Process Support and Visual Adaptation to Assist Visual Trend Analytics in Managing Transportation Innovations

Dirk Burkhardt[1(✉)], Kawa Nazemi[1], and Egils Ginters[2]

[1] Human-Computer Interaction and Visual Analytics Group, Darmstadt University of Applied Sciences, Darmstadt, Germany
{dirk.burkhardt,kawa.nazemi}@h-da.de
[2] Riga Technical University, Kalku Str. 1, Riga 1048, Latvia
egils.ginters@rtu.lv

Abstract. In the domain of mobility and logistics, a variety of new technologies and business ideas are arising. Beside technologies that aim on ecologically and economic transportation, such as electric engines, there are also fundamental different approaches like central packaging stations or deliveries via drones. Yet, there is a growing need for analytical systems that enable identifying new technologies, innovations, business models etc. and give also the opportunity to rate those in perspective of business relevance. Commonly adaptive systems investigate only the users' behavior, while a process-related supports could assist to solve an analytical task more efficient and effective. In this article an approach that enables non-experts to perform visual trend analysis through an advanced process support based on process mining is described. This allow us to calculate a process model based on events, which is the baseline for process support feature calculation. These features and the process model enable to assist non-expert users in complex analytical tasks.

Keywords: Adaptive visualization · Transportation and logistics · Process mining

1 Introduction

The digitalization is a challenging task for almost every economic area. New upcoming technologies coupled with novel business ideas lead to innovative and revolutionizing business solutions. Many of them have the potential to change markets fundamentally and imperil primarily smaller players like small and medium sized enterprises (SME). The transportation and logistics domain is one of these fields, where a variety of new technologies and business ideas arise. However, in the current state it is rather difficult to estimate which of these innovations will be important and successful [10]. Beside technologies that aim on ecologic and economic transportation, such as alternative engines e.g. for parcel trucks, there are also fundamental different approaches like central packaging station or deliveries via drones. Yet there is a growing need for tools that enable identifying new technologies, ideas and furthermore, give also the opportunity to rate those in perspective of (future) relevance. The analytical challenges that

E. Ginters et al. (Eds.): ICTE ToL 2019, LNITI, pp. 319–327, 2020.
https://doi.org/10.1007/978-3-030-39688-6_40

have to be faced particularly are: (1) What technologies, business ideas, strategies etc. are coming up? (2) How have these upcoming technologies, business ideas, strategies etc. to be rated in perspective of future relevance? And (3), what of these upcoming technologies, business ideas, strategies etc. are relevant for the own market and more important the own business? These narrowed questions are substantial to answer as early as possible to be able to react on market changes.

Indeed, solutions for trend analysis, especially on patent data are still standard and often used. However, these tools use patent data only, which limits the gathered insights significantly on those that can be IPR protected. New business ideas or technological procedures cannot be registered and can therewith not identified with this data fundament. Even more, the patent registration procedures take approx. two years, which results in identifying trends that are at least two years old as well. Next to the data side, most of these tools are expert tools and require knowledge about the domain, technical understanding as well es economic experience to be able to perform useful analysis. Such kinds of experts are rare in SMEs and are therewith critical, so that most of these tools are not suitable for SMEs.

Therefore, in this chapter a different solution is proposed. First of all, a modular and scalable processing pipeline is defined that enables visual trend analysis on a variety of data fundaments for innovation management. Standard in that field is patent analysis, but due the long patent registration process and restriction on patents on technologies only, the focus will be on other data sources. Very important are research data from e.g. (open access) digital libraries and web data, such as news from enterprises or market magazines/blogs/portals. With this data and a highly interactive visual analytical solutions for trend analysis [3–5], technology and innovation foresight is possible to identify early market trends. However, to perform sufficient analysis professional knowledge about how to do an analysis would be still required. Therefore, in the chapter is described a novel graphical assistance feature that supports users via process mining [1, 2] in the analysis process. This allows even non-analysts or non-experts to observe the market for relevant innovations and rate them in regards of relevance. This is, as already mentioned, in particular required in the transportation domain, where a high number of small and medium-sized players are active and do not have the budget for intensive analytical investigations. A system that allows an assisted analysis enables these enterprises to be aware of market trends and support strategic decision making. The main contribution is a general model for supporting users based on the events by the mined process and the support features. This model is further implemented in a visual trend analytics system to assist users.

2 Design of a Process-Based User Supporting System

Common system that enable a process support consider a statically defined process where users are guided through the different process steps. A very basic implementation are installation wizards, which run during setup of e.g. windows programs [7]. The advantage of such process-driven programs is that the user just needs to follow each step so that there is only a low risk of failing it. In complex visual analytical system with high number of analysis features a static process model would not work, however, a process support would enable especially non-expert users in using those systems.

To enable process support, a dynamic approach is described that is based on process mining [1]. Van der Aalst originally introduced process mining to structure and optimize business processes, but the principle idea is modified to apply it on visual analytical system with the goal of providing adaptation functionality to assist the user.

2.1 General Process Support Methodology

The general process support methodology owns three major phases: (1) The training phase by analyzing data, (2) the process and support feature mining phase and (3) the supported data analysis phase. Thereby the first and third phase are identical and use overall the same system and data, but due to trained process model and support feature model the user could be supported by system meanwhile performing the data analysis. But the general analysis capabilities are still the same. The methodology overview is given in Fig. 1.

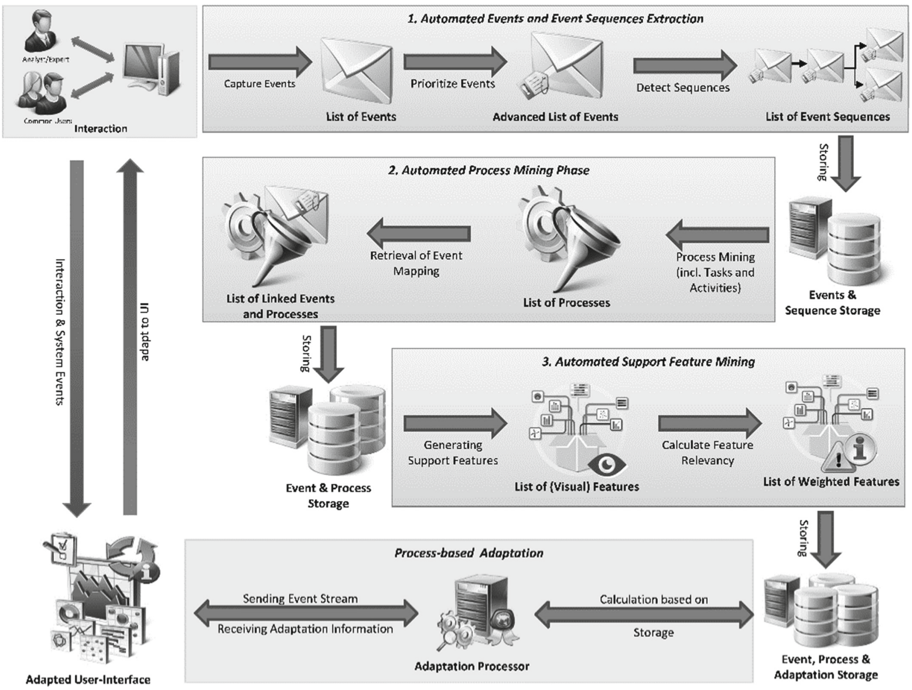


Fig. 1. Generalized interaction and activity processing model to support the user during his analysis tasks.

In the first phase, the training phase, the user already can analyze data with all analysis capabilities, but without any system support. In this phase, the potential users are restricted to analysts and experts who know how to use the system and how to perform a sufficient analysis. While the users are working with the system, the system logs any actions – the major focus lays on interaction events, such as mouse clicks and keyboard input, but also internal system states are logged.

In the second phase, the system performs the mining. These phase can be divided in three major steps: (1) Event extraction and event sequence determination, (2) the process mining to generate a process model and (3) the support feature mining where, relating to the generated process, concrete support features are determined and generated, such as recommendation and interface adaptations to certain databases, visualizations or filter criteria while performing a specific process task.

2.2 General Model Definition

The general Processing mining is majorly based on the definition of van der Aalst [1] with just some minor modification for different usage purpose. To perform the process mining, first an even log L is defined. The event log consists of a variety of structured events of the system, such as:

- System state events, particularly when internal actions happened such as configuration file read or reload of data is initialized;
- Interaction events of the user, particularly when the user clicks on graphical element. Important to mention is that it needs to be distinguished between i.e. the mouse click itself on a graphical spot and the meaning of the interaction such as an item selection, the choosing of a specific data source, the selection of a specific visualization etc.;
- Data events, such as changing, reloading, or filtering of data;
- Visualization events, such as parametrization of a visual layout, reordering of elements, focus or selection of concrete entities;
- Explicit task and activity events, which “bookmark” the beginning of specific real analysis action;
- Adaptation events, that are result of the used process support.

Consequently, the amount of all possible events $\mathcal{E}$ is defines, whereas an event $e \in \mathcal{E}$ is characterized by a set of attributes defined as Θ with a set of attribute names, specifically: an iteratively given *id* (e.g. “151451519”); a *timestamp* (e.g. “2019-09-24 08:32:12.250”); an *activity* (e.g. “visualization selected”); a *resource* (e.g. “icon slice chart”); a *target* (e.g. “slice chart visualization”); a *priority* for the event (e.g. “high”); a list of further *properties*, demanding on the event behavior such as document ID, which are only important for support feature mining stage later on.

For any event $e \in \mathcal{E}$ and name $n \in \Theta$, $\#_n(e)$ is the value of attribute n for event e. If event e does not have an attribute named n, then $\#_n(e) = \perp$ (null value). For convenience the following standard attributes are assumed: $\#_{id}(e)$ is the *id* associated to event e; $\#_{timestamp}(e)$ is the *timestamp* associated to event e; $\#_{activity}(e)$ is the activity associated to event e; $\#_{resource}(e)$ is the resource associated to event e; $\#_{target}(e)$ is the target associated to event e; $\#_{priority}(e)$ is the priority associated to event e; $\#_{properties}(e)$ is the *list of further properties* associated to event e.

To name the events for any event $e \in \mathcal{E}$ is defined, whereas $\underline{e}$ is the name of the event that $\underline{e} = \#_{activity}(e)$. Additionally, it is defined that $\mathcal{A}$ is the amount of all possible activities, so that counts $a \in \mathcal{A}$.

Van der Aalst [1] mentions that a process consists of cases and cases consist of events, which finally would mean that an event log consists of cases, and cases consist of event. But due the fact that annotations for cases are not exist, is defined that a process can consists of logical processes (for a better distinguishing they can be named as subprocesses) and those consist of events. A logical process helps only to structure/cluster a bigger process in smaller units, but basically this alignment is from the principle processing point of view not existing and not required. Finally, this results in the definition that a process consists of events only.

Therewith the event log can be defined with let $\mathcal{A}$ be a set of activity names. The trace σ is a sequence of activities, i.e. $\sigma \in \mathcal{A}^*$. The event log L is a multi-set of traces over $\mathcal{A}$, i.e. $L \in \mathbb{B}(\mathcal{A}^*)$ while $\sigma \in L$ implies $\sigma \neq \emptyset$. In regards of the process mining, the log-based ordering relations are defined first. Therefore, $a, b \in \mathcal{A}$ and has to count:

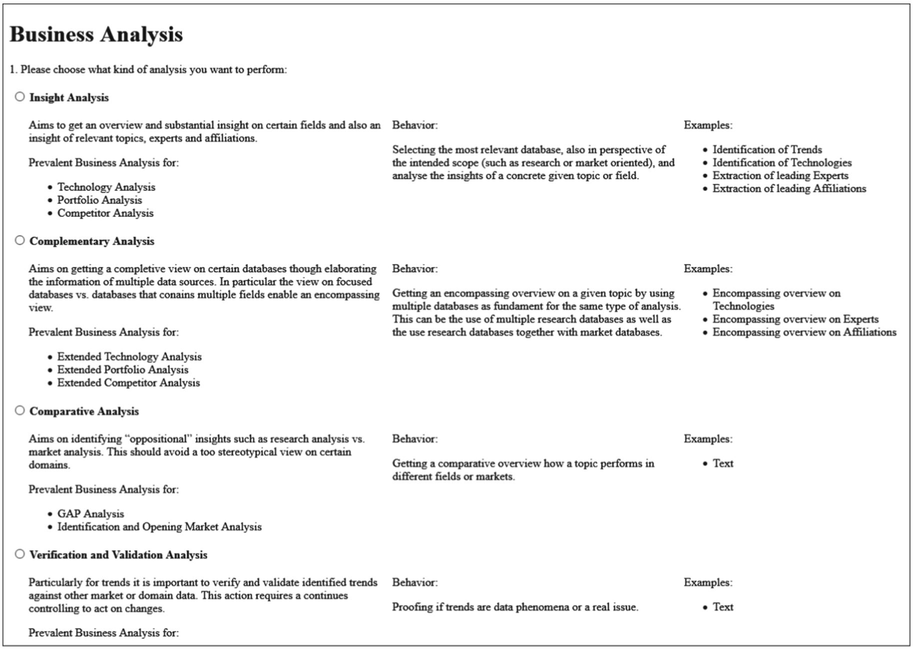
- $a >_L b$ if and only if there is a trace $\sigma = \langle t_1, t_2, t_3, \ldots, t_n \rangle$ and $i = \{1, \ldots, n-1\}$ such that $\sigma \in L$ and $t_i = a$ and $t_{i+1} = b$;
- $a \rightarrow_L b$ if and only if $a >_L b$ and $b \not>_L a$;
- $a \#_L b$ if and only if $a \not>_L b$ and $b \not>_L a$ and $a \parallel_L b$ if and only if $a >_L b$ and $b >_L a$.

For the process mining, the definition of the α-algorithm [1, 2] is considered and L is an event log over $T \subseteq \mathcal{A}$. Therefore $\alpha(L)$ is defined as follow:

$$
\begin{aligned}
&T_L = \{t \in T \mid \exists_{\sigma \in L}\, t \in \sigma\}, T_I = \{t \in T \mid \exists_{\sigma \in L}\, t = \mathit{first}(\sigma)\},\\
&T_O = \{t \in T \mid \exists_{\sigma \in L} t = \mathrm{last}(\sigma)\},\\
&X_L = \{(A, B) | \mathrm{A} \subseteq T_L \wedge \mathrm{A} \neq \emptyset \wedge \mathrm{B} \subseteq T_L \wedge \mathrm{B} \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B} \mathrm{a} \rightarrow_L \mathrm{b} \wedge\\
&\forall_{a_1, a_2 \in A} a_1 \#_L a_2 \wedge \forall_{b_1, b_2 \in B} b_1 \#_L b_2\},\\
&Y_L = \left\{(A, B) \in X_L \mid \forall_{(A', B') \in X_L} A \subseteq A' \wedge B' \subseteq B' \Rightarrow (A, B) = (A', B')\right\} \quad (1)\\
&P_L = \{\mid (\mathrm{A}, \mathrm{B}) \in Y_L\} \cup \{i_L, o_L\},\\
&F_L = \left\{\left(a, p_{(\mathrm{A},\mathrm{B})}\right) \mid (\mathrm{A}, \mathrm{B}) \in Y_L \wedge a \in A\right\} \cup \left\{\left(p_{(A,B)}, b\right) | (\mathrm{A}, \mathrm{B}) \in Y_L \wedge b \in B\right\} \cup\\
&\{(i_L, t) \mid t \in T_I\} \cup \{(t, o_L) \mid t \in T_O\}, \text{and}\\
&\alpha(L) = (P_L, T_L, F_L).
\end{aligned}
$$

2.3 Explicit Versus Implicit Task Selection

The idea is majorly based of a passive system that most of time only observes the user's interaction and system status and presents support wherever it may help the user. However, for major task, particularly at the very beginning of the analysis, it is almost impossible to detect the analysis purpose automatically or gather the information implicitly and therewith a guidance is almost impossible as well. To solve this beginning challenge, an explicit task selection as well is provided. Figure 2 illustrates the explicit task selection at the very beginning, but even during the analysis the system can request if different activities may be possible and cannot be identified via the implicit event observations.



Business Analysis

1. Please choose what kind of analysis you want to perform:

○ **Insight Analysis**

Aims to get an overview and substantial insight on certain fields and also an insight of relevant topics, experts and affiliations.

Prevalent Business Analysis for:

- Technology Analysis
- Portfolio Analysis
- Competitor Analysis

Behavior:

Selecting the most relevant database, also in perspective of the intended scope (such as research or market oriented), and analyse the insights of a concrete given topic or field.

Examples:

- Identification of Trends
- Identification of Technologies
- Extraction of leading Experts
- Extraction of leading Affiliations

○ **Complementary Analysis**

Aims on getting a completive view on certain databases though elaborating the information of multiple data sources. In particular the view on focused databases vs. databases that conains multiple fields enable an encompassing view.

Prevalent Business Analysis for:

- Extended Technology Analysis
- Extended Portfolio Analysis
- Extended Competitor Analysis

Behavior:

Getting an encompassing overview on a given topic by using multiple databases as fundament for the same type of analysis. This can be the use of multiple research databases as well as the use research databases together with market databases.

Examples:

- Encompassing overview on Technologies
- Encompassing overview on Experts
- Encompassing overview on Affiliations

○ **Comparative Analysis**

Aims on identifying "oppositional" insights such as research analysis vs. market analysis. This should avoid a too stereotypical view on certain domains.

Prevalent Business Analysis for:

- GAP Analysis
- Identification and Opening Market Analysis

Behavior:

Getting a comparative overview how a topic performs in different fields or markets.

Examples:

- Text

○ **Verification and Validation Analysis**

Particularly for trends it is important to verify and validate identified trends against other market or domain data. This action requires a continues controlling to act on changes.

Prevalent Business Analysis for:

Behavior:

Proofing if trends are data phenomena or a real issue.

Examples:

- Text

Fig. 2. Example screenshot of an explicit task selection.

A challenge is that the cases (see [1]) are ignored, which could bring a better structuring in the process, but yet it is aimed on an almost automatic approach where not stringent such tasks are explicitly named. It would be impossible to generate a solid process model when users are not giving regularly their intended analysis task or activity. In consequence, additional special events are defined, that are used as bookmarks for concrete beginnings of tasks and activities.

2.4 User Support Features

To enable an advanced analysis based on a process model, different support features are considered. So, the process mining is not the major result, it is more the fundament to use it as baseline to support the user in solving tasks, since the system "knows" the analysis procedures. The following listing is not complete, it should only outline the most important support features, which specifically make use from of the mined process model.

User Guidance. One of the major advancements of the system is to support users in their analysis work. Therefore, the mined process is used as baseline, to identify the current activity. Based on previous analysis activities, it could be furthermore extracted what aspects (i.e. data source, visualization, filter criteria etc.) are most often used and can now be suggested. If a concrete support feature could be identified toward the current user activity, a hint is shown to the user as suggestion – here the user can accept or decline it (see Fig. 3). If an aspect has a high probability, the system automatically performs the change – but, this function is very restrictively used, since it can confuse the user if e.g. visualization changes within a session without any notice.

The major goal of the user guidance features is to provide the user an invisible assistance through the analysis routine until the analysis goal is achieved. In previous works a workflow visualization based on BPMN [6, 7] was used to show the procedure as orientation, but first this requires extra space on the screen and second it was mostly hard understood due to the unknown notation and presentation. As future idea it will be tested to use abstracted process visualizations that just show e.g. the five most important analysis steps, which normally means that the mined process has to be simplified and abstracted manually.

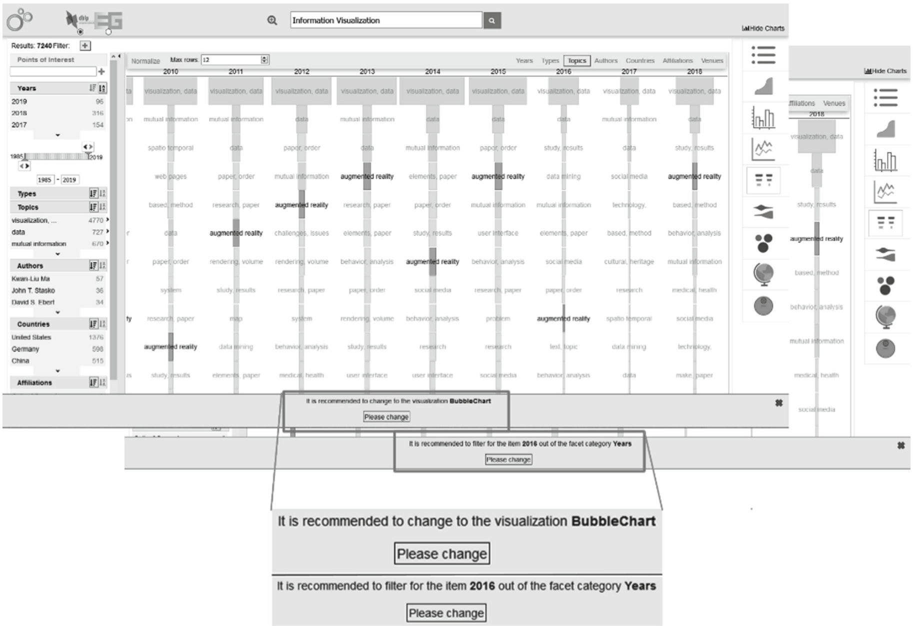


Fig. 3. Guiding the analyst through showing hints.

User-Interface Adaptation. The more the current analysis fits to an already learned analysis, the better the system can support the user via user-interface adaptation. Hereby the system notices the "intention" of the user and initiate changes on visual variables, visual layouts or the entire interface on behalf of the current performed activity. Thereby the functionalities are similar to those that are also currently used to guide the user (see section before), but the focus is more on changing visual aspects to gain a specific view on the data. Two examples are shown in Fig. 4, where the adaptation switches from a single view to a comparative view, because the system recognizes that the user wants to compare different data. And in Fig. 3, the system notices to change visualization layout or rather a specific filter, to show a specific data piece that most likely will be of interest for the user.

It is to mention that there are two different adaptation capabilities to consider: (1) a soft switch or rather a hint and (2) a direct adaptation. The soft switch is still too low to estimate that the change will almost face the user's intention. The second method is used, when the system could predict with a close to total probability what the user intents to. This last method is often used only at the beginning of an analysis. In future research, the individual user behavior should be given greater consideration by including user-based adaption capabilities (likewise described in [8, 9]).

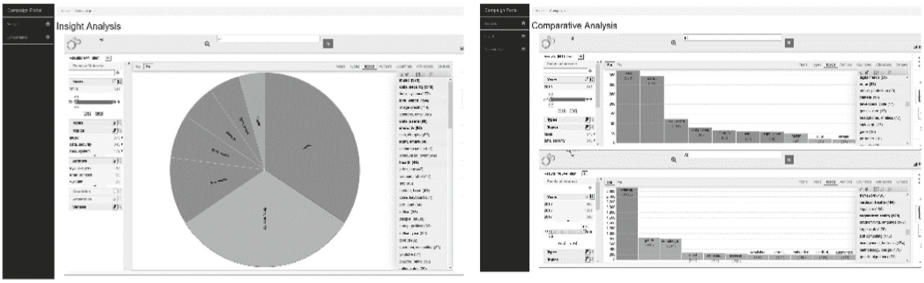


Fig. 4. Adaptation of the User-Interface demanding on required functionalities.

Interactive and Dynamic Interface. To be able to perform detailed analysis, the interface has an important role, since a solid analysis stays and falls with the analytical possibilities. So, the more interactive and from the analytical perspective dynamical an interface is, the more options it enables to analyze the data and find (unexpected) insights. Even more in previous works is outlined that a user support has only significant advantages if the system is highly dynamic and interactive and provides so many ways to e.g. analyze data, that a non-professional user is overstrained quickly. Here a process-driven support can help to structure the analysis process and recommends useful tools when the user really could need them.

3 Conclusion

The digitalization is a challenge in many domains where upcoming new technologies, business ideas, strategies etc. can revolutionize entire businesses. Particularly the transportation and logistics domain have huge variety of changes – to name just a few, there are significant optimizations of the transportation processes, new upcoming logistic ideas such as the delivery via drones or alternative delivery via packaging stations, but also the development of new transportation vehicles that more sustainable. To identify such trends, it is important to use analytical systems, that enable the identification of new trends, next to a rating in perspective of relevance and prediction of the impact on the own business. Even if those tools do exist, they are almost expert tools for professionals, whose are rare in SME. Therefore, in this chapter a new approach was described to assist and support particularly non-expert users in performing such (visual) trend analysis. As major fundament, process mining is used to generate common analysis processes and additionally extract supporting features that are commonly used by users. On behalf of these processing, a process support can be provided that guides and assist especially non-experts through the complex analysis tasks.

Acknowledgements. This work was partially funded by the Hessen State Ministry for Higher Education, Research and the Arts within the program “Forschung für die Praxis” and was conducted within the research group on Human-Computer Interaction and Visual Analytics (https://www.vis.h-da.de).

References

1. van der Aalst, W.: Process Mining: Data Science in Action. Springer, Heidelberg (2016)
2. Weijters, A., van der Aalst, W.: Rediscovering workflow models from event-based data using little thumb. Integr. Comput. Aided Eng. **10**(2), 163–190 (2003)
3. Nazemi, K., Burkhardt, D.: Visual analytics for analyzing technological trends from text. In: Proceedings of 23rd International Conference Information Visualisation (IV2019), pp. 191–200. IEEE (2019)
4. Nazemi, K., Retz, R., Burkhardt, D., Kuijper, A., Kohlhammer, J., Fellner, D.: Visual trend analysis with digital libraries. In: Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business, pp. 14:1–14:8. ACM (2015)
5. Nazemi, K., Burkhardt, D.: A visual analytics approach for analyzing technological trends in technology and innovation management. In: Advances in Visual Computing, pp. 283–294. Springer, Cham (2019)
6. Burkhardt, D., Nazemi, K, Kohlhammer, J.: Visual process support to assist users in policy making. In: Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling, pp. 149–162. IGI Global, Hershey (2014)
7. Burkhardt, D., Ruppert, T., Nazemi, K.: Towards process-oriented information visualization for supporting users. In: Proceedings of 15th International Conference on Interactive Collaborative Learning (ICL), pp. 1–8. IEEE (2012)
8. Burkhardt, D., Nazemi, K.: Dynamic process support based on users’ behavior. In: 15th International Conference on Interactive Collaborative Learning (ICL), pp. 1–6. IEEE (2012)
9. Nazemi, K.: Conceptual model of adaptive semantics visualization. In: Adaptive Semantics Visualization. Studies in Computational Intelligence, vol. 646, pp. 193–297. Springer (2016)
10. Kayikci, Y.: Sustainability impact of digitization in logistics. In: Procedia Manufacturing, vol. 21, pp. 782–789. Elsevier (2018)