HAL
open science

Space-Time Cube Operations in Process Mining

Dina Bayomie, Lukas Pfahlsberger, Kate Revoredo, Jan Mendling

To cite this version:

Dina Bayomie, Lukas Pfahlsberger, Kate Revoredo, Jan Mendling. Space-Time Cube Operations in Process Mining. 13th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modeling (PoEM 2020), Nov 2020, Riga, Latvia. pp.405-414, 10.1007/978-3-030-63479-7_28 . hal-03434663

HAL Id: hal-03434663

https://inria.hal.science/hal-03434663v1

Submitted on 18 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Space-Time Cube Operations in Process Mining

Dina Bayomie[0000−0002−2549−6407], Lukas Pfahlsberger[0000−0002−1367−9441], Kate Revoredo[0000−0001−8914−9132], Jan Mendling[0000−0002−7260−524X]

Institute for Information Business,
Vienna University of Economics and Business (WU), Vienna, Austria
{dina.sayed.bayomie.sobh}@wu.ac.at {lukas.pfahlsberger}@wu.ac.at
{kate.revoredo}@wu.ac.at {jan.mendling}@wu.ac.at

Abstract. Process mining techniques provide data-driven visualizations that help gaining multi-perspective insights into business processes. These techniques build on a variety of algorithms, however without any explicit reference to the spectrum of potential analysis of operations. For this reason, it is unclear if the state of the art of process mining has missed opportunities to develop techniques that could be of potential value to an analyst. In this paper, we refer to research on information visualization where this problem has been addressed from a more general angle. More specifically, we use the framework defined for space-time cube operations to explore to which extent process mining instantiates these operations. To this end, we refer to most widely used commercial process mining tools and analyze their analysis operations. We find that the majority of the operations are already supported by the tools, but there are still unsupported ones, which exhibit opportunities for future research and tool innovation.

Keywords: Process mining · space-time cubes · information visualization

1 Introduction

Organizations increasingly use process mining techniques for analyzing their processes based on process execution data (i.e. event-log) [1,2]. These techniques can be used to discover the process model, check conformance between the data and the process model, explore variants, analyze bottlenecks or predict process performance indicators. The broad spectrum of techniques results from the different inputs (event-log, process model, etc.) and the different outputs that are generated (various types of metrics, charts, etc.). For instance, process discovery techniques use event-logs only to construct a process model that best fits the data, while conformance checking techniques use event-logs and a process model to project the results onto the model. All process mining techniques have in common that they generate visualizations from process data. These visualizations are beneficial to process analysts for process optimization, monitoring, and decision making.

One of the current challenges of process mining is the richness and diversity of techniques. The 6.9 distribution of ProM from May 2020 alone contains 269 packages and most of them provide separate analysis functionality. On the one hand, it is unclear which techniques offer comparable insights. On the other hand, it is unclear if

opportunities for analysis techniques have been missed. What is absent at this stage is a framework that helps organizing the spectrum of techniques and inspiring the development of new techniques.

In this paper, we address this research gap by referring to research on information visualization. More specifically, we build on the space-time cube proposed in [3] as a framework for identifying abstraction operations that produce two-dimensional visualizations from multi-dimensional data. Using this framework, we analyze to which extent commercial process mining tools instantiate those operations. We find that the majority of the operations are already supported by the tools, but there are still unsupported ones, which exhibit opportunities for future research and tool innovation.

The remainder of this paper is organized as follows. Section 2 describes the background of process mining techniques and space-time cube operations. Section 3 presents our findings of analyzing the spectrum of analysis techniques offered by two commercial process mining tools. Finally, Section 4 concludes our work and provides directions for future research.

2 Background

This section discusses the background of process mining and gives an overview of the space-time cube and its operations.

2.1 Process mining and process analytics

Process mining techniques enable companies to discover, monitor, and improve real business processes by extracting information from event-logs of process-aware systems (compare Figure 1). In this way, it becomes possible to visualize and analyze control flows, bottlenecks, resource utilization, social networks, and organizational hierarchies in direct association with the process execution [1] and various other kinds of process analytics using case-spanning aggregations. Typically, dedicated process mining tools like Celonis and Disco offer specialized analysis and visualizations of event-logs produced by other systems. Each tool provides a partially overlapping and partially individual set of analyzing functionalities with varying characteristics and naming conventions which makes it challenging for analysts to make a comparison.

Especially the *time* dimension plays an important role during the analysis since processes are usually time-related and therefore subject to constant temporal changes. The underlying phenomenon is defined by the term *process drift* [4,5]. A process drift can occur in accordance to seasonal or recurring patterns like “office hours, lunch breaks, and weekends” [1] as well as factors attributable to the economic, political, cultural or nature-related environment. A drift implies that the results of a process analysis and the corresponding visualization can change significantly over time. For this reason, any systematic visualization approach has to explicitly capture this temporal dimension and provide operations that help exploring and investigating this specific dimension.

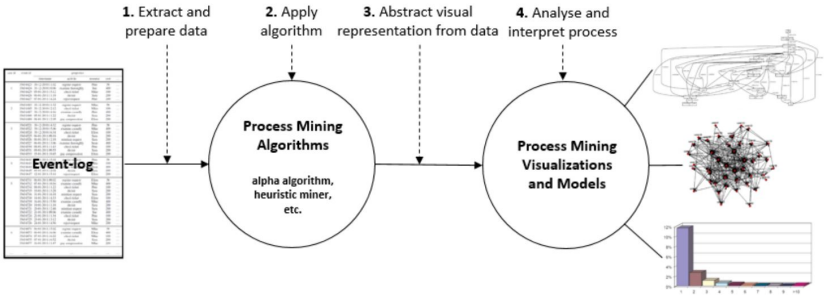


Fig. 1: General steps in process mining from an event-log to a visualization

2.2 The space-time cube

The visualization of temporal information is discussed from a general angle in research on information visualization. In this paper, we build upon the descriptive framework for temporal data visualization of generalized space-time cubes by Bach et al. [6]. The framework provides a comprehensive view of the alternatives to visualize temporal datasets. A generalized space-time cube is defined as a three-dimensional Euclidean space that consists of a time dimension as well as a two-dimensional spatial space. In the context of processes, the spatial space can be interpreted as additional process-related data like process activities, organizational units, associated costs, persons, etc.. The idea of this framework is to offer a temporal data visualization by abstracting it in multiple ways so that it is representable on a screen or on paper. Mind that it is not explaining the underlying data structures and algorithmic transformation procedures [6,3].

Space-time cubes contain space-time objects. These objects instantiate the dimensions of the space-time cube. They can be represented as "(i) space-time volumes (of which a complete space-time cube is an example), (ii) space-time surfaces (planar and non-planar), (iii) space-time curves, (iv) points as well as (v) sets of disconnected volumes, surfaces, curves, and points" [6].

Different two-dimensional visualizations can be calculated from space-time objects. Corresponding operations transform an existing space-time object of a certain type into a different one. This transformation process can repeat itself multiple times and *completes* once the space-time object provides the desired visualization.

2.3 Space-time cube operations

Next, we describe selected space-time cube operations that fit the purpose of transforming space-time objects into 2D visualizations. In general, there are four space-time cube operation categories *extraction* (selection of a subset), *flattening* (aggregating into a lower dimension), *content transformation* (content change without adapting its geometry), and *geometry transformation* (rearranging the data spatially without changing its content) [6].

The most common operations from the *extraction* category are *time cutting* and *space cutting*. This operations perform an orthogonal planar cut to the time axis for time extraction and a parallel planar cut for space extraction (Figure 2). Further detailing of the data can be achieved by applying *time and space drilling*. These drillings not only cut a planar, but extracts linear (orthogonal or parallel) data from the space-time cube. In contrast, the *time and space chopping* operations not only cut a tight snapshot out of the cube but a certain temporal or spacial range. For instance, time chopping would be expressed as data restricted to e.g. the year 2020, whereas space chopping would consider the entire time period but just e.g. the information regarding a certain country [6].

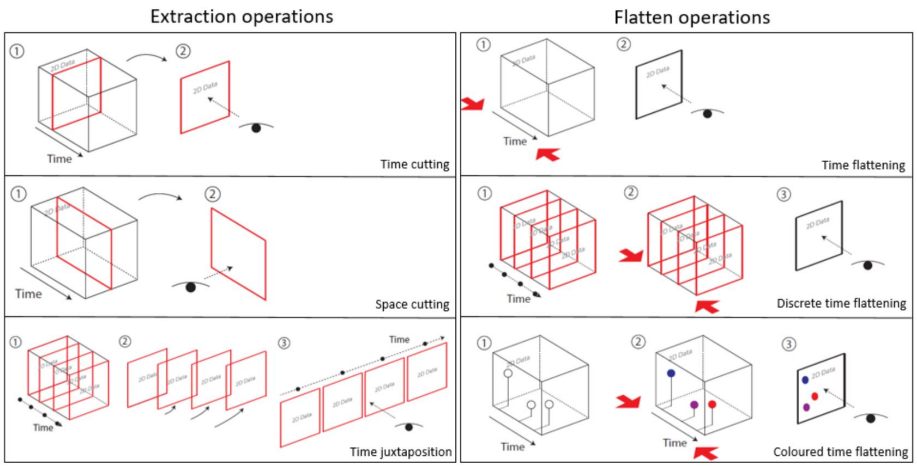


Fig. 2: Selected extraction (left) and flatten (right) operations inside the space-time cube [6]

Every operation that aggregates data into a lower dimension belongs to the *flattening* category. When applying the *time or space flattening* operation, all data along the time or space axis are coalesced into a single 2D illustration. To exemplify, a time-lapse picture would be a flattened 2D representation of multiple pictures along the time axis. Figure 2 (top right) illustrates this operation. As an alternative, *discrete time flattening* uses multiple slices and flattens them afterwards (Figure 2 (middle right)). By assigning time slices to specific colors *colored time flattening* offers a third possibility (Figure 2 (bottom right)).

The category *content transformation* contains multiple operations to present data in different ways without changing its geometry. Figure 2 already showed one type, namely *colored time flattening*. This operation can be complement with operations like labelling, shading, stabilizing, bundling, filtering, or aggregation [6].

Further, *space and time shifting* from the category *geometry transformation* rotates, scales, bends, unfolds or translates space-time objects inside the Euclidean space without changing its content. Thereby, it gets possible to place multiple slices of data next

to each other in order to provide a better comparison detached from certain spacial or temporal attributes [6].

Space-time cube operations usually occur as a combination of several individual operations, called *compound operations*. For instance, the previously introduced *discrete time flattening* (Figure 2) is a combination out of multiple *time cutting* operations and a subsequent *time flattening* operation. Which operations to use is based on the underlying visualization requirement of the user and can have any desired order. The outcome of a compound operation can have a static or a dynamic character. Whereas a static outcome is represented e.g., as a graphic on paper media, a dynamic outcome usually occurs on computer screens with adjustable sliders or buttons to change the visualization dynamically [6].

2.4 Related work

Now, we want to emphasize related topics in both *information systems* and *information visualization* to embed them into the broader context. Process cubes provide a way to abstract the description of event-log data by decomposing and visualizing it in different dimensions [4]. Based on the axes of the cube, the dimensions can be rolled up, split up, sliced, or diced. These operations are tightly related to OLAP-cubes which are commonly used for reporting on process-aware systems. Still, traditional OLAP-cubes are unsuitable for process-centric analysis since events have to be temporally ordered and associated with specific process cases [7]. In comparison to space-time cubes, both, process- and OLAP-cubes, can be constituted by an infinite number of dimensions. This can be seen as the key difference together with the fact that space-time cubes focus on the transformation operations of visualizations on a conceptual level, while process- and OLAP-cubes are data structures that connect to a technical implementation. In turn, the space-time cube is related to other taxonomies, which it extends, including value flow maps [8], planning polygons [9] as well as related taxonomies from Aigner et al. [10] and Andrienko et al. [11].

3 Space-time cube operations in process mining

Process mining techniques, like e.g., process discovery, conformance checking, or process enhancement [1] provide helpful insights into process behavior. However, it stays unclear which underlying data-operations (or algorithmic implementation) certain process mining tools execute on the event log to visualize the behavior. In order to abstract those data-operations from their technical execution, we conduct an explorative experiment and focus on process discovery as well as conformance checking techniques.

Figure 3 illustrates the procedure by using the cube as a descriptive and abstract element to explain the data-operations performed. The data of the event log can be seen as the input for building up the space-time cube whereas the concrete visualization, after applying space-time cube operations, represents the output.

To be more precise, the event-log data (left side of Figure 3) determines how to set up the cube and subsequently, which operations can be performed. Essential is the existence of timestamps within the event log since they are assigned to the time axis of

the cube. A space dimension can be represented by any of the associated event attributes from the event log, e.g., activity, resources, etc. Consequently, each event represents a dedicated cell inside the cube which is explicitly assigned to both time and space axis.

The output shown on the right side of Figure 3 illustrates possible process visualizations, e.g. a process model as a visualization output for process discovery techniques.

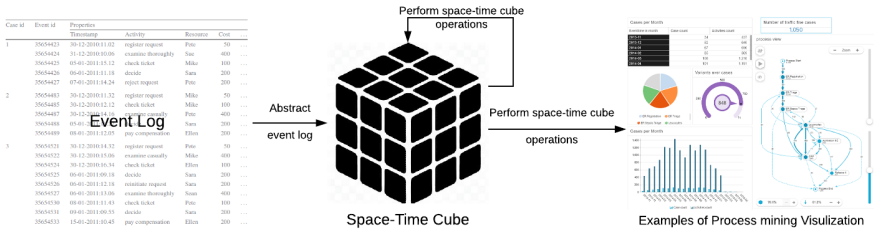


Fig. 3: General overview of space-time cube operations on an event-log

3.1 Exploratory setting of the experiment

We conduct an exploratory experiment to investigate the space-time cube operations utilized in process mining by commercial process mining tools. By doing that, we used the tools Celonis[1] and Disco[2] since they are the most commonly used tools in the industry and also provide academic licenses. Nevertheless, the goal of this study is not associated with the comparison of specific tools. Rather, we want to provide inductive evidence that the space-time cube operations are generally applicable in process mining. This also implies that the number of supported space-time cube operations in process mining could increase by examining additional tools.

Celonis supports the process analyst with a wide range of process mining techniques to analyze the process behavior itself and reporting capabilities, i.e., charts and aggregated KPIs. We used the Celonis Intelligent Business Cloud - Academic Version. Disco, on the other hand, provides the process analyst with quick insights into the process behavior through its discovered model. It also provides the analyst with a statistical view over all the event data attributes and case explorer view to investigate the cases in detail. We used Disco - academic license - version 1.9.9. Both tools offer filter and selection functionalities to perform profound investigation analysis over logs.

In order to prove the applicability of the introduced space-time cube operations in process mining, we utilized a publicly available event-log of the Italian state road traffic fine management process[3] with 150.370 cases and 231 process variants. We chose

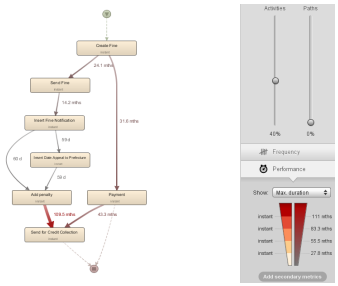
[1] https://www.celonis.com/
[2] https://fluxicon.com/disco/
[3] https://data.4tu.nl/repository/uuid:270fd440-1057-4fb9-89a9-b699b47990f5

this event-log because it has a suitable size and variety of process variants to generate appropriate visualizations. However, the data content itself is not relevant as it does not affect the results of this study.

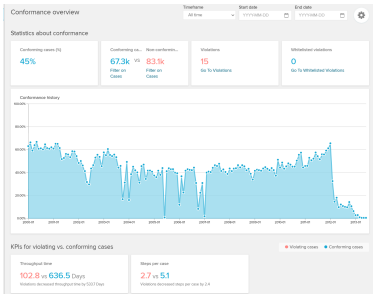
3.2 Preliminary results

In the following paragraph we share the results of this experiment by demonstrating four selected examples of the process mining techniques process discovery (a), conformance checking (b), variant analysis (c) and performance analysis (d).

The **process discovery (a)** technique implemented in Disco applies the space-time cube operations *space cutting* and *time flattening*. It first performs *space cutting* over the activities and then over the paths. Finally, it executes *time flattening* to aggregate the activities' relationships in one model over the log's whole time window. We used Disco for this analysis and the output visualization is depicted in Figure 4a.

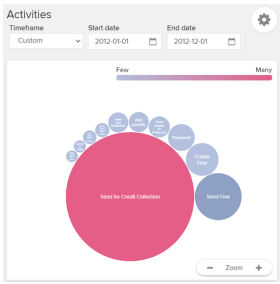


(a) Process discovery using Disco



(b) Conformance checking using Celonis

(c) Variant analysis using Celonis



(d) Activity explorer using Celonis

Fig. 4: Results from the exploratory experiment with Celonis and Disco

For **conformance checking (b)**, the space-time cube operations which can be derived are *oblique chopping*, *space shifting*, and *time flattening*. The goal of conformance checking is to detect the deviations and to measure log fitness. The technique performs *oblique chopping* over the whole event log to extract the conforming and non-conforming cases, then *space shifting* to use both the cases in computing the fitness.

Finally, *time flattening* to generate the time series chart. We used Celonis for this analysis, and the output visualizations are depicted in Figure 4b.

To detect and **analyze the variants (c)** within the log, the technique performs *oblique cutting*, *space shifting*, and *space flattening*. It executes *oblique cutting* over the activity sequences, then *space shifting* to show the difference between the behavior and interactive space flattening. Figure 4c depicts the results obtained using Celonis.

For **performance analysis (d)**, we explore the traffic fine process activities' density over a specific year using Celonis. First, the *time cutting* operation over a specific year (e.g., get only the cases that occurred in 2012) was applied. Then, it performs *space flattening* over the activities, and finally, *space coloring* to indicate the density of the activity within the selected cases. Figure 4d depicts the resulting graphic for the year 2012.

	Space-time operations		**PM**
Extraction	Point Extraction		✓
	Drilling	Time	✓
		Space	✓
		Oblique	✓
		Planar curvilinear	-
		Non- Planar	-
	Cutting	Time	✓
		Linear space	✓
		Oblique	✓
		Curvilinear space	-
		Non-planar	-
	Chopping	Time	✓
		Linear space	✓
		Oblique	✓
		Curvilinear space	-
		Non-planar	-
Flattening	Planar flattening	Time	✓
		Space	✓
		Oblique	-
	Non planar flattening		-

	Space-time operations		**PM**
Geometry	Transformation Shifting	Time	✓
		Space	✓
	Rotation	Yaw	-
		Pitch and Roll	-
	Scaling	Time	-
		Space	-
	Bending		-
	Unfolding		-
Content Transformation	Filing Interpolation	Time	-
		Space	-
		Volume	-
	Coloring	Time	✓
		Space	✓
		Other	✓
	Time Labeling		✓
	Repositioning	Stabilizing	-
		Bundling	-
	Shading		-
	Filtering		✓
	Aggregation		✓

Table 1: The support of the space-time cube operation in process mining by analyzing Disco and Celonis

The result of the confrontation of the process mining techniques using Celonis and Disco with space-time cube operations is summarized in Table 1. The table is divided by the four space-time cube operation categories (see section 2.3). Then in each category, the available space-time cube operations are listed and we indicate if current process

mining techniques support them (✓) or not (−). A checkmark in column *PM* of Table 1 can be interpreted as a proof that the associated space-time cube operation is generally applicable in process mining. In contrast to this, a minus symbol means that we could not proof its applicability in our experiment. This should not imply that it is generally not possible.

3.3 Discussion

Table 1 shows that some space-time cube operations satisfy the process mining techniques more than others. Process mining techniques support most of the operations in *extraction* and *flattening* categories. Nevertheless, all space-time cube operations from those categories with curvilinear character (planar or linear) could not be appropriately assigned. Moreover, oblique extraction operations represent a challenge. The oblique flattening is not supported because oblique structures always imply a dynamic perspective change parallel to the time and space dimensions, which is currently not implemented in Celonis or Disco. Further, the space-time cube category *geometry* offered the least conformity since we could not find an underlying use case for bending, unfolding, scaling, or rotation in process mining. The category *content transformation* can be seen as an enhancement to the remaining three since it offers additional refinement possibilities for visualizations. Tools like Celonis and Disco already offer many visualization refinements to present results in a different light. However, this is limited to the most common ones like coloring, labeling, filtering, and aggregating due to complexity.

Process mining techniques perform component operations, in which different space-time basic operations are applied in sequence, to generate its output. The extraction and flattening operations are the essential operations that are applied in the performed component operations. Additionally, Celions and Disco provide dynamic-interactive operations on the visualization through the usages of the filterings.

4 Conclusion

In this paper, we presented a conceptual approach that explains how temporal event-log data can be visualized in process mining. By doing that, we leveraged the conceptual taxonomy of Bach et al. [6], which describes certain space-time cube operations and transferred their underlying assumptions to the visualization in process mining. This provides a new perspective on how to demonstrate the visualization possibilities of process mining but also on how to standardize its underlying operations for a better cross-domain understanding. We conducted an exploratory study using Celonis and Disco applied to a real event log.

The preliminary results we found with our study indicate some directions for future work. The main one concerns exploring the operations not yet addressed by process mining techniques. Especially the content transformation operations, which mainly focus on different ways of visualizing the findings, and therefore process mining could benefit from them. Also, investigate more if the different geometry operations are applicable. Another direction for investigation relies on the observation that process mining techniques usually use compound operations. The study should investigate existing

compound operations and the impact of the order in which the operations are executed. This study focused on the two commercial tools Celonis and Disco which can also be seen as the major limitation of this experiment. Moreover, the linkage between space-time cube operations and their actual algorithmic implementation is missing. Further investigation can be done using scientifically oriented tools like ProM[4] but also different commercial tools positioned in dedicated niches to uncover novel visualization approaches.

References

1. W. van der Aalst, *Process Mining - Data Science in Action*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2 ed., 2016.
2. M. Dumas, M. L. Rosa, J. Mendling, and H. A. Reijers, *Fundamentals of Business Process Management, Second Edition*. Springer, 2018.
3. B. Bach, E. Pietriga, and J.-D. Fekete, "Visualizing dynamic networks with matrix cubes," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, (New York, NY, USA), p. 877–886, Association for Computing Machinery, 2014.
4. W. M. P. van der Aalst, "Process cubes: Slicing, dicing, rolling up and drilling down event data for process mining," in *Asia Pacific Business Process Management* (M. Song, M. T. Wynn, and J. Liu, eds.), (Cham), pp. 1–22, Springer International Publishing, 2013.
5. A. Yeshchenko, C. Di Ciccio, J. Mendling, and A. Polyvyanyy, "Comprehensive process drift detection with visual analytics," in *Conceptual Modeling* (A. H. F. Laender, B. Pernici, E.-P. Lim, and J. P. M. de Oliveira, eds.), (Cham), pp. 119–135, Springer International Publishing, 2019.
6. B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale, "A descriptive framework for temporal data visualizations based on generalized space-time cubes," *Computer Graphics Forum*, vol. 36.
7. A. Bolt, M. de Leoni, W. M. P. van der Aalst, and P. Gorissen, "Business process reporting using process mining, analytic workflows and process cubes: A case study in education," in *Data-Driven Process Discovery and Analysis* (P. Ceravolo and S. Rinderle-Ma, eds.), (Cham), pp. 28–53, Springer International Publishing, 2017.
8. N. Andrienko and G. Andrienko, "Interactive visual tools to explore spatio-temporal variation," in *Proceedings of the working conference on Advanced visual interfaces*, pp. 417–420, 2004.
9. Poonam Shanbhag, P. Rheingans, and M. desJardins, "Temporal visualization of planning polygons for efficient partitioning of geo-spatial data," in *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 211–218, 2005.
10. W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski, "Visualizing time-oriented data—a systematic view," *Computers & Graphics*, vol. 31, no. 3, pp. 401–409, 2007.
11. G. Andrienko, N. Andrienko, and M. Heurich, "An event-based conceptual model for context-aware movement analysis," *International Journal of Geographical Information Science*, vol. 25, no. 9, pp. 1347–1370, 2011.

[4] `http://www.promtools.org/doku.php`