LNCS 10650

Heinrich C. Mayr
Giancarlo Guizzardi
Hui Ma
Oscar Pastor (Eds.)

Conceptual Modeling

36th International Conference, ER 2017
Valencia, Spain, November 6–9, 2017
Proceedings

Springer

Lecture Notes in Computer Science 10650

More information about this series at http://www.springer.com/series/7409

Heinrich C. Mayr · Giancarlo Guizzardi
Hui Ma · Oscar Pastor (Eds.)

Conceptual Modeling

36th International Conference, ER 2017
Valencia, Spain, November 6–9, 2017
Proceedings

Springer

Editors
Heinrich C. Mayr
University of Klagenfurt
Klagenfurt
Austria

Hui Ma
Victoria University of Wellington
Wellington
New Zealand

Giancarlo Guizzardi
Free University of Bozen-Bolzano
Bozen-Bolzano
Italy

Oscar Pastor
Valencia University of Technology
Valencia
Spain

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-69903-5 ISBN 978-3-319-69904-2 (eBook)
https://doi.org/10.1007/978-3-319-69904-2

Library of Congress Control Number: 2017957732

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

Printed on acid-free paper

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

Preface

The International Conference on Conceptual Modeling (ER) is the leading global forum for current research on conceptual modeling (CM) and trendsetting CM applications. The topics of interest span the entire spectrum of CM: theoretical and ontological foundation, methods and tools for developing and communicating conceptual models and meta models, techniques for transforming conceptual models into effective implementations, and the impact of CM techniques on databases, business strategies, and information systems development. The ER conference series has been held at a variety of superb locations, rotating in successive years between Europe, the Asia-Pacific region, and the Americas, and attracting an international community of scholars.

This volume contains the research and technical papers comprising the main program of ER 2017 in its 36th conference edition held during November 6–9, 2017, in the beautiful city of Valencia, Spain. More than 450 researchers from all over the world followed our call for papers and submitted 153 papers about their latest research results. Each paper was carefully reviewed by at least three members of the Program Committee, which consisted of renowned scientists from more than 40 nations. Finally, 28 papers, i.e., about 18%, were selected as full papers to be presented at the conference and to be included in this volume. An additional 10 submissions were accepted as short papers. The quality of these 38 papers is a tribute to the authors and also to the reviewers who guided any necessary improvements.

Focal points of these papers are: (1) CM and ontologies in the context of requirements analysis, business processes, and other domains; (2) foundations of CM, for example, regarding multi-level modeling; (3) CM methodology with a broad spectrum of innovative answers to interesting research questions; (4) ontologies; and (5) model efficiency.

This volume would not have materialized without the support of many people. First, we are very grateful to all the authors for their continuous commitment and intensive work. Second, we would like to thank the Program Committee members and additional reviewers for providing timely and in-depth assessments. Furthermore, we thank all the people and sponsors who helped in the organization of ER 2017. Without all that effort there would have been no substance for this volume and no success for ER 2017. Last but not least, we are greatly indebted to the five invited speakers, Prof. Lois Delcambre (USA), Prof. Josef Mitterer (Austria), Prof. Antoni Olivé (Spain), Francisco Garcia-Moran (Spain), and Prof. Yair Wand (Canada), for accepting our invitation to address this conference.

September 2017

Heinrich C. Mayr
Giancarlo Guizzardi
Hui Ma
Oscar Pastor

Organization

Program Committee

Jacky Akoka	CNAM and TEM, France
Raian Ali	Bournemouth University, UK
Joao Paulo Almeida	Federal University of Espirito Santo, Brazil
Yuan An	Drexel University, USA
Joao Araujo	Universidade Nova de Lisboa, Portugal
Alessandro Artale	Free University of Bolzano-Bozen, Italy
Claudia P. Ayala	Technical University of Catalunya, Spain
Fatma Başak Aydemir	Utrecht University, The Netherlands
Doo-Hwan Bae	KAIST, South Korea
Fernanda Araujo Baiao	UNIRIO, Brazil
Zhifeng Bao	RMIT University, Australia
Judith Barrios Albornoz	University of Los Andes, Colombia
Ladjel Bellatreche	LIAS/ENSMA, France
Nelly Bencomo	Aston University, UK
Kawtar Benghazi	Universidad de Granada, Spain
Sandro Bimonte	IRSTEA, France
Mokrane Bouzeghoub	UVSQ/CNRS, France
Shawn Bowers	Gonzaga University, USA
Stephane Bressan	National University of Singapore
Cristina Cabanillas	Wirtschaftsuniversität Wien, Austria
Diego Calvanese	Free University of Bozen-Bolzano, Italy
Maria Luiza Campos	Federal University of Rio de Janeiro, Brazil
Luca Cernuzzi	Universidad Católica, Asunción, Paraguay
Vinay Chaudhri	Independent Researcher, USA
Roger Chiang	AIS, USA
Suphamit Chittayasothorn	King Mongkut's Institute of Technology Ladkrabang, Thailand
Dickson K.W. Chiu	The University of Hong Kong, SAR China
Byron Choi	Hong Kong Baptist University, SAR China
Isabelle Comyn-Wattiau	ESSEC Business School, France
Nelly Condori-Fernández	Universidade da Coruña, Spain
Dolors Costal	Universitat Politècnica de Catalunya, Spain
Alfredo Cuzzocrea	ICAR-CNR and University of Calabria, Italy
Fabiano Dalpiaz	Utrecht University, The Netherlands
Karen Davis	Miami University, USA
Valeria De Antonellis	University of Brescia, Italy
Sergio De Cesare	Brunel University, UK
Jose Luis de La Vara	Carlos III University of Madrid, Spain

Hwan-Seung Yong	Ewha Womans University, South Korea
Eric Yu	University of Toronto, Canada
Yanchun Zhang	Victoria University, Australia
Iyad Zikra	Stockholm University, Sweden

Additional Reviewers

Abdelahad, Corina
Artale, Alessandro
Asprino, Luigi
Baek, Youngmin
Bianchini, Devis
Borges, Marcos R.S.
Bork, Dominik
Böhmer, Kristof
Calvanese, Diego
Chu, Lisa
Chuah, Seong Ping
Corman, Julien
Estrada Torres, Irene Bedilia
Feltus, Christophe
Fernandes, Filipe
Fernandez, Pablo
García, José María
Huemer, Christian
Kaczmarek, Monika
Khodabandelou, Ghazaleh
Khouri, Selma
Koh, Judice
Köpke, Julius
Mazak, Alexandra
Melchiori, Michele
Nalchigar, Soroosh
Noguera, Manuel
Oriol, Xavier
Ouhammou, Yassine
Panach Navarrete, Jose Ignacio
Piras, Luca
Pittl, Benedikt
Prince Sales, Tiago
Razo-Zapata, Iván S.
Russo, Alessandro
Sangat, Prajwol
Shin, Donghwan
Silveira, Denis
Souza, Vitor E. Silva
Stertz, Florian
Troya, Javier
Unger, Moshe
Walch, Michael
Wally, Bernhard
Winter, Karolin
Wolny, Sabine

Sponsors

Main Organizers

Invited Talks

Conceptual Modeling?
When We are Awash in Information?

Lois Delcambre

Computer Science Department,
Portland State University
lmd@pdx.edu

Abstract. We challenge the traditional who/what/why of conceptual modeling of information in a world where structured data is ubiquitous.

Who (defines conceptual models?) Analysts? Developers? Ontology specialists? All of the above. But non-traditional users such as scientists, journalists, educators, and almost anyone with data to share are being empowered to define their own information with easy to use data storage and web management systems.

What (is being modeled?) A database as part of an information system or software system? Information that supports a business process? Definitely. But some users define their structured information directly – for display and processing.

Why (is a conceptual model defined?) To describe information and processing of an information system or a software system? To promote collaboration and communication? To increase understanding of a domain? To document a system? Certainly. But let's consider the goals of people who define and publish their own structured information directly; perhaps we can use a conceptual model to offer them useful functionality for their information (e.g., for browsing, mapping, calculations).

We suggest that domain users are doing conceptual modeling. And we believe that they can relate their conceptual model to a domain model when they are enticed by sophisticated information widgets that can select, display, and process their information. We also highlight a problem that has been present since conceptual models (or database schemas) were first created: information of interest to a user might be present in the "data" (such as "Oregon" being part of someone's address) or in the "schema" such as "Oregon" or "California" being attribute names (for a sport fishing registry). Finally, we show that users (who understand their own information) can perform schema integration, including complex operations such as pivot and unpivot, when guided with examples (of the widgets) using sample data.

Conceptual Modeling: Philosophical Considerations

Josef Mitterer

Alpen-Adria-Universität Klagenfurt
josef.mitterer@aau.at

Abstract. The underlying philosophies of Conceptual Modeling vary between Critical Realism and Ontological Constructivism and fit into the philosophical panorama: *There are distinctions and therefore we make them* (Realism) — *We make distinctions and therefore they are* (Idealism/Constructivism).

The presupposition of dichotomies between language and world, description and object, between what we talk and what we talk about, helps to freeze, dogmatize and fundamentalize the *status quo* into a "real" world and "its" representations.

Claims of representing the real world remain irrelevant as long as consensus prevails. When conflicts arise, the world and other potential decision criteria in a beyond of discourse stay mute: the criteria fail and the opposing parties get into a stalemate… In a recent conversation the ontologist Barry Smith said on how he would deal with competing ontologies: "I try to win."

Proposing an alternative philosophy of change requires a shift in the vocabulary and in the direction of discourse: Instead of advocating a dichotomy between a fixed/independent world and privileged representations, a philosophy of change favors relations between *so far* and *from now on*. The object of a description relates to the description of the object like the description *so far* to the description *from now on*. Every description of the object changes the object into a new object of further descriptions.

Philosophical ontologists try to transcend the "here and now" into the past and future. I opt for transparence rather than transcendence. The world, the reality is nothing but the present state of things.

IT Professionals and Conceptual Modeling

Francisco Garcia Moran

European Commission, BECH C3/631, Luxembourg
francisco.garcia-moran@ec.europa.eu

Abstract. IT professionals, explicitly or implicitly, develop conceptual models when trying to produce a high level description of the fundamental principles and the main functionalities of the “systems” (understood in the most general way: Enterprise Architecture, Infrastructure Blueprints, Information Systems, Database Systems, etc.) they want to implement. They do it because they want

1. enhance the understanding of the “users”,
2. facilitate the dialogue among system’s stakeholders,
3. provide system designers with an input to produce system specifications at different levels, and
4. document the system for future reference and collaboration activities.

There are several relevant questions to IT practitioners about the use of conceptual modelling that the author will try to cover in his presentation on his more than 40 years of professional experience in the public sector as well as his conversations with hundreds of IT professionals in the public and private sectors:

1. Why “conceptual modeling” is considered by many IT professionals as “too theoretical” or “too heavy”?
2. Which are the barriers and facilitators for its more formal adoption?
3. Is there a contradiction between “being agile” (for instance using agile development methodologies like Scrum) and the formal use of conceptual models?
4. What can be done about it?

The author will try to illustrate the answer to some of the above mentioned questions based on the results on an informal survey filled in by many of his contacts in public and private sectors.

Classification and Science

Yair Wand

Sauder School of Business, The University of British Columbia,
Vancouver, BC, Canada
yair.wand@ubc.ca

Abstract. Classifying phenomena is deeply intertwined with cognition and human information processing. Therefore, identifying classes is a central aspect of information technology (IT). Choosing a "good" set of classes is both theoretically and practically important. Two cognitive principles underlie the cognitive approach to classification. First, classes encapsulate inferences about the properties of their instances – in other words, knowing a category can "tell" us more about an instance that required to identify the category it belongs to. Second, collections of classes should provide economy of storage. This leads to a view of classes as carriers of domain knowledge in the form of inferences about situations, which is more than "containers" for information.

We discuss how this view can be used to model scientific theories. We explain how the principles can be used to guide the choice of collections of classes. We show how the approach can be used in scientific discourse by applying it to one of the most well-known areas of physics – the electromagnetic equations as developed originally by Maxwell. The example shows how the classification based approach can be generally applied to scientific problems and that it has two advantages. First, it can provide a simpler and more informative account of the sample phenomena. Second, the classification principles can lead to questions to be asked to help resolve differences between observations and predictions. This means that the resolution of problems can be framed in terms of changes to classification structures, and to principles suggesting how such changes might occur.

Contents

Conceptual Modeling and Requirements

Foundations

Mining Goal Refinement Patterns: Distilling Know-How from Data

Metta Santiputri[1], Novarun Deb[2], Muhammad Asjad Khan[3], Aditya Ghose[3(✉)], Hoa Dam[3], and Nabendu Chaki[2]

[1] Department of Informatics, State Polytechnic of Batam, Batam 29461, Indonesia
metta@polibatam.ac.id

[2] Department of Computer Science and Engineering, University of Calcutta, Kolkata, India
novarun@acm.org, nabendu@ieee.org

[3] Decision Systems Lab, School of Computing and IT, University of Wollongong, Wollongong, NSW 2522, Australia
{aditya,hoa}@uow.edu.au

Abstract. Goal models play an important role by providing a hierarchic representation of stakeholder intent, and by providing a representation of lower-level subgoals that must be achieved to enable the achievement of higher-level goals. A goal model can be viewed as a composition of a number of *goal refinement patterns* that relate parent goals to subgoals. In this paper, we offer a means for mining these patterns from enterprise event logs and a technique to leverage vector representations of words and phrases to compose these patterns to obtain complete goal models. The resulting machinery can be quiote powerful in its ability to mine *know-how* or *constitutive norms*. We offer an empirical evaluation using both real-life and synthetic datasets.

Keywords: Goal model mining · Goal refinement · Know-how

1 Introduction

Goal models play a critical role in requirements engineering, by providing a hierarchic representation of statements of stakeholder intent, with goals higher in the hierarchy (parent goals) related to goals lower in the hierarchy (sub-goals) via AND- or OR-refinement links. Goal models encode important knowledge about feasible, available alternatives for realizing stakeholder intent represented at varying levels of abstraction. A number of prominent frameworks leverage goal models, including KAOS [8], i* [25] and Tropos [4].

There is a growing realization that data analytics (this term being liberally interpreted to denote a broad repertoire of machine learning, data mining and natural language processing techniques) have an important role to play in software engineering in general, and requirements engineering in particular. In that spirit, this paper addresses the question: *can enterprise goal models be mined*

H.C. Mayr et al. (Eds.): ER 2017, LNCS 10650, pp. 69–76, 2017.
https://doi.org/10.1007/978-3-319-69904-2_6

from readily available enterprise data? It is useful to distinguish, at this point, the exercise of mining *goal models* from the exercise of mining *goals*. That latter problem is arguably more difficult, since user goals or stakeholder intent are often never manifested in enterprise data, and are often not explicitly articulated either. Knowledge about how a goal might be refined into lower-level sub-goals is a different matter altogether. Goal refinements that have been deployed before (either explicitly or implicitly) are ultimately manifested in operational data. Our intent in this paper is to leverage data of this form.

Mining goal models adds value in a number of ways. *First*, it offers a way around the *model acquisition bottleneck* (where the high investments associated with careful modeling often prevents businesses from leveraging the full value of goal modeling). While our approach does not guarantee that all models mined will be correct and accurate, it does ensure that the goal models (or model fragments) that are mined can be quickly deployed with minimal editing (the requirement for oversight and editing by analysts remains). Overall, the approach improves the productivity of modelers/analysts; instead of starting with a "blank sheet", our machinery generates "first draft" models or model fragments that can be composed to obtain usable models. *Second*, our approach could potentially improve model quality, by mining execution histories from which "undesirable" executions have been filtered out. *Third*, model *anti-patterns* can be mined from "undesirable" execution data. *Fourth*, this machinery can be used for *goal conformance* checking.

Goal models can also be viewed as statements of *know-how*, where an AND-decomposition provides the know-how for achieving a parent goal by satisfying a set of sub-goals. Mining know-how patterns is independently useful. In particular, it permits us to use goal models as *effectors*, where a goal model is used to specify the desired state of the enterprise while decomposition via a sequence of know-how patterns enables us to identify the operational interventions which would help realize the desired state of the enterprise.

AND-refinement patterns can also be viewed as *constitutive norms* [3]. A constitutive norms specifies how the act of achieving conditions c_1, c_2, c_n *counts as* achieving condition c (we can also, without loss of generality, replace conditions with goals or actions). For instance, the acts of putting a tea bag in a cup followed by puring hot water into the cup *counts as* making tea. The account we offer in this paper can thus be also viewed as an account of constitutive norm mining.

We address two problems in this paper. First, we address the *goal refinement pattern mining problem*, where a goal refinement pattern is of the form $sg_1, sg_2, \ldots sg_n \rightarrow G$ where G is the parent goal while each sg_i is a sub-goal, and where the statement is that the act of achieving each sub-goal conjointly leads to the achievement of the parent goal. These latter are referred to as AND-refinement patterns, and are the main focus of this paper (OR-refinement patterns can be mined via small variants of the techniques discussed here, but a full discussion is omitted due to space constraints). Second, we address the problem of composing individual goal refinement patterns into *goal trees* (more generally goal graphs) which describe not only how a goal is refined into subgoals, but also how these subgoals can be further refined into sub-subgoals and so on.

We present the general approach in Sect. 2. The identification of goal refinement patterns involves mining event logs (partitioned by levels of abstraction) that leverage *temporal correlation patterns* between goals and subgoals (recall that an *event log* is a collection of time-stamped events). The composition of goal refinement patterns relies on matching subgoal in one refinement pattern with the parent goal of another such pattern - we use word2vec [20] to identify *semantic similarity* between words and phrases that appear in the goals and subgoals for this purpose. We briefly summarize the empirical evaluation contained in the full version of the paper in Sect. 3, and position this proposal in the context of related work in Sect. 4.

2 General Approach

Temporal correlation patterns relating goals and subgoals: A goal and its subgoals are typically related via *temporal correlation patterns* which impose temporal constraints on the achievement of the parent goal relative to the achievement of the subgoals. One such pattern (and the one we will leverage in the empirical evalation in this paper) requires that event denoting the achievement of the parent goal occur immediately or soon, after the events denoting the achievement of the subgoals (the event denoting the making of a cup of tea occurs immediately after the events denoting the placing of a teabag in a cup and the pouring of hot water into the cup). We shall call these *sequential correlations*. Other examples of temporal correlation patterns leverage relations from Allen's Interval Algebra [2]. In some settings, we might require the interval over which each subgoal is achieved be included entirely (using the **during** relationship from the Interval Algebra) in the interval over which the parent goal is achieved. In some settings it might make sense to relate these intervals using the **meets, finishes** or **is equal to** relations from Interval Algebra.

Mining goal refinement patterns from multi-layered event logs: Independent of which temporal correlation pattern applies in a given setting, it is critical that the input event logs are partitioned into layers based on different levels of abstraction. A key assumption underpinning this proposal is that events denoting the achievement of parent goals appear in a log of more abstract events, while events denoting the achievement of subgoals appear in logs of more refined (or lower-level) events. In other words, we assume a hierarchy of levels $L_1, L_2, \ldots$ such that L_i is always at a higher level of abstraction than L_{i+1}. The idea is that goal refinement always occurs between goals manifested by events in adjacent levels in this hierarchy. The key question to address now is: How do we obtain this partitioning/hierarchy? Possible strategies include:

- Leveraging part-whole relationships between objects: We know that a photo, a front page, an embedded chip, a visa or an expiry date are parts of a more

abstract object called a passport. Any event involving the passport photo, or a visa etc. will belong to a lower level in the hierarchy than any event involving the passport.
- Leveraging the source of the data: We know that any event from a process log is likely to be lower in an abstraction hierarchy than any event in a message log. Similarly, events that manifest in the IT infrastructure are typically lower in abstraction than events that involve applications, which in turn are lower level than events concerning business services.
- Leveraging the organizational hierarchy: We know that events associated with roles lower in the organizational hierarchy will likely be lower in the abstraction hierarchy than events associated with roles higher in the organizational hierarchy. The intuition is that employees in a business unit are usually tasked with achieving lower-level goals than the manager of that business unit. Indeed, the goals of the manager rely on the achievement of the subgoals that the employees in that unit are tasked to achieve. The employee-level goals can thus be viewed as AND-refinements of the manager-level goals.

With the abstraction hierarchy of events thus obtained, our task in now to mine (temporal) sequential correlations between events in adjacent levels of the abstraction hierarchy. Thus a passport photo check, a passport validity check, a visa check and a passport stamping event would be followed soon after by a higher-level event indicating that an immigration check has been completed. We would expect to see this pattern repeated frequently. If this frequency meets a user-specified threshold, we conclude that it is indicative of a goal refinement pattern.

Composing goal refinement patterns: The challenge in composing goal refinement patterns to obtain goal models (or goal trees) is the difficulty in relating semantically similar, but syntactically highly distinct, specifications of goals and subgoals. For instance, a subgoal might be represented in natural language as: *log labour hours for billing*. Quite separately, we might find a mined goal refinement pattern for a parent goal represented textually as: *track technician time for charging the customer*. Human intuition suggests that these two goals are semantically quite similar, and any available know-how for the latter would also be useful for the former. Our strategy is to use a state-of-the-art machinery for vector encoding of words and phrases, called *word2vec* [20] which is effective in identifying semantic similarity. Word2vec learns vector representations of words and phrases such that semantically similar ones are projected in close proximity to each other in the vector space. Given a pair of phrases, word2vec returns a real-valued measure of semantic similarity (the higher the value, the more similar the phrases are). By setting an appropriate threshold for the similarity measure (this will require domain-specific tuning), we can connect a phrase describing a subgoal in one goal refinement pattern with a phrase describing a parent goal in another goal refinement pattern.

3 Evaluation

In this section, we briefly summarize the empirical evaluation results presented in the full version of the paper.

Two distinct strategies were evaluated: (1) Sequential pattern mining for leveraging temporal correlations patterns (specifically sequential correlation patterns) between goals andsub-goals and (2) word2vec for evaluating goal-subgoal similarity.

Two distinct datasets were used for the evaluation:

- A synthetic dataset consisting of an event log of a telephone repair process[1]
- A real-life dataset consisting of data from the BPI Challenge 2015 (BPIC'15)[2] which features building permit application process in five Dutch municipalities from year 2010 until 2015.

The BIDE+ algorithm was used for sequential pattern mining.

The evaluation using Google's pre-trained word2vec model was particularly interesting [20]. Word2vec includes word vectors for a vocabulary of 3 million words and phrases that has been trained on approximately 100 billion words from a Google News dataset. Although for this evaluation, we used a pre-trained model, training a model with a smaller but more targeted and domain-specific corpora is not hard. We have done this but have not achieved results thus far that surpass the results we have obtained using the pre-trained model. We took the goal refinement patterns obtained in the evaluation using the phone repair scenario described above (8 in total), and extended these with a repertoire of 40 additional goal refinement patterns (this was necessary to be able to further refine the sub-goals initially obtained from the mining of a 2-level event log.

The Word2Vec metric tends to place two words close to each other if they are semantically similar. We found, for instance, that 'Print repair receipt for the customer' and 'Print customer service repair order' have a high similarity score even though the phrases use different vocabulary to explain the same subgoal. The notion of similarity used here is just cosine distance (dot product of vectors). It is closer to 1 if the phrases are semantically similar. For two completely dissimilar phrases, the similarity is closer to 0. For instance, update issue status to "in repair" and "dissemble the phone components" refer to two very different goals and are very far apart semantically thus receiving a score of 0.130887. In some cases like'log labor hours for billing' and'Track technician time for charging the customer' the score is neither too high nor too low. We can use a certain threshold e.g. (0.60) to filter cases where we not fully confident of a semantic match.

Overall, the results of the empirical evaluation (contained in the full version of the paper and ommitted here due to space constraints) suggest that the combination of techniques proposed here provide a promising basis for goal model mining.

---

1 http://www.processmining.org/_media/tutorial/repairexample.zip.

2 https://www.win.tue.nl/bpi/doku.php?id=2015:challenge.

4 Related Work

A considerable amount of research has been reported applying data mining techniques in requirements engineering. Zawawy et al. have proposed a root-cause analysis framework [26] that mines natively generated log data to establish the relationship between a requirement and the pre- and post-conditions associated with that requirement. In [13], the authors have proposed techniques for mining dependencies from message logs and task-dependency correlations from process logs. There have been very interesting industrial and commercial applications of mining requirements from event logs. Formal verification of control systems have been performed by mining temporal requirements from simulation traces [16]. REQAnalytics [10], proposed by Garcia and Paiva, mines the usage statistics of a website and provides a roadmap for the evolution of the website's requirements specification. ACon [17] is another data mining technique that tries to address the inconsistencies that affect the contextual requirements of a system at runtime.

Sequential pattern mining has been frequently used for extracting statistically relevant patterns or sequences of values in data sets. StrProM [15], for instance, uses the Heuristics Miner algorithm to generate prefix-trees from the data stream and continuously prunes these trees to extract sequences of events. Sohrabi and Ghods use bit-wise compression techniques to represent the data sequence as a 3-dimensional array and extract frequently occurring patterns from this compressed array [23]. Hassani et al. have proposed the PIVOTMiner [14] which considers activities as interval-based events rather than the conventional single-point events. Some researchers have also tried to improve the legacy sequential mining algorithm PrefixSpan (like [5, 21]). Sequential pattern mining has also been used in interesting applications that range from detecting user behavior from online surveys to mining electronic medical records and inferring the efficacy of medicines [24]. A detailed survey of sequential pattern mining algorithms is available in [1].

Previously workflow logs used to be mined for extracting the control flow within an organization and, hence, extensively used for developing process models. Schönig and his group have proposed a framework to extract the organisational structure of business processes by mining human resource allocation information from event logs [22].

Also in prior work, non-functional requirements have been extracted from text [7].

5 Conclusion

The ability to mine goal models has important implications for requirements engineering, as well as a wide variety of other settings that benefit from goal modeling. The machinery that we present can therefore provide useful directions for future research and development. This machinery can also be used to mine know-how which can support enterprise innovation strategies in significant ways. The empirical evaluation presented in the paper is preliminary in nature, but provides evidence that suggests that there is merit in pursuing this general approach. This work can be extended in a number of interesting ways. For instance, evidence of goal update in operational data could be used to reverse engineer goal models from data using intuitions from belief revision [6,11] or belief merging [18,19]. Viewing softgoals as optimization objectives (as has been done in [9]) could provide the basis for correlating goals and softgoals. Techniques for discovering process designs from legacy artefacts [12] could form the basis for an alternative approach to mining goal models.

References

1. Abbasghorbani, S., Tavoli, R.: Survey on sequential pattern mining algorithms. In: 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI), pp. 1153–1164 (2015)
2. Allen, J.F.: Maintaining knowledge about temporal intervals. Commun. ACM **26**(11), 832–843 (1983)
3. Boella, G., Broersen, J., van der Torre, L.: Reasoning about constitutive norms, counts-as conditionals, institutions, deadlines and violations. In: Bui, T.D., Ho, T.V., Ha, Q.T. (eds.) PRIMA 2008. LNCS, vol. 5357, pp. 86–97. Springer, Heidelberg (2008). doi:10.1007/978-3-540-89674-6_12
4. Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F., Mylopoulos, J.: Tropos: an agent-oriented software development methodology. Auton. Agent. Multi-Agent Syst. **8**(3), 203–236 (2004)
5. Chaudhari, M., Mehta, C.: Extension of prefix span approach with GRC constraints for sequential pattern mining. In: International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 2496–2498 (2016)
6. Chopra, S., Ghose, A., Meyer, T.: Non-prioritized ranked belief change. J. Philos. Logic **32**(4), 417–443 (2003)
7. Cleland-Huang, J., Settimi, R., Zou, X., Solc, P.: The detection and classification of non-functional requirements with application to early aspects. In: 14th IEEE International Conference on Requirements Engineering, pp. 39–48. IEEE (2006)
8. Darimont, R., Delor, E., Massonet, P., van Lamsweerde, A.: GRAIL/KAOS: an environment for goal-driven requirements engineering. In: Proceedings of the 19th International Conference on Software Engineering, pp. 612–613. ACM (1997)
9. Dasgupta, A., Ghose, A.K.: Implementing reactive BDI agents with user-given constraints and objectives. Int. J. Agent-Oriented Softw. Eng. **4**(2), 141–154 (2010)
10. Garcia, J.E., Paiva, A.C.: Maintaining requirements using web usage data. In: International Conference on ENTERprise Information Systems/International Conference on Project MANagement/International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN /HCist. Procedia Comput. Sci. **100**, 626–633 (2016)
11. Ghose, A., Goebel, R.: Belief states as default theories: studies in non-prioritized belief change. In: ECAI, vol. 98, pp. 8–12 (1998)
12. Ghose, A., Koliadis, G., Chueng, A.: Rapid business process discovery (R-BPD). In: Parent, C., Schewe, K.-D., Storey, V.C., Thalheim, B. (eds.) ER 2007. LNCS, vol. 4801, pp. 391–406. Springer, Heidelberg (2007). doi:10.1007/978-3-540-75563-0_27

13. Ghose, A., Santiputri, M., Saraswati, A., Dam, H.K.: Data-driven requirements modeling: Some initial results with i*. In: Tenth Asia-Pacific Conference on Conceptual Modelling (APCCM), pp. 55–64 (2014)
14. Hassani, M., Lu, Y., Wischnewsky, J., Seidl, T.: A geometric approach for mining sequential patterns in interval-based data streams. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 2128–2135 (2016)
15. Hassani, M., Siccha, S., Richter, F., Seidl, T.: Efficient process discovery from event streams using sequential pattern mining. In: IEEE Symposium Series on Computational Intelligence, pp. 1366–1373 (2015)
16. Jin, X., Donze, A., Deshmukh, J.V., Seshia, S.A.: Mining requirements from closed-loop control models. IEEE Trans. Comput. Aided Des. Integr. Circ. Syst. **34**(11), 1704–1717 (2015)
17. Knauss, A., Damian, D., Franch, X., Rook, A., Müller, H.A., Thomo, A.: ACon: a learning-based approach to deal with uncertainty in contextual requirements at run time. Inf. Softw. Technol. **70**, 85–99 (2016). Elsevier
18. Meyer, T., Ghose, A., Chopra, S.: Social choice, merging, and elections. In: Benferhat, S., Besnard, P. (eds.) ECSQARU 2001. LNCS, vol. 2143, pp. 466–477. Springer, Heidelberg (2001). doi:10.1007/3-540-44652-4_41
19. Meyer, T., Ghose, A., Chopra, S.: Syntactic representations of semantic merging operations. In: Ishizuka, M., Sattar, A. (eds.) PRICAI 2002. LNCS (LNAI), vol. 2417, p. 620. Springer, Heidelberg (2002). doi:10.1007/3-540-45683-X_88
20. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
21. Patel, R., Chaudhari, T.: A review on sequential pattern mining using pattern growth approach. In: International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 1424–1427 (2016)
22. Schönig, S., Cabanillas, C., Jablonski, S., Mendling, J.: Mining the organisational perspective in agile business processes. In: Gaaloul, K., Schmidt, R., Nurcan, S., Guerreiro, S., Ma, Q. (eds.) CAISE 2015. LNBIP, vol. 214, pp. 37–52. Springer, Cham (2015). doi:10.1007/978-3-319-19237-6_3
23. Sohrabi, M.K., Ghods, V.: CUSE: a novel cube-based approach for sequential pattern mining. In: 4th International Symposium on Computational and Business Intelligence (ISCBI), pp. 186–190 (2016)
24. Uragaki, K., Hosaka, T., Arahori, Y., Kushima, M.: Sequential pattern mining on electronic medical records with handling time intervals and the efficacy of medicines. In: IEEE Symposium on Computers and Communication (ISCC), pp. 20–25 (2016)
25. Yu, E.S.: Towards modelling and reasoning support for early-phase requirements engineering. In: Proceedings of the Third IEEE International Symposium on Requirements Engineering, pp. 226–235. IEEE (1997)
26. Zawawy, H., Mankovskii, S., Kontogiannis, K., Mylopoulos, J.: Mining software logs for goal-driven root cause analysis. In: The Art and Science of Analyzing Software Data, pp. 519–554 (2015)