# Lecture Notes in Business Information Processing 323

Series Editors

More information about this series at http://www.springer.com/series/7911

Han van der Aa

# Comparing and Aligning Process Representations

## Foundations and Technical Solutions

Springer

Han van der Aa
Humboldt University of Berlin
Berlin, Berlin
Germany

This book is a revised version of the PhD dissertation written by the author at the Vrije Universiteit Amsterdam. ISBN: 9402808671

ISSN 1865-1348 ISSN 1865-1356 (electronic)
Lecture Notes in Business Information Processing
ISBN 978-3-319-94633-7 ISBN 978-3-319-94634-4 (eBook)
https://doi.org/10.1007/978-3-319-94634-4

Library of Congress Control Number: 2018947442



Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

Processes within organizations can be highly complex chains of inter-related steps, involving numerous stakeholders and information systems. Owing to this complexity, having access to the right information is vital to the proper execution and effective management of an organization's business processes. A major challenge in this regard is that information on a single process is often spread out over numerous models, documents, and systems. This phenomenon results from efforts to provide a variety of process stakeholders with the information that is relevant to them, in a suitable format. However, this disintegration of process information also has considerable disadvantages for organizations. In particular, it can lead to severe maintenance issues, reduced execution efficiency, and negative effects on the quality of process results. Against this background, this doctoral thesis focuses on the spread of process information in organizations and, in particular, on the mitigation of the negative aspects associated with this phenomenon. The main contributions of this thesis are five techniques that focus on the alignment and comparison of process information from different informational artifacts. Each of these techniques tackles a specific scenario involving multiple informational artifacts that contain process information in different representation formats. Among others, we present automated techniques for the detection of inconsistencies between process models and textual process descriptions, the alignment of process performance measurements to process models, conformance-checking in the context of uncertainty, and the matching of process models through the analysis of event-log information. We demonstrate the efficacy and usefulness of these techniques through quantitative evaluations involving data obtained from real-world settings. Altogether, the presented work provides important contributions for the analysis, comparison, and alignment of process information in various representation formats through the development of novel concepts and techniques. The contributions, furthermore, provide a means for organizations to improve the efficiency and quality of their processes.

May 2018 Han van der Aa

*The original version of the book frontmatter has been revised: A note has been added to the imprint page. The erratum to the book front matter is available at https://doi.org/10.1007/978-3-319-94634-4_9*

# Acknowledgments

The majority of the research presented in this book was conducted as part of my time as a PhD candidate. Looking back at those years, it is clear that I owe a lot to a great number of even greater people who helped and supported me.

First of all, I am immensely grateful to my supervisors, Hajo Reijers and Henrik Leopold. Hajo: thanks for inviting me to be the first member of your group, for supporting my development as a researcher and a teacher, for giving me the freedom to pursue my own research interests, and for the great collaboration. Henrik: thanks for the countless hours you invested in me, for the many trips we took together, and for guiding me in the academic world as well as in other matters of life.

Although I am only at the start of my academic career, I have already had the opportunity to collaborate with numerous excellent researchers from all over Europe and beyond. My gratitude goes out to the co-authors of the publications included in this work: to Adela del-Río-Ortega, Manuel Resinas, and Antonio Ruiz-Cortés from the University of Seville, to Felix Mannhardt from the Eindhoven University of Technology, to Jan Mendling from the Vienna University of Business and Economics, and to Roee Shraga, Tomer Sagi, and Avigdor Gal from the Technion – Israel Institute of Technology. Furthermore, I want to express my great appreciation for Jan and Avi, who invited me to visit their research groups in Vienna and Haifa, respectively.

Closer to home, I am grateful to my former colleagues at the Vrije Universiteit Amsterdam with special mention of Banu, Unal, Dennis, Inge, Ermeson, Isaac, Iris, and Jelmer. Furthermore, thanks go out to the administrative staff, in particular to Caroline and Mojca, for providing me with support as well as all the office essentials and not-so-essentials that allow academics to conduct their work. Outside of the academic world, I am most grateful to my friends, especially Justine, Jennifer, Franzi, and Henrik from Amsterdam, Frank, Koen, Jeroen G., and Jeroen R. from my hometown, as well as Tijn and Wim from Eindhoven. The final dedication goes out to my mother, father, brother, and the rest of my family, for their support and encouragement during my research endeavors.

# Contents


**1 Introduction** . . . . . . . . . . 1
1.1 Motivation . . . . . . . . . . 1
1.2 Contributions. . . . . . . . . . 2
1.3 Outline . . . . . . . . . . 5

**2 Background** . . . . . . . . . . 9
2.1 Process Information in Organizations . . . . . . . . . . 10
2.1.1 Business Process Management. . . . . . . . . . 10
2.1.2 Process Information . . . . . . . . . . 11
2.1.3 Capturing Process Information. . . . . . . . . . 12
2.1.4 The Problem of Fragmented Process Information . . . . . . . . . . 15
2.2 Natural Language Processing. . . . . . . . . . 19
2.2.1 Part-of-Speech Tagging . . . . . . . . . . 20
2.2.2 Sentence Parsing . . . . . . . . . . 23
2.2.3 Reference Resolution . . . . . . . . . . 25
2.2.4 Natural Language Ambiguity . . . . . . . . . . 26
2.2.5 Applications in Business Process Management . . . . . . . . . . 27
2.3 Matching . . . . . . . . . . 28
2.3.1 Matching Example. . . . . . . . . . 29
2.3.2 The Matching Task . . . . . . . . . . 30
2.3.3 Similarity Measures . . . . . . . . . . 31
2.3.4 Matching in Business Process Management. . . . . . . . . . 36

**3 Comparing Process Models to Textual Process Descriptions** . . . . . . . . . . 39
3.1 Problem Illustration . . . . . . . . . . 40
3.2 Inconsistency-Detection Approach . . . . . . . . . . 41
3.2.1 Linguistic Analysis . . . . . . . . . . 42
3.2.2 Activity-Sentence Alignment. . . . . . . . . . 45
3.2.3 Inconsistency Detection . . . . . . . . . . 48
3.3 Evaluation. . . . . . . . . . 52
3.3.1 Test Collection . . . . . . . . . . 52
3.3.2 Setup . . . . . . . . . . 53
3.3.3 Results. . . . . . . . . . 55
3.3.4 Discussion . . . . . . . . . . 58
3.4 Limitations . . . . . . . . . . 59
3.5 Related Work . . . . . . . . . . 60
3.6 Summary . . . . . . . . . . 61

**4 Conformance Checking Based on Uncertain Event-Activity Mappings** . . . . 63
4.1 Problem Illustration . . . . 64
4.2 Conformance-Checking Technique . . . . 66
4.2.1 Capturing Mapping Uncertainty Using Behavioral Spaces . . . 66
4.2.2 Using Behavioral Spaces for Conformance Checking . . . . 68
4.3 Evaluation . . . . 70
4.3.1 Test Collection . . . . 70
4.3.2 Setup . . . . 70
4.3.3 Results . . . . 72
4.4 Limitations . . . . 73
4.5 Related Work . . . . 74
4.6 Summary . . . . 75

**5 Dealing with Ambiguity in Textual Process Descriptions** . . . . 77
5.1 Problem Illustration . . . . 78
5.2 Capturing Ambiguity Using Behavioral Spaces . . . . 79
5.2.1 Parsing Textual Process Descriptions . . . . 80
5.2.2 Computing Statement Interpretations . . . . 82
5.2.3 Generating a Behavioral Space . . . . 86
5.3 Conformance Checking Using Behavioral Spaces . . . . 88
5.4 Pruning Behavioral Spaces Based on Information Gain . . . . 90
5.4.1 Conformance-Checking Uncertainty . . . . 90
5.4.2 Information Gain . . . . 91
5.5 Evaluation . . . . 92
5.5.1 Test Collection . . . . 93
5.5.2 Conformance Evaluation . . . . 93
5.5.3 Pruning Evaluation . . . . 96
5.6 Limitations . . . . 98
5.7 Related Work . . . . 100
5.8 Summary . . . . 101

**6 Transforming and Aligning Process Performance Indicators** . . . . 103
6.1 Problem Illustration . . . . 104
6.2 Template-Based PPI Definitions . . . . 106
6.2.1 Template Slots . . . . 106
6.2.2 Slot Domains . . . . 108
6.3 Transformation Approach . . . . 111
6.3.1 Semantic Annotation . . . . 112
6.3.2 Domain Value Resolution . . . . 115
6.3.3 Operationalization and Extensibility . . . . 118
6.4 Evaluation . . . . 119
6.4.1 Test Collection . . . . 120
6.4.2 Setup . . . . 120
6.4.3 Results . . . . 122

6.5 Limitations . . . . . 124
6.6 Related Work . . . . . 124
6.7 Summary . . . . . 125

**7 Process Model Matching Using Event-Log Information** . . . . . 127
7.1 Problem Illustration . . . . . 128
7.2 Event Log-Based Matching . . . . . 129
7.2.1 Positional Similarity . . . . . 130
7.2.2 Occurrence Similarity . . . . . 130
7.2.3 Duration Similarity . . . . . 131
7.2.4 Attribute Name Similarity . . . . . 132
7.2.5 Attribute Value Similarity . . . . . 133
7.2.6 Prerequisites Similarity . . . . . 134
7.2.7 Operationalization . . . . . 135
7.3 Evaluation . . . . . 135
7.3.1 Test Collection . . . . . 135
7.3.2 Setup . . . . . 136
7.3.3 Results . . . . . 137
7.4 Limitations . . . . . 140
7.5 Related Work . . . . . 141
7.6 Summary . . . . . 141

**8 Conclusion** . . . . . 143
8.1 Summary of Results . . . . . 143
8.2 Implications . . . . . 145
8.2.1 Implications for Practice . . . . . 146
8.2.2 Implications for Research . . . . . 147
8.3 Future Research . . . . . 148
8.3.1 Technical Perspective . . . . . 149

**Erratum to: Comparing and Aligning Process Representations** . . . . . E1

**References** . . . . . 151

# Acronyms

| | |
|---|---|
| BPM | Business Process Management |
| BPMN | Business Process Model and Notation |
| CFG | Context-Free Grammar |
| FLM | First-Line Matcher |
| IDF | Inverse Document Frequency |
| IG | Information Gain |
| IS | Information Systems |
| IT | Information Technology |
| LSH | Locality Sensitive Hashing |
| HMM | Hidden Markov Model |
| MWBM | Maximum Weighted Bipartite-graph Match |
| NLP | Natural Language Processing |
| ORE | Ontobuilder Research Environment |
| PCFG | Probabilistic Context-Free Grammar |
| PPI | Process Performance Indicator |
| SCOR | Supply Chain Operations Reference |
| SLM | Second-Line Matcher |
| 7PMG | Seven Process Modeling Guidelines |
| TTF | Task-Technology Fit |

# List of Figures


Fig. 2.1 Overview of a generic alignment task . . . . . . . . . . . . . . . . . . . . . . 9
Fig. 2.2 Hierarchy between data, information, and knowledge (adapted from [55]) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
Fig. 2.3 Main constructs in Task-Technology Fit (TTF) theory . . . . . . . . . . . 13
Fig. 2.4 Loan application process with fragmented process information . . . . . 15
Fig. 2.5 Example of a parse tree. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
Fig. 2.6 Two process models and their correspondences (adapted from [147]) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
Fig. 3.1 A textual and a model-based description of a bicycle manufacturing process . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41
Fig. 3.2 Overview of the proposed approach. . . . . . . . . . . . . . . . . . . . . . . . 42
Fig. 3.3 Simplified parse tree for sentence . . . . . . . . . . . . . . . . . . . . . . . . . 44
Fig. 3.4 Correspondences disallowed by ordering constraints . . . . . . . . . . . . 47
Fig. 3.5 Precision-recall graph for the detection of missing activities (activity level) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56
Fig. 3.6 Precision-recall graph for the detection of model-text pairs with missing activities (process level) . . . . . . . . . . . . . . . . . . . . . . 56
Fig. 4.1 Process model for a simplified order handling process . . . . . . . . . . . 65
Fig. 4.2 Overview of the evaluation setup . . . . . . . . . . . . . . . . . . . . . . . . . 71
Fig. 4.3 Evaluation results for deterministic conformance checking . . . . . . . . 72
Fig. 5.1 Exemplary description of a claims handling process. . . . . . . . . . . . . 78
Fig. 5.2 Steps involved to construct a behavioral space from a textual description . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
Fig. 5.3 A behavioral space as a collection of *m* process interpretations . . . . . 87
Fig. 5.4 Visualization of three sets of conforming traces for cases with scope ambiguity. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96
Fig. 5.5 Comparison of uncertainty resolution using max. IG and random selection. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 98
Fig. 6.1 Process model for an order handling process . . . . . . . . . . . . . . . . 104
Fig. 6.2 Semantic concepts in a PPI definition . . . . . . . . . . . . . . . . . . . . . . 107
Fig. 6.3 Overview of the proposed transformation approach . . . . . . . . . . . . . 111
Fig. 6.4 Fragment of a semantic prior . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
Fig. 7.1 Two process models and their correspondences. . . . . . . . . . . . . . . . 128
Fig. 7.2 Recall scores for top-*k* results . . . . . . . . . . . . . . . . . . . . . . . . . . 139

# List of Tables


Table 1.1 Overview of the five presented techniques . . . . . . . . . . . . . . . . . . 3
Table 2.1 Informational artifacts for a purchasing approval process (from [9]). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16
Table 2.2 Overview of the Penn Treebank tagset (from [127, p. 131]) . . . . . . 21
Table 2.3 Possible tags for an exemplary sentence, with correct tags in bold . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
Table 2.4 Exemplary Stanford dependencies (from [70]) . . . . . . . . . . . . . . . 25
Table 2.5 Comparison of syntactic and semantic similarity scores . . . . . . . . . 33
Table 2.6 Overview of semantic word relations. . . . . . . . . . . . . . . . . . . . . . 34
Table 3.1 Main Stanford Dependencies used for anaphora resolution . . . . . . . 43
Table 3.2 Examplary similarity matrices with correspondences in bold. . . . . . 50
Table 3.3 Fragment of the similarity matrix for the running example . . . . . . . 51
Table 3.4 Overview of the test collection . . . . . . . . . . . . . . . . . . . . . . . . . 53
Table 3.5 Predictor performance evaluation results . . . . . . . . . . . . . . . . . . 57
Table 3.6 Highest $F_1$-measures for different configurations and predictors. . . . 57
Table 5.1 Activities in the running example . . . . . . . . . . . . . . . . . . . . . . . 80
Table 5.2 Exemplary outcomes of behavioral statement parsing . . . . . . . . . . 81
Table 5.3 Parallel indicators used in [93] and their classification . . . . . . . . . . 85
Table 5.4 Overview of the test collection . . . . . . . . . . . . . . . . . . . . . . . . . 93
Table 5.5 Evaluation results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 95
Table 6.1 PPIs for the order handling example . . . . . . . . . . . . . . . . . . . . . 104
Table 6.2 Example of a structured notation for PPI2 . . . . . . . . . . . . . . . . . . 105
Table 6.3 PPI templates and examples . . . . . . . . . . . . . . . . . . . . . . . . . . . 107
Table 6.4 Domains associated with template slots . . . . . . . . . . . . . . . . . . . 109
Table 6.5 Tag set used for semantic annotation. . . . . . . . . . . . . . . . . . . . . . 113
Table 6.6 Domain value resolution for PPI2 . . . . . . . . . . . . . . . . . . . . . . . . 116
Table 6.7 Overview of the test collection . . . . . . . . . . . . . . . . . . . . . . . . . 120
Table 6.8 Evaluation results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 122
Table 7.1 Characteristics of the test collection. . . . . . . . . . . . . . . . . . . . . . . 136
Table 7.2 Evaluation results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 138