Malik and ImranSarwar Bajwa. Back to origin: Transformation of business process models to business rules. In Marcello La Rosa and Pnina Soffer, editors, Business Process Management Workshops, volume 132 of Lecture Notes in Business Information Processing, pages 611-622. Springer Berlin Heidelberg, 2013.

[MDM13] Monika Malinova, Remco M. Dijkman, and Jan Mendling. Automatic extraction of process categories from process model collections. In Lohmann et al. [LSW14], pages 430-441.

[Men08] J. Mendling. Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness, volume 6 of LNBIP. Springer, 2008.

[Moo09] Daniel L. Moody. The physics of notations: Toward a scientific basis for constructing visual notations in software engineering. IEEE Trans. Software Eng., 35(6):756-779, 2009.

[MRR10] J. Mendling, H. A. Reijers, and J. Recker. Activity Labeling in Process Modeling: Empirical Insights and Recommendations. Information Systems, 35(4):467-482, 2010.

[MRvdA10] J. Mendling, H. A. Reijers, and W. M. P. van der Aalst. Seven Process Modeling Guidelines (7PMG). Information and

Software Technology, 52(2):127-136, 2010.

[PLM13] Fabian Pittke, Henrik Leopold, and Jan Mendling. Spotting terminology deficiencies in process model repositories. In Enterprise, Business-Process and Information Systems Modeling, 2013.

[PLM14] Fabian Pittke, Henrik Leopold, and Jan Mendling. When language meets language: Anti patterns resulting from mixing natural and modeling language. In 5th International Workshop on Process Model Collections: Management and Reuse (PMC-MR 2014). Proceedings, LNBIP. Springer, 2014.

[PLM15] Fabian Pittke, Henrik Leopold, and Jan Mendling. Automatic detection and resolution of lexical ambiguity in process models. IEEE Transactions on Software Engineering, 2015.

[PSFGP10] María Poveda, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Péerez. Common pitfalls in ontology development. In Current Topics in Artificial Intelligence, pages 91-100. Springer, 2010.

[PW11] N. Peters and M. Weidlich. Automatic Generation of Glossaries for Process Modelling Support. Enterprise Modelling and Information Systems Architectures, 6(1):30-46, 2011.

[QAR11] Mu Qiao, Rama Akkiraju, and Aubrey J. Rembert. Towards eficient business process clustering and retrieval: Combining language modeling and structure matching. In Stefanie Rinderle-Ma, Farouk Toumani, and Karsten Wolf, editors, Business Process Management - 9th International Conference, BPM 2011, Clermont-Ferrand, France, August 30 - September 2, 2011. Proceedings, volume 6896 of Lecture Notes in Computer Science, pages 199-214. Springer, 2011.

[RMD11] Hajo A. Reijers, Jan Mendling, and Remco M. Dijkman. Human and automatic modularizations of process models to enhance their comprehension. Inf. Syst., 36(5):881-897, 2011.

[RMKL12] Stefanie Rinderle-Ma, Sonja Kabicher, and Linh Thao Ly. Activity-oriented clustering techniques in large process and compliance rule repositories. In Business Process Management Workshops, pages 14-25. Springer, 2012.

[Ros06] M. Rosemann. Potential Pitfalls of Process Modeling: Part A. Business Process Management Journal, 12(2):249-254, 2006.

[SDMW10] S. Smirnov, R.M. Dijkman, J. Mendling, and M. Weske. Meronymy-Based Aggregation of Activities in Business Process Models. In ER 2010, volume 6412 of LNCS, pages 1-14. Springer, 2010.

[Sil11] Bruce Silver. BPMN Method and Style, with BPMN Implementer's Guide. Cody-Cassidy Press, 2nd edition, January 2011.

[SP10] A. Sinha and A. Paradkar. Use Cases to Process Specifications in Business Process Modeling Notation. In 2010 IEEE International Conference on Web Services, pages 473-480. IEEE, 2010.

[SRW12] Sergey Smirnov, Hajo A. Reijers, and Mathias Weske. From fine-grained to abstract process models: A semantic approach. Inf. Syst., 37(8):784-797, 2012.

[SRWN12] Sergey Smirnov, Hajo A. Reijers, Mathias Weske, and Thijs Nugteren. Business process model abstraction: a definition, catalog, and survey. Distributed and Parallel Databases, 30(1):63-99, 2012.

[SWM12] Sergey Smirnov, Matthias Weidlich, and Jan Mendling. Business process model abstraction based on synthesis from consistent behavioural profiles. International Journal of Cooperative Information Systems, 21, 2012.

[SWMW12] Sergey Smirnov, Matthias Weidlich, Jan Mendling, and Mathias Weske. Action patterns in business process model repositories. Computers in Industry, 63, 2012.

[TF07] Oliver Thomas and Michael Fellmann. Semantic business process management: Ontology-based process modeling using event-driven process chains. IBIS, 4:29-44, 2007.

[van00] W.M.P. van der Aalst. Business Process Management, volume 1806 of LNCS, chapter Workow Verification: Finding Control-Flow Errors Using Petri-Net-Based Techniques, pages 161-183. Springer, 2000.

[vdA13] Wil MP van der Aalst. Business process management: A comprehensive survey. ISRN Software Engineering, 2013, 2013.

[vdVGvdR97] Bram van der Vos, Jon Atle Gulla, and Reind van de Riet. Verification of conceptual models based on linguistic knowledge. Data & Knowledge Engineering, 21(2):147-163, 1997.

[WDM10] M. Weidlich, R.M. Dijkman, and J. Mendling. The ICoP Framework: Identification of Correspondences between Process Models. In CAiSE 2010, volume 6051 of LNCS, pages 483-498. Springer, 2010.

[Wes12] Mathias Weske. Business Process Management: Concepts, Languages, Architectures. Springer, 2nd edition, 2012.

[WHM10] I. Weber, J. Ho_mann, and J. Mendling. Beyond Soundness: on the Verification of Semantic Business Process Models. Distributed and Parallel Databases, 27(3):271-343, 2010.

[WMW11] Matthias Weidlich, Jan Mendling, and Mathias Weske. Eficient consistency measurement based on behavioral profiles of process models. IEEE Trans. Software Eng., 37(3):410-429, 2011.

[WRMR11] Barbara Weber, Manfred Reichert, Jan Mendling, and Hajo A. Reijers. Refactoring large process model repositories. Computers in Industry, 62(5):467-486, 2011.

[WRR08] Barbara Weber, Manfred Reichert, and Stefanie Rinderle-Ma. Change patterns and change support features - enhancing exibility in process-aware information systems. Data Knowl. Eng., 66(3):438-466, 2008.

[WW90] Y. Wand and R. Weber. Studies in Bunge's Treatise on Basic Philosophy, chapter Mario Bunge's Ontology as a Formal Foundation for Information Systems Concepts, pages 123-149. the Poznan Studies in the Philosophy of the Sciences and the Humanities. Rodopi, 1990.

[WW95] Y. Wand and R. Weber. On the deep structure of information systems. Information Systems Journal, 5:203-223, 1995.

[WW02] Y. Wand and R. Weber. Research Commentary: Information Systems and Conceptual Modeling – A Research Agenda. Information Systems Research, 13(4):363-376, 2002.

[YDG12] Zhiqiang Yan, Remco Dijkman, and Paul Grefen. Fast business process similarity search. Distributed and Parallel Databases, 30:105-144, 2012.

[ZWW+10] Haiping Zha, Jianmin Wang, Lijie Wen, Chaokun Wang, and Jiaguang Sun. A workow net similarity measure based on transition adjacency relations. Computers in Industry, 61(5):463-471, 2010.