Carmona, J. The alignment of formal, structured and unstructured process descriptions. A: International Conference on Application and Theory of Petri Nets and Concurrency. "Application and Theory of Petri Nets and Concurrency, 38th International Conference, PETRI NETS 2017: Zaragoza, Spain, June 25–30, 2017: proceedings". Berlín: Springer, 2017, p. 3-11. The final authenticated version is available online at https://doi.org/10.1007/978-3-319-57861-3_1

The Alignment of Formal, Structured and Unstructured Process Descriptions

Josep Carmona

Universitat Politècnica de Catalunya, Barcelona (Spain)
jcarmona@cs.upc.edu

Abstract. Nowadays organizations are experimenting a drift on the way processes are managed. On the one hand, formal notations like Petri nets or Business Process Model and Notation (BPMN) enable the unambiguous reasoning and automation of designed processes. This way of eliciting processes by manual design, which stemmed decades ago, will still be an important actor in the future. On the other hand, regulations require organizations to store their process executions in structured representations, so that they are known and can be analyzed. Finally, due to the different nature of stakeholders within an organization (ranging from the most technical members, e.g., developers, to less technical), textual descriptions of processes are also maintained to enable that everyone in the organization understands their processes.
In this paper I will describe techniques for facilitating the interconnection between these three process representations. This requires interdisciplinary research to connect several fields: business process management, formal methods, natural language processing and process mining.

1 Introduction

With the aim of having individuals from various levels examine their operations, organizations maintain different representations of their processes: while *textual descriptions* of processes are well-suited for non-technical members, they are less appropriate for describing precise aspects of the underlying process [1]. In contrast, *formal and graphical process notations* are unambiguous representations which can be the basis for automating the corresponding processes within the organization [2], but they are oriented to specialized members. Recent studies have not concluded a clear superiority between neither of the two aforementioned notations [3,4]. Finally, the current trend to store all kinds of digital data has made organizations to become more than ever data-oriented, thus dependent on the available techniques to extract value from the data. Process mining is an emerging field which focuses on analyzing the *event logs* corresponding to process executions, with the purpose of extracting, analyzing and enhancing evidence-based process models [5].

In this context, due to the evolving nature of processes, there is a high risk of having deviations between these three different representations, a problem that may have serious consequences for any organization [6]. To have these different descriptions aligned to ensure that everybody shares the same version of the process is not only a desired feature, but also a real challenge originated by the contrasting nature of each process representation.

Likewise, organizations need to keep track of the deviations between different versions of the same process under the same representation (e.g., the winter sales process vs. the summer sales process, or the incorporation of a new form of payment in a process), to bound the flexibility and variability of a running process, or simply to be aware of the evolution of a process over time.

In this paper I will provide an overview of the milestones and current challenges that arise when trying to align these three different process descriptions. I will mainly focus on the algorithmic support for computing alignments across different process descriptions, and only will briefly discuss the case where the process descriptions are the same.

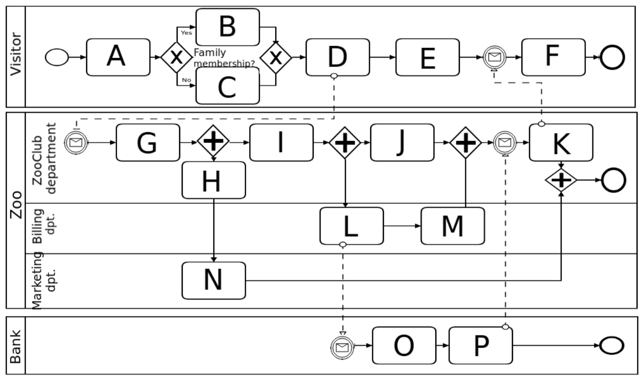
2 Descriptions of Processes

Here we informally describe three types of descriptions to report processes used in organizations. The reader can find a wider view in (which also includes spreadsheets and business rules) in [6].

Graphical Models. There exist a plethora of formal and graphical notations to model processes, like BPMNs [7], EPCs [8], Petri Nets [9], YAWL [10], and many others. In this paper, we will informally use one of them: BPMN. BPMN models are composed from three types of nodes: events, activities and gateways. *Events* (represented as circles) denote something that happens (e.g., time, messages, . . .), rather than *Activities* which are something that is done (represented as rounded-corner rectangles). Finally, the *gateways*, represented as diamond shapes, are used to route the control flow. These elements can be partitioned into pools or lanes, to group activities performed by the same actor (person, department, institution, etc). An example of BPMN is shown in Figure 1. We consider graphical models a formal process description.

Textual Descriptions. Textual descriptions of processes can often be found in organizations [1]. A possibility is to use *written use cases* [11], but those already introduce some structuring that limits the flexibility of the description. Instead, unrestricted textual descriptions like the one shown in Figure 1 can be created by anyone with knowledge on the process. In general, textual descriptions assume a linear description of the sequence of tasks carried out, while concurrency, iteration and other control-flow patterns are expressed in a less precise manner. We consider text as an unstructured process description.

Event Logs. Event logs represent the footprints left by process executions, stored by an information system [5]. As minimal requirement, event logs are formed from *events*, that assign activities to process executions (cases). Additionally, other information can be associated to an event like its timestamp, resource, cost, etc. Part of an event log is reported in Table 1. While in the two previous descriptions, the process is explicitly described, an event log describe implicitly the process, by providing example of its possible executions. We regard event logs as an structured process description.



Tasks:

A: Decide individual or family ticket.
B: Prepare personal information.
C: Prepare family's information.
D: Send information to the ZooClub department.
E: Wait for card.
F: Go home.
G: Enter information into the system
H: Forward information to the marketing department.
I: Send request to billing department.
J: Wait for payment.
K: Deliver ZooClub card.
L: Enter information into the system
M: Send payment request
N: Mail ZooClub magazine
O: Process payment information
P: Charge account

(1) *When a visitor wants to become a member of Barcelona's ZooClub, the following steps must be taken.* (2) *First of all, the customer must decide whether he wants an individual or family membership.* (3) *If he wants an individual membership, he must prepare his personal information.* (4) *If he wants a family membership instead, he should prepare the information for its spouse and spawn as well.* (5) *The customer must then give this information to the ZooClub department.* (6) *The ZooClub enters the visitor's personal data into the system and takes the payment request to the Billing department.* (7) *The ZooClub department also forwards the visitor's information to the marketing department.* (8) *On receiving the request, the billing department also enters the visitor's personal data into their local database.* (9) *After that, the billing department sends the payment request to the bank.* (10) *The bank processes the payment information and, if everything is correct, charges the payment into user's account.* (11) *Once the payment is confirmed, the ZooClub department can print the card and deliver it to the visitor.* (12) *In the meantime, the Marketing department makes a request to mail the Zoo Club's magazine to the visitor's home.* (13) *Once the visitor receives the card, he can go home.*

Fig. 1. Graphical and Textual description of the Zoo process.

3 The Alignment Graph: an Algorithmic Tour

Van der Aa *et al.* recently reported on the fragmentation organizations have with respect to the description of the processes [6]. This will only get worse in the near future, since process-related data in any of the considered forms will become ubiquitous. Hence, automation will be a crucial element to allow stakeholders to catch up with the evolving and flexible nature of processes.

In particular, algorithms for the computation of alignments between (different) process representations will be needed. One possibility, that it is not contemplated here, is to use transformations between the different representations.

Event	Case ID	Activity	Timestamp	Person	Dept.	Data
1	1	A	10-04-2015 9:08am	Miquel	–	–
2	2	A	10-04-2015 10:03am	Sandra	–	–
3	2	B	10-04-2015 10:05am	Sandra	–	Personal
4	1	C	10-04-2015 9:09am	Miquel	–	Family
5	1	D	10-04-2015 9:10am	Miquel	–	–
6	1	G	10-04-2015 9:12am	Ruth	ZooClub	Family
7	1	H	10-04-2015 9:12am	Ruth	ZooClub	Family
8	2	D	10-04-2015 10:06am	Sandra	ZooClub	Personal
9	1	I	10-04-2015 9:18am	Ruth	ZooClub	–
10	1	N	10-04-2015 10:03am	Pere	Marketing	–
11	1	L	10-04-2015 11:32am	Teresa	Billing	34567-e
12	1	J	10-04-2015 2:01pm	Ruth	–	–
13	1	M	10-04-2015 7:06pm	Teresa	–	–

Table 1. Part of an event log for the Zoo process.

For instance, there are mature techniques that transform an event log into a process model [5]. Likewise, recent techniques have appeared to transform a textual description into a process model [1] and back [12]. Although they represent a very useful toolbox that may help into integrating different sources of process information, these transformations do not always guarantee the preservation of the main aspects of the original process description.

For the three process representations described in the previous section, we now show the main techniques available for facilitating the matching between process representations. Figure 3 summarizes them into the *alignment graph*.

Formal Models vs. Event logs. The seminal work in [13] proposed the notion of alignment between process models described as Petri nets and event logs, and developed a technique to compute optimal alignments for a particular class of process models. For each trace σ in the log, the approach consists on exploring the synchronous product of model's state space and σ. In the exploration, the shortest path is computed using the A^* algorithm, once costs for model and log moves are defined. Several optimizations have been proposed to the basic approach: for instance, the use of Integer Linear Programming (ILP) techniques on each visited state to prune the search space [13]. Alternatively, an approach based on partial orders which verbalizes the differences computed has been proposed in [14]. Recently, an approach fully based on ILP has been presented, which significantly reduces the complexity of computing alignments [15] at the expense of dropping the optimality guarantee. Some heuristics that cannot guarantee always the derivation of real alignments but work well in practice can be found in the literature [16,17,18].

Textual Descriptions vs. Formal Models. The seminal work [19,20] was the first one in proposing an algorithm for aligning textual descriptions and process models, with the particular aim of detecting inconsistencies between both representations. The technique uses a linguistic analysis (NLP) previous to a *best-first search* technique to compute an optimal alignment. In contrast to [19,20], the approach in [21] encodes the problem of computing an alignment as the resolution of an ILP model, representing a significant reduction (of several orders of magnitude) in the time requirements for computing an alignment.

Textual Descriptions vs. Event logs. This is a less explored field. However, techniques applied in related problems may be applicable here. For instance, linguistic techniques for extracting the temporal relations between the main events in a text can be used to derive the behavioral patterns [22,23], which can then be compared to the *log-based ordering relations* of a log [5]. Those can be the inputs to ILP matching techniques similar to the ones applied to the previous problems.

The previous family of techniques focused on techniques for aligning across different process descriptions. For completeness, we now report some of the techniques used for aligning process descriptions on the same notation.

Formal Models vs. Formal Models. There has been a plethora of techniques in the last decade to facilitate the matching between process models. For BPMN notation, for instance, the reader can find a good summary in [24]. The techniques have been extensively applied in the context of process model repositories, e.g. [25]. Overall, the techniques range from graph-edit distance, event structures, behavioral profiles and many more.

Event Logs vs. Event Logs. In the last years some contributions have focused into aligning event logs, in order to extract differences that may represent expected or unexpected process variations. The work in [26] uses event structures to verbalize differences, while less fine-grained techniques can also be used by comparing log-based ordering relations. On a different perspective, the use of concept-drift techniques based on statistical tests together with adaptive windowing [27] can also be used to detect inconsistencies between event logs [28].

Textual Descriptions vs. Textual Descriptions. Again, in the scope of textual descriptions of processes this is a less explored family of techniques. Due to the widespread use of textual documentations of processes in organization, techniques for automatically providing inconsistencies between textual descriptions can be a very important tool to improve the understandability of the processes [4]. As for previous techniques that need to deal with textual descriptions, the use of linguistic analysis as an input for later matching techniques (e.g., ILP) may be a promising direction. On a more general setting, computing the semantic similarity between two texts is a classical task in NLP and Information Retrieval (IR) fields. In IR, this is typically tackled by term-frequency based approaches, that compare distributions of words in the documents, so documents are considered similar if they contain similar words in a similar distribution. In NLP field, approaches based on n-gram occurrences in both texts have also been used to evaluate results of Machine Translation (BLEU [29]) or Summarization (ROUGE [30] systems, by comparing them to gold standard human-produced documents for the same tasks. More recently, more accurate semantic comparison

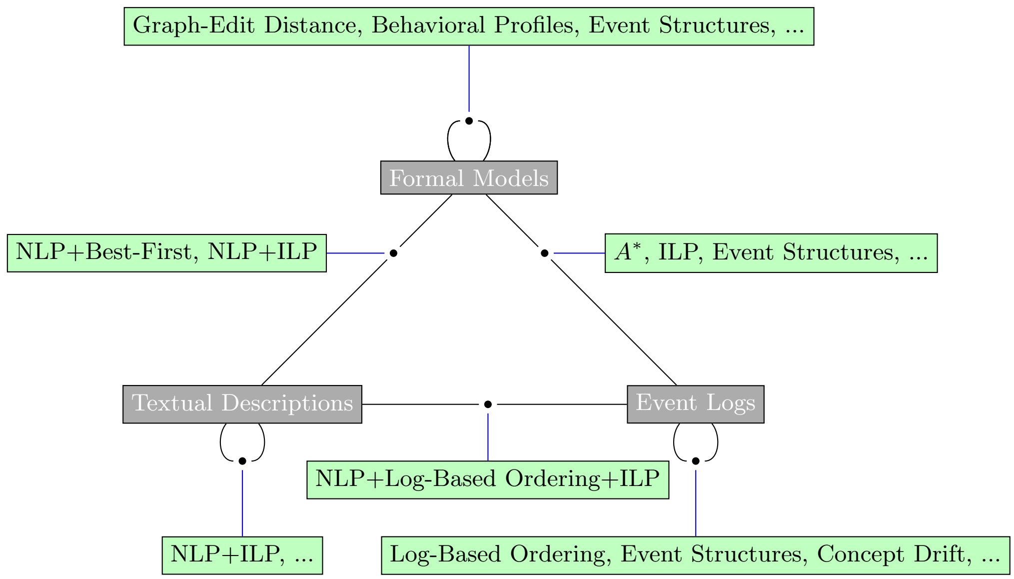


Fig. 2. The Alignment Graph.

of texts has been approached by tasks such as Textual Entailment [31] (decide whether one text implies the other or not) and Semantic Textual Similarity [32] (decide to which extent two texts are equivalent –they say the same, one is contained in the other, one implies the other, they talk about the same topic but do not say the same, they are unrelated...). For this more advanced comparison, heavier NLP machinery is required (syntactic parsers, semantic analyzers, ontologies, word embeddings, etc).

4 Outlook

In this paper I have summarized the current algorithmic support for the alignment of different process descriptions. Overall, these techniques often need to combine several disciplines, like linguistic analysis, graph-based techniques, mathematical optimization, statistics, machine learning, to name a few.

In spite of some successful cases, most of the techniques need to be improved or reconsidered in some particular scenarios: when the quality of the alignments derived needs to be secured, when the techniques are meant to be applicable on the large, or in an online setting.

The progress in the techniques enumerated in this paper will have a direct impact in the way organizations deal with their processes, enabling the continuous awareness of the processes by any agent. For particular fields like healthcare, education, administration and similar, the influence can be even stronger, due to the enormous importance processes have in parallel with the heterogeneity of the existing stakeholders.

Acknowledgements

I would like to thank to the organization of the Petri Nets and ACSD conferences for both the invitation to give a keynote and to write an article in the conference proceedings.

I would like to thank some researchers that collaborate with my group on some of the topics of this paper: Han van der Aa, Andrea Burattin, Thomas Chatain, Boudewijn van Dongen, Henrik Leopold, Hajo A. Reijers and Barbara Weber. Likewise, I would like to thank the local collaborators LLuís Padró, Josep Sànchez-Ferreres, David Sanchez-Charles and Farbod Taymouri.

This work has been partially supported by funds from the Spanish Ministry for Economy and Competitiveness (MINECO), the European Union (FEDER funds) under grant COMMAS (ref. TIN2013-46181-C2-1-R).

References

1. Leopold, H., Mendling, J., Polyvyanyy, A.: Supporting process model validation through natural language generation. IEEE Trans. Software Eng. **40**(8) (2014) 818–840
2. Dumas, M., Rosa, M.L., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer (2013)
3. Ottensooser, A., Fekete, A.D.: Comparing Readability of Graphical and Sentential Process Design Notations: Data Analysis Report. University of Sydney, School of Information Technologies (2010)
4. Ottensooser, A., Fekete, A., Reijers, H.A., Mendling, J., Menictas, C.: Making sense of business process descriptions: An experimental comparison of graphical and textual notations. Journal of Systems and Software **85**(3) (2012) 596–606
5. van der Aalst, W.M.P.: Process Mining - Discovery, Conformance and Enhancement of Business Processes. Springer (2011)
6. van der Aa, H., Leopold, H., Mannhardt, F., Reijers, H.A.: On the fragmentation of process information: Challenges, solutions, and outlook. In: Enterprise, Business-Process and Information Systems Modeling - 16th International Conference, BPMDS 2015, 20th International Conference, EMMSAD 2015, Held at CAiSE 2015, Stockholm, Sweden, June 8-9, 2015, Proceedings. (2015) 3–18
7. OMG: Business process model and notation (bpmn) version 2.0 (2011)
8. Scheer, A.W.: Aris–Business Process Modeling. Springer-Verlag (1998)
9. Murata, T.: Petri nets: Properties, analysis and applications. Proceedings of the IEEE **77**(4) (April 1989) 541–574
10. van der Aalst, W.M.P., ter Hofstede, A.H.M.: Yawl: yet another workflow language. Inf. Syst. **30**(4) (2005) 245–275
11. Cockburn, A.: Writing Effective Use Cases. Addison-Wesley Longman (2000)
12. Friedrich, F., Mendling, J., Puhlmann, F.: Process model generation from natural language text. In: Advanced Information Systems Engineering - 23rd International Conference, CAiSE 2011, London, UK, June 20-24, 2011. Proceedings. (2011) 482–496

13. Adriansyah, A.: Aligning observed and modeled behavior. PhD thesis, Technische Universiteit Eindhoven (2014)
14. Armas-Cervantes, A., Baldan, P., Dumas, M., García-Bañuelos, L.: Diagnosing behavioral differences between business process models: An approach based on event structures. Inf. Syst. **56** (2016) 304–325
15. van Dongen, B., Carmona, J., Chatain, T., Taymouri, F.: Aligning modeled and observed behavior: A compromise between complexity and quality. In: Advanced Information Systems Engineering - 29th International Conference, CAiSE 2017, Essen, Germany, June 12-16, 2017. Proceedings. (2017) In Press
16. Rozinat, A., van der Aalst, W.M.P.: Conformance checking of processes based on monitoring real behavior. Inf. Syst. **33**(1) (2008) 64–95
17. Taymouri, F., Carmona, J.: A recursive paradigm for aligning observed behavior of large structured process models. In: Business Process Management - 14th International Conference, BPM 2016, Rio de Janeiro, Brazil, September 18-22, 2016. Proceedings. (2016) 197–214
18. Verbeek, H.M.W., van der Aalst, W.M.P.: Merging alignments for decomposed replay. In: Application and Theory of Petri Nets and Concurrency - 37th International Conference, PETRI NETS 2016, Toruń, Poland, June 19-24, 2016. Proceedings. (2016) 219–239
19. van der Aa, H., Leopold, H., Reijers, H.A.: Detecting inconsistencies between process models and textual descriptions. In: Business Process Management - 13th International Conference, BPM 2015, Innsbruck, Austria, August 31 - September 3, 2015, Proceedings. (2015) 90–105
20. van der Aa, H., Leopold, H., Reijers, H.A.: Comparing textual descriptions to process models – the automatic detection of inconsistencies. Information Systems (2016) In Press.
21. Sànchez-Ferreres, J., Carmona, J., Padró, L.: Aligning textual and graphical descriptions of processes through ilp techniques. In: Advanced Information Systems Engineering - 29th International Conference, CAiSE 2017, Essen, Germany, June 12-16, 2017. Proceedings. (2017) In Press
22. UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M., Pustejovsky, J.: Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, Georgia, USA, Association for Computational Linguistics (June 2013) 1–9
23. Mirza, P.: Extracting Temporal and Causal Relations between Events. PhD thesis, International Doctorate School in Information and Communication Technologies, University of Trento (2016)
24. Becker, M., Laue, R.: A comparative survey of business process similarity measures. Comput. Ind. **63**(2) (February 2012) 148–167
25. Rosa, M.L., Dumas, M., Ekanayake, C.C., García-Bañuelos, L., Recker, J., ter Hofstede, A.H.M.: Detecting approximate clones in business process model repositories. Inf. Syst. **49** (2015) 102–125
26. van Beest, N.R.T.P., Dumas, M., García-Bañuelos, L., Rosa, M.L.: Log delta analysis: Interpretable differencing of business process event logs. In: Business Process Management - 13th International Conference, BPM 2015, Innsbruck, Austria, August 31 - September 3, 2015, Proceedings. (2015) 386–405
27. Bifet, A., Gavaldà, R.: Learning from time-changing data with adaptive windowing. In: Proceedings of the Seventh SIAM International Conference on Data Mining, April 26-28, 2007, Minneapolis, Minnesota, USA. (2007) 443–448

28. Maaradji, A., Dumas, M., Rosa, M.L., Ostovar, A.: Fast and accurate business process drift detection. In: Business Process Management - 13th International Conference, BPM 2015, Innsbruck, Austria, August 31 - September 3, 2015, Proceedings. (2015) 406–422
29. Papineni, K., Roukos, S., Ward, T., Zhu, W.: Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA. (2002) 311–318
30. Lin, C.Y.: Rouge: A package for automatic evaluation of summaries. In: Proceedings of the ACL-04 Workshop. Association for Computational Linguistics. (2004) 74–81
31. Negri, M., Marchetti, A., Mehdad, Y., Bentivogli, L., Giampiccolo, D.: Semeval-2012 task 8: Cross-lingual textual entailment for content synchronization. In: Proceedings of *SEM 2012. (2012)
32. Agirre, E., Banea, C., Cardie, C., Cer, D.M., Diab, M.T., Gonzalez-Agirre, A., Guo, W., Lopez-Gazpio, I., Maritxalar, M., Mihalcea, R., Rigau, G., Uria, L., Wiebe, J.: Semeval-2015 task 2: Semantic textual similarity, english, spanish and pilot on interpretability. In: Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, Denver, Colorado, USA, June 4-5, 2015. (2015) 252–263