

Article

Process Discovery in Business Process Management Optimization

Paweł Dymora ^{1,*}, Maciej Koryl ² and Mirosław Mazurek ¹

¹ Faculty of Electrical and Computer Engineering, Rzeszów University of Technology, 35 959 Rzeszów, al. Powstańców Warszawy 12, Poland

² Asseco Poland S.A., ul. Olchowa 14, 35 322 Rzeszów, Poland

* Correspondence: Pawel.Dymora@prz.edu.pl

Received: 13 July 2019; Accepted: 27 August 2019; Published: 29 August 2019



Abstract: Appropriate business processes management (BPM) within an organization can help attain organizational goals. It is particularly important to effectively manage the lifecycle of these processes for organizational effectiveness in improving ever-growing performance and competitiveness-building across the company. This paper presents a process discovery and how we can use it in a broader framework supporting self-organization in BPM. Process discovery is intrinsically associated with the process lifecycle. We have made a pre-evaluation of the usefulness of our facts using a generated log file. We also compared visualizations of the outcomes of our approach with different cases and showed performance characteristics of the cash loan sales process.

Keywords: business processes management (BPM); process discovery; BPMN; CMMN; process-aware information system; business process redesign (BPR)

1. Introduction

An optimal approach to policy setting, allowing system architects to automate application processes, is a crucial issue in today's software engineering and complex engineering systems. For many years business process management (BPM) approach has been used to analyze business processes in many organizations. Traditional business process management methods assume the following five stages in the business process cycle: Process design consisting of creating a new or changing an existing process model; creating a process instance, its configuration and startup; executing; monitoring the execution of the process instance; and analyzing the performance of the process for improvement [1–3].

According to Reference [4], the development of business process models plays an essential role in the management of business processes. Achieving these models can take up to 60% of the project time. This time can be shortened by automatic or semi-automatic process model generation methods. The authors of Reference [4] used one of the existing algorithms of process detection or execution of a composition algorithm based on an activity graph, which generates a process model directly from an input event log file. The proposed approach allows process operators to take part in process modeling. It can also support business analysts or process designers in visualizing the workflow without the need to design the model in a graphical editor.

In recent years, an essential change in the trend of transaction processing may be observed, as follows: The traditional BPM approach has been replaced with the Case Management one with process discovery methods, which are more and more often associated with the process lifecycle. Such an approach allows the transformation of input data from event logs registered during process execution into process models [1–6]. This transformation happens without the existence of any prior information about how the intended process should look. The built-in model usually takes the form

of a graph, as follows: Petri net, BPMN (Business Process Model and Notation), EPC (Event-driven Process Chain diagram), or UML (Unified Modeling Language) activity diagram [2,3,5]. Process model validation methods consist of a comparison of an existing process model to the actual event log records. Model validation can be used to verify that the process instances recorded in the event log are compatible with the model, and vice versa. Methods of extending the process model include the expansion and improvement of the current process model using contextual information on the real implementation of activities registered in the event log. Meanwhile, process model validation evaluates the compatibility of the model with actual process data and the development of the process model is aimed at modifying or enlarging the existing model [1,3,7,8].

The connection between well-grounded social theories, like the social emergence theory, and techniques of data mining, primarily process mining, may result in the development of a system supporting modern organization in doing process-aware activity in a dynamically changing business environment [2,3,8,9].

2. Related Works

Process discovery is a growing and promising scientific area, which concentrates on the development of theory and techniques of gathering and representing knowledge about real processes execution and the other rules driving the organization activity. The recent twenty years were fruitful in novel research in this area [10]. A systematic survey presented in Reference [10] shows that most active research subjects are focused on process mining algorithms working out and optimization, with the conformance checking techniques and software architecture and tools enhancement. As regards the domains where those efforts are applied, the health segment is very significant and then communication, IT, manufacture, education, logistics, and finances. Our research domain is the finance domain, wherein we have a long time experience and constant access to extensive real-world data.

The authors of Reference [11], after analysis of 705 papers, showed that the primary type of process mining activity studied is “discovery” (71%) and “categorical prediction” is the challenge most often taken (25%). They also report that the “graph structure-based” technique is the most frequently used one (38%) and they concluded that computational intelligence and machine learning techniques are still rather loosely tied with the process mining discipline. “Evolutionary computation” (9%) and “decision tree” (6%) are the most applied. Many authors, e.g. Reference [12], prove that the process representation most often used is the Petri net because of its well-grounded theory and its computability, and then the Business Process Model and Notation (BPMN) because of its broad application in tools and its everyday use in the business world. Some authors present the discussion on abstractions and process representations to reflect on the gap between process mining literature and commercial process mining tools [12]. That discussion facilitates users to select an appropriate process discovery technique.

In References [3,6,13], the authors present some of the essential methodologies of process model generation. The first group is a set of methods which use the up-front design of the process model. As an example, the technique which allows the generation of process models from text descriptions based on natural language or methods which allow for creation of new models from existing models, and in particular from a set of existing process diagram or through translation from other representations, such as UML, data models, or other diagrams, may be used. Those methods require additional investment to design other intermediate models. The authors also distinguish methods for generating imperative process models (focusing on explicit process definition) based on declarative models (implicit definitions, which are described by directives and policies), as well as hybrid solutions in which a hybrid process model combines data-oriented declarative specification and flow-oriented imperative control.

The second group mentioned by the same authors contains methods based on event log analysis. The event log includes a set of execution sequences, based on which the relationship between the different activities may be determined. Such real data can be used to build a process model directly

from its working implementation. Based on isolated use cases, archival data, and negative logs (which represent workflow sequences not allowed during the process execution), such method allows for optimal redesign of the initial business process model. One of the first steps in this method is for the developer to determine the initial and final state of the process and the variants of transitions performed by actors on specific resources. A particular business process consists of multiple use cases. Such a structure can be defined by a set of n -element vectors that describe the order and number of executions for each task. Then, the event log is generated and analyzed. The event log is a set of traces of events that can be defined as an ordered sequence of activities in the process. Traces are described by the use case identifier, the event identifier, the timestamp, the activity name, the resource, and the cost. Based on such a record, it is possible to determine many useful statistics, such as the number of events in use cases, the duration of the case, etc. Next, the mining-driven approach may be applied to generate a BPMN diagram using the event log generated from the business process model; in our case, it is a cash loan process. There are many process discovery algorithms which can be used in the BPMN diagram building process, for example, the α -algorithm (abstraction-based algorithm), the heuristic miner (heuristic-based algorithm), the ILP Miner (language-based algorithm), or the inductive miner (inductive-based algorithm).

Some authors (e.g., Reference [14]) raise the significant problem of the quality of event logs. They notice that processes discovered using logs of low quality will also be of low quality, resulting in incorrect business decisions taken upon them. Authors of Reference [14] propose the use of methods based on autoencoders, which are a class of neural networks, to reconstruct the leaking values in logs. They report a significant improvement of the model quality after applying that method.

A considerable amount of research in the process discovery area is done under the adaptive case management (ACM) umbrella. ACM [15,16] is a relatively novel approach in the BPM world, which operates in many connected areas such as knowledge extraction, knowledge storage and sharing, empowerment of workers, advanced collaboration, adaptability, and guidance techniques [17]. Our work supports a few of them, especially knowledge extraction and storage, and superior collaboration done by our omnichannel business model (OBM). The great challenge of today's systems is the multichannel access to shared data with the growing significance of mobile channels. This question is increasingly raised by researchers, e.g., Reference [18]. Social network aspects of process management are studied in Reference [18], where the authors introduce a framework for supporting case management in social networking environments. The structure can capture and formalize communication acts between cooperating people, making them usable in future processes.

Since the business process is recognized as a set of related events and decisions, trying to respond to the increasingly demanding expectations of the markets, it is necessary to implement mechanisms of the dynamic redesign of basic workflows in business processes to improve performance indicators. Many authors focus their work on the possibility of introducing business process redesign (BPR) mechanisms [19]. For example, in Reference [5], the authors focus on the presentation of the evaluation mechanism that measures the effectiveness of the BPMN model and its ability to effectively transform it into a directed acyclic graph and to determine the average efficiency of process execution. The proposed mechanism evaluates the model type, complexity indicators, and the ability to standardize and optimize candidate process models, while at the same time allowing users to set the desired complexity thresholds. The authors of Reference [20] propose process reengineering ontology—based knowledge map methodology (PROM) to reduce the failure ratio and solve BPR problems. They use analytical hierarchy processing to identify and prioritize the processes of the business to be re-designed. That technique enables creation of process-aware software and process-driven applications (PDA) based on the ontology and knowledge maps.

As an alternative to the classical process model-driven approach, some authors propose activity modeling and simulation based on agents (agent-based simulation, ABS). According to Reference [21], ABS will get more attention in the field of BPM similarly to other areas like, e.g., social sciences. Following non-standard solutions, some authors use the axiomatic fuzzy set (AFS) theory to convert

event logs into fuzzy sets, resulting in a powerful tool to model and store human knowledge and behavior [22]. In recent years, such an approach is more and more often applied in various disciplines—business intelligence, financial analysis, clinical tests, etc. The still existing drawback of that direction is the complicated implementation, especially in systems where performance requirements are high. Similarly, the adaptation of latent Dirichlet allocation (LDA), proposed by the authors of Reference [23] as an unsupervised machine learning technique to automatically detect and assign labels to activities, encountered difficulties in an implementation, but it seems to be a promising direction since this framework does not require any human intervention.

Our work combines a few techniques described above and a few of them we consider to embed into our framework in the future. First of all, we use well-grounded methods of process mining as the primary source of knowledge in the system. We connect them with existing BPM and ACM tools used in enterprises and an omnichannel cooperation platform, obtaining the solution called the omnichannel business model (OBM). The solution (a) allows catching and storage of the knowledge from the cloud of information hidden in the enterprise resources and (b) can automate future processes and general enterprise activity using the unhidden knowledge and BPR techniques.

3. New Business Management Model

A process-aware information system (PAIS) is a kind of software which supports organization in business process automation involving many accessible assets, such as data sources, software applications, and primarily enterprise employees and customers. PAISs usually use up-front designed scenarios to execute an ordered set of activities leading to the goal of the process.

Process mining is a data mining technique which enables the ability to discover knowledge about real processes in an organization hidden in the data produced during its regular activity. It automatically constructs models, usually using a Petri net or BPMN notation, taking the stream of events registered by the enterprise software as an input [2,3]. Three main aspects of business process development are usually covered by the process mining, as follows: Process discovery, which gives an outlook of the real flows of processes, conformance checking, which allows for finding gaps between assumptions and the reality, and process improvement, which helps to minimize those gaps.

In this paper, we demonstrate selected process mining techniques used to discover process models from event logs, find logs, and model deviations, and demonstrate the performance characteristics of the process of a cash loan sale as an example. We also suggest a new approach for building a software system which aims to help the community capture its emergent behavior. Using the analytical approach, we have sketched the required system model called the omnichannel business model (OBM), which is shown as a component model in Figure 1. The OBM is an example of a PAIS which provides a unique emergence-aware logic embedded together with domain-specific enterprise logic. That additional logic is provided by a few software modules, described in greater detail in Reference [9] and shortly sketched below.

Adaptive case management engine (ACM) is the business process management software aiding a novel approach to support people doing a process-aware activity, where the process is not described up-front, but is orchestrated on-the-fly by workers doing their job with the domain knowledge. The process mining engine (PME) is a module which enables discovery of process models and provides their verification and expansion based on data from event logs which describe real, not hypothetical, business processes. The business rules management system (BRMS) is the module used to automate flow in the decision points of processes. Key performance indicators (KPI) are the set of procedures which calculate efficiency indicators and make them available to workers. The social module (SM) engages cooperating parties in formal and informal communication, increasing the level of social awareness. The business process management engine (BPM) enables automation of activity sequences discovered by the system. The knowledge engine is the main part of the solution which coordinates the whole system and makes the gathered knowledge maximally usable [9].

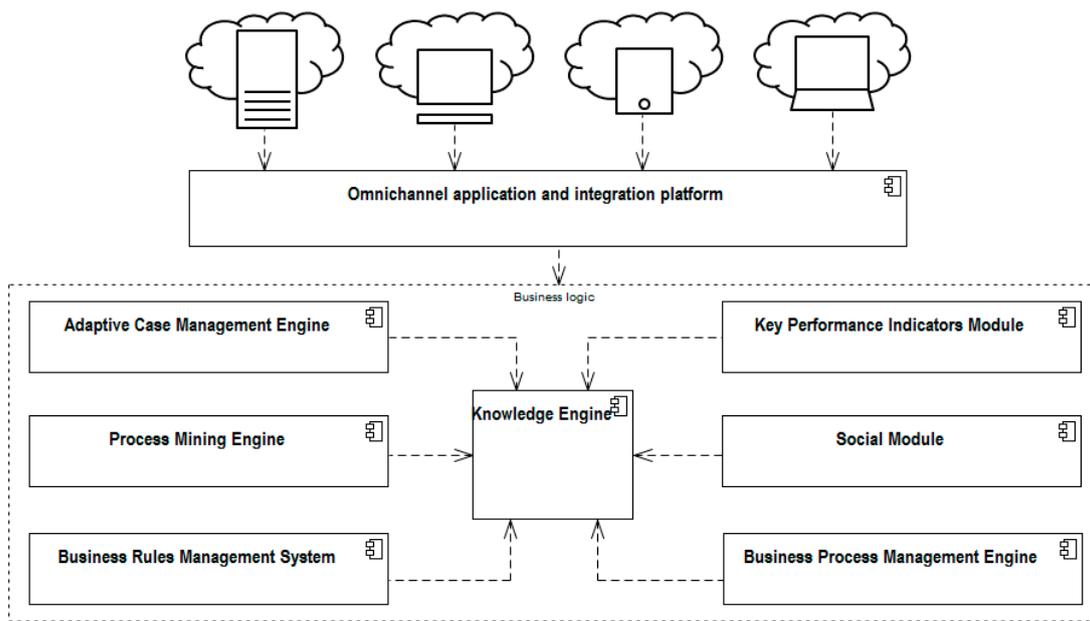


Figure 1. Proposed omnichannel business model.

4. Modeling the Process of a Cash Loan Sale

In this paper, we present means of data mining discovery in terms of the omnichannel business model, an emergent system class which is characterized by a self-organization feature. We create a model of the real business process of a cash loan and then we discuss key steps of process mining and cost estimation. The cash credit sale process is usually tied with high levels of some efficiency indicators, such as cost efficiency or the speed of decisions. This requirement has a critical impact on the level of automation offered by the system [1,6,7]. The discussed process is built on the set of use cases (UC), shown in Figure 2. Each use case may be performed by one or more business roles fulfilled by one or more persons (connections between actors and use cases do not unambiguously appoint who is responsible for the use case; e.g., in the figure, some of them may be performed by two roles). The model also neither orders use cases nor defines which are necessary to achieve the goal of the process.

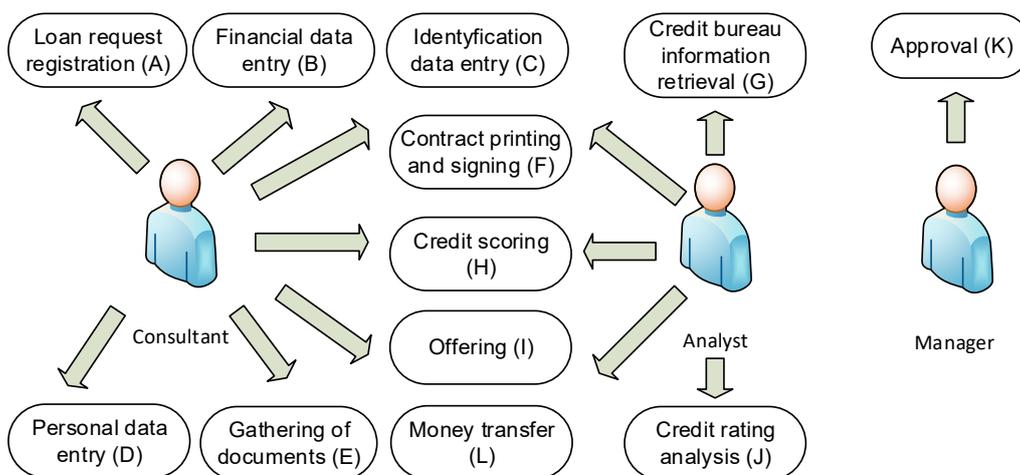


Figure 2. Use cases (UC) in the process of a cash loan.

In the real business environment, many possible traces may be used to move from the starting point of the process to its final state. Even many of the use cases have constraints disabling their use in

some circumstances. Typically, a few of them may be used in a different order, resulting in hundreds of possible solutions as regards the whole population of process traces. Usually, this puzzle is solved (better or worse) by business analysts, process engineers, or other highly skilled workers.

In our model, we propose discussion of three main event paths. A first variant called passive-defensive (L_1), the second is called active-defensive (L_2), and the third is called active-offensive (L_3), where the first part of the name concerns level of pro-active behavior of sellers (the passive seller only follows customer needs, the active seller tries to maximize transaction) and the second part concerns the level of the financial risk involved (the offensive variant involves higher risk expecting higher sale) These variants are defined as follows:

$$L_1 = \langle A, C, D, B, G, H, E, J, K, F, L \rangle,$$

$$L_2 = \langle C, D, B, G, H, I, E, J, K, F, L \rangle,$$

$$L_3 = \langle B, I, C, G, D, H, I, E, J, K, F, L \rangle.$$

In the literature, many different notations for process models may be found. At the first step of the analysis, Petri nets are used, especially its subclass known as WorkFlow nets (WF-nets) [3–6,13,24–26]. A WF-net is a Petri net with a defined starting point for the process and a separate ending point for the process [2,4]. All nodes are on a path from source to sink. Figure 3 presents a WF-net generated for our model of the cash loan process. The model is designed to describe the handling of an application for a new cash loan where customers may apply for a money transfer from a bank. As Figure 3 shows, the process may start by a loan request registration (A) request or other activities like financial data entry (B) or identification data entry (C). Each action is represented by a transition that is a square. Transitions are interconnected by points that model the potential process status. Each point is defended by a circle. In a Petri net, a transition is activated, i.e., a proper action may be performed if all input points have a token. Transition loan request registration (A) has only one input point (start), and this point initially includes a token to represent the request for compensation. Other transitions like financial data entry (B) or identification data entry (C) have two input points. The transition uses one token from each of its input points and creates one token for each of its output points.

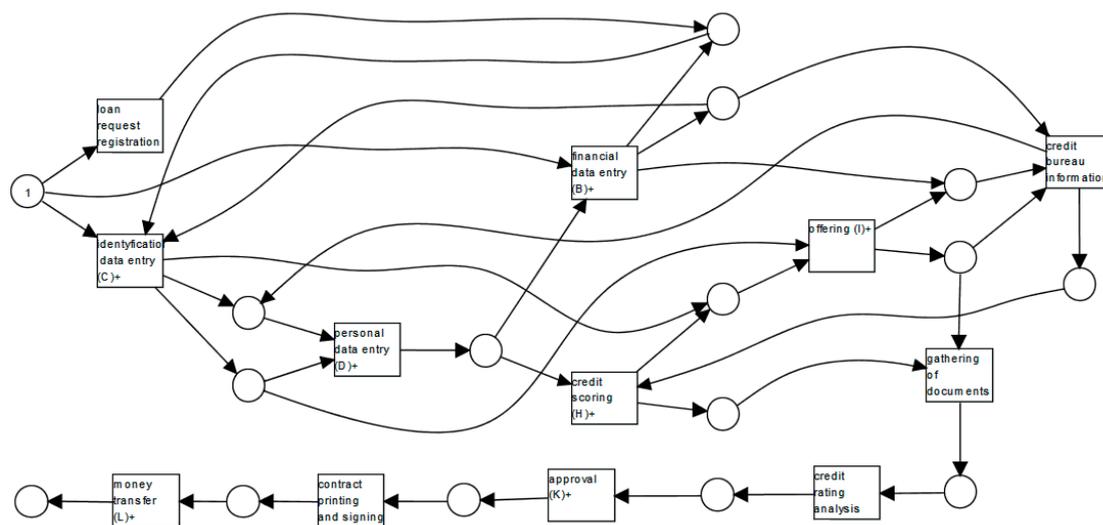


Figure 3. WF-net discovered for L_1, L_2, L_3 .

Therefore, starting the transition, i.e., loan request registration (A), causes the termination of the token from the initial point of entry and the creation of the token for the output place. Tokens are presented as black dots. The setting of tokens in specific points, in this case, the request state, is defined as marking. In our WF-net model, we can distinguish activities which may be processed by different job roles, i.e., consultant, analyst, and manager. At many use cases, the same activity may also be

processed by a different person with the same job role. Additionally, some activities like contract printing and signing (F), credit scoring (H), and offering (I) may be realized by the consultant job role as an analyst. Let us notice that the activity approval (K) is assigned only to the manager. The process ends after paying the money transfer (L) activity of the cash loan request.

In Figure 4, we present the same process model in terms of the BPMN diagram [1–3]. The business process model and notation (BPMN) uses explicit gateways instead of points for simulating the logic of control flow. The diamonds with an “x” mark indicate XOR split/join gateways, whereas diamonds with a “+” mark indicate AND split/join gateways. The presented graphs were obtained with the use of RapidMiner Studio, a popular mining tool, with the ProM extension (RapidProM). This software is an upgradeable structure that handles a large number of process mining techniques, which are prepared and distributed as extensions and plug-ins [2–7,13,26].

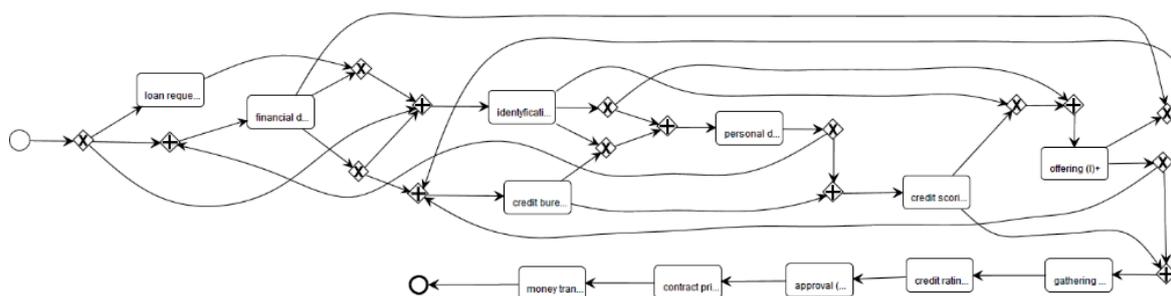


Figure 4. Cash loan process in terms of BPMN.

5. Process Discovery and Aspects of the Emergence

Process mining and process discovery is a new trend in scientific research. These techniques focus on the extraction of knowledge about a (business, system) process from its execution logs. Process mining enables system engineers to see the system (process, application) from different perspectives, such as the process (or control flow) perspective and/or the performance, data, social network, and organizational perspectives. We can say that process mining, as well as process analysis and modeling, is an entirely new research discipline that can be placed in the field between data mining and machine learning [1–7,27]. By redesigning business processes, organizations make significant changes to improve their performance indicators.

Following the presented idea of the omnichannel business model (OBM), especially its process mining engine (PME) module, we can implement in it the process discovery techniques in order to confirm the adequacy of the conceptual model of the system. Conformance checking is used to verify real system functions with recorded activity flows and to quantify and diagnose its deviations. Such knowledge may improve an existing process model by modifying or extending the a priori model to reflect reality, sufficiently retrieving an event log, and creating a model with no use of any a priori information.

In the analyzed system, we aim to discover new rules in existing event logs of real processes in order to improve them. These new rules and new activity flows in the process may improve the first system model constructed at the beginning of the system (or application) implementation.

5.1. Methodology

In the approach proposed, existing enterprise-class tools are combined with the sophisticated methods of process discovery, resulting in the comprehensive solution, which may be used in the real enterprise to aid its business activity. The methodology used contains the following:

- (1) A deep study on process mining techniques and the selection of the most appropriate ones, as regards the incorporation into the solution and the ability to work in real business circumstances.

- Both process discovery and social network discovery methods are used to gather the knowledge present in the organization;
- (2) Creating the working omnichannel business model (OBM) solution, which connects the discovery tools, the BPM/ACM tools, and other enterprise tools, and controls/transforms the flow of data produced and required by them;
 - (3) Implementation of business process redesign (BPR) mechanisms to utilize the knowledge discovered;
 - (4) Real-world long-time tests of the system with careful analysis of the results and improvement of the solution, if required; and
 - (5) Assessment of the applicability of methods, algorithms, and tools selected in support of the typical process-sensitive enterprise.

The omnichannel business model (OBM) solution provides the process mining engine (PME) module. This module is responsible for the process discovery and it tightly cooperates with the knowledge engine (KE), the central component of the system. PME acquires information resulting from calculations performed in other parts of the system, processes it, and gathers knowledge in the form required for further automation of processes. For example, the PME module receives information about the BPMN models resulting from the actual activity of the process participants and confronts this information with the indicators derived from the KPI module and the knowledge about the reputation of the participants from other modules. The PME module creates a BPMN model and the resulting knowledge allows it to issue a recommendation on the most beneficial ways to continue the process. In particular, the Alpha algorithm, which records events and creates a BPMN diagram as a Petri net showing the action stored in the log [4,13]. If the event log provides information about resources, individuals may also discover resource-related models, e.g., a social network showing how people work together in an organization [2–4]. After the BPMN model is generated, it has to be enriched with information regarding the effectiveness of possible paths. The model is traversed using paths from all successful cases, and transitions between activities are marked by a unique value. In the simple solution, each transition used gets one point to sum, while in more a sophisticated one, each granted point is multiplied by a value which is a result of the function of measured KPIs of the case [2,3,25]. The activity performer choosing the next step in the process sees punctuation of each reachable path. The punctuation serves as a hint which helps them to select the best processing path.

The most important tasks of the PME module include the following:

- Data transformation between models operating in particular modules (e.g., historical data from Camunda API, event log in XES format, process model in PNML format, Camunda API for process model management, etc.);
- Storage of the proposed knowledge base, created based on the performed tasks;
- Storage of the procedural knowledge base resulting from the discovery of processes;
- The categorization of process participants according to different criteria (reputation, performance, internship, function, etc.);
- Marking and suggesting the most suitable paths;
- Awarding contractual awards resulting from reputation analysis and KPIs (awarding reputation strengthens the level of cooperation that determines emergence).

5.2. Event Log of the Cash Loan Process

In the area of process mining, the carrying out of a process is called a trace or a case and is a sequence of events. A set of traces is called an event log and is a key concept in the field of process mining. Every event log, trace, or event can contain data attributes which must at least describe the executed activity type and may also include other information, e.g., the resource or time information. Every type of activity is also called an event class [1–3].

Figure 5 presents a small fragment of an event log correlated to the process of handling the cash loan requests. Each line in the table shows one event of some business case. We can distinguish a few attributes like case id and event id, which are strongly related to the activity attribute, timestamp, resource, and cost. Each event id is a unique value grouped in a specific case and assigned to an activity realized in the process by a particular resource (user, worker). The activity set contains 12 different actions (see Figure 2) which may be performed by a specific job role, e.g. consultant (resource: Pawel, Mirek, Maciek), analyst (resource: Andrzej, Marek), and also manager (resource: Dominik). Let us notice that some activities may be performed only by a specific job role—only the manager may do approval activity or only the analyst (Marek or Andrzej) may do, e.g., credit scoring, but there are a few activities common for consultant and analyst, e.g., money transfer, offering, or contract printing and signing. An important attribute is a timestamp, which informs when the activity was started. In this event log, the actions are treated as atomic and the table does not disclose the time period of the actions. The table also shows the costs associated with the events. As we can see, two of the same activities may be triggered by different resources in a separate case in almost the same time. However, it is a different business case. Our event log contains 19 cases and 267 events.

Case ID	Event ID	dd-MM-yyyy:HH.mm	Activity	Resource	Costs
1	69463315	06-03-2017:10.12	loan request registration (A)	Pawel	50
1	69463316	06-03-2017:10.22	identification data entry (C)	Pawel	50
1	69463317	06-03-2017:10.32	personal data entry (D)	Pawel	50
1	69463318	07-03-2017:12.43	financial data entry (B)	Mirek	75
1	69463319	07-03-2017:13.23	credit bureau information retrieval (G)	Andrzej	100
1	69463320	09-03-2017:11.25	credit scoring (H)	Andrzej	200
1	69463321	10-03-2017:12.02	gathering of documents (E)	Maciek	250
1	69463322	13-03-2017:09.12	credit rating analysis (J)	Marek	150
1	69463323	14-03-2017:16.19	approval (K)	Dominik	400
1	69463324	15-03-2017:11.44	contract printing and signing (F)	Maciek	100
1	69463325	17-03-2017:14.24	money transfer (L)	Maciek	50
2	69463330	06-03-2017:10.32	identification data entry (C)	Mirek	50
2	69463331	06-03-2017:10.55	personal data entry (D)	Mirek	50
2	69463332	06-03-2017:11.32	financial data entry (B)	Pawel	75
2	69463333	07-03-2017:09.33	credit bureau information retrieval (G)	Marek	100
2	69463334	07-03-2017:11.23	credit scoring (H)	Marek	200
2	69463335	08-03-2017:09.52	offering (I)	Pawel	90
2	69463336	09-03-2017:13.32	gathering of documents (E)	Mirek	250
2	69463337	10-03-2017:13.27	credit rating analysis (J)	Marek	150
2	69463338	13-03-2017:11.29	approval (K)	Dominik	400
2	69463339	14-03-2017:09.24	contract printing and signing (F)	Maciek	100
2	69463340	15-03-2017:14.34	money transfer (L)	Maciek	50

Figure 5. Part of the event log of a cash loan process.

In Table 1, some aggregated information is presented about our event log of the cash loan process. For each case, we can obtain the number of its events, the resource count involved in the activities performed, and, one of the most important things, the case duration time, which is obtained from the start of the first activity in the case through to the completion of the last one. By using timestamps in the event log, we can find bottlenecks, service levels, throughput times, and frequencies. By analyzing the event log of the modeled process information about resources, decision rules, quality metrics, etc., may be obtained.

Table 1. The event log statistics.

Case ID	Events Number	Duration Time	Resources
1	11	11 days, 4 h	6
2	11	9 days, 4 h	5
3	12	14 days, 2 h	6
4	12	11 days, 3 h	6
5	12	14 days, 17 h	4
6	12	17 days, 23 h	4
7	12	15 days, 19 h	6
8	10	13 days, 1 h	5
9	10	12 days, 5 h	5
10	10	13 days, 20 h	5
11	10	10 days, 21 h	6
12	15	19 days, 21 h	6
13	25	27 days, 20 h	6
14	15	14 days, 20 h	6
15	20	20 days, 22 h	6
16	15	12 days, 20 h	6
17	20	21 days, 2 h	6
18	20	23 days, 22 h	6
19	15	16 days, 18 h	6

6. Results

In the accordance with the conceptual model of the system presented in the Figure 1, where the process mining engine is one of its modules, the critical responsibility of the system is to improve business processes using process mining as a tool that gives continuous insight into actors' behavior and enables the opportunity for online improvement [1,2,7].

In the beginning, we assumed that in our example, only three main processing paths would be used, as the outcome of process engineer work told us. After system implementation in the real banking environment, we logged system events in order to obtain hidden information about actual activity flows and other system business parameters. Next, we performed a sophisticated analysis of the event log (which refers to the process mining engine functions in Figure 1). After execution of the Alpha algorithm in a RapidProM environment, which provides a Petri net and a marking [2,4,13,24,26], we discovered some new activity traces (Figure 6).

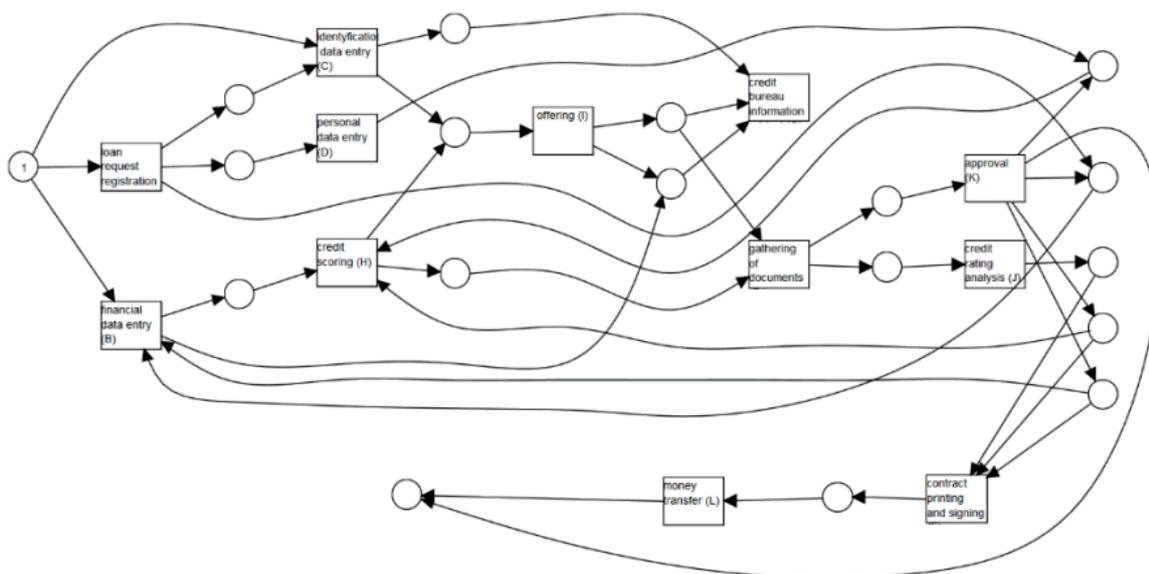


Figure 6. Petri net for a cash loan process recorded in the event log.

We can now update our first model with the obtained knowledge. Furthermore, on the base of the Petri net, with the use of the Petri net to BPMN framework in the RapidProM tool, we created the BPMN graph for a cash loan process recorded in the event log (Figure 7) [13,26].

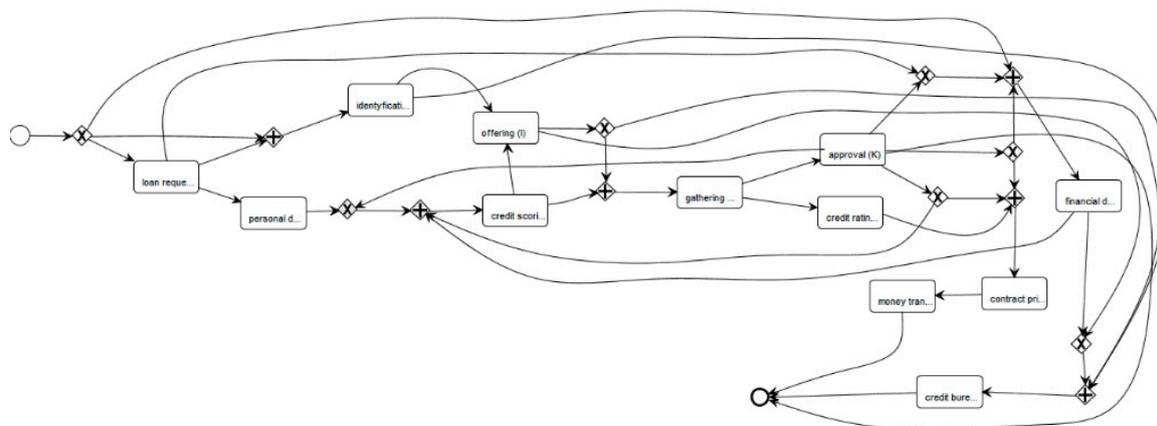


Figure 7. BPMN graph for a cash loan process recorded in the event log.

The efficiency of processes in an organization can be determined in various ways. Usually, three dimensions are identified, as follows: Time, cost, and quality. Each of these efficiency dimensions can be assigned different key performance indicators (KPIs) [2,7,25]. For further analyses, we should determine the activity costs using a causal activity matrix. To do that, we must discover activity clusters based on the cash loan process recorded in the event log. The clustered view of the event log is a product of that step. Clustering algorithms cause partitioning or hierarchical division of objects, in which the most different objects are distinguished in each grouping [2,3,24]. Most clustering algorithms expect a distinction that defines how the two objects are dissimilar. Such a dissimilarity measure is implemented using a distance function. By clustering an event log, a more consistent group of tracks can be achieved. This results in more intelligible exploration of process models. In this process, clusters are created for and associated with every activity in the event log. All direct predecessors and successors of activity will be added to the associated cluster. Clustering refers to building groups of entities that are comparable to each other and not equivalent to entities from other clusters. This mechanism is a technique for identifying information from untagged data. Clustering may be very helpful in many scenarios, e.g., to find groups of resources with similar job behavior [2,7,24].

After clustering, we discovered a causal activity matrix from an event log. Every single cell in the matrix may obtain a value from 1 or -1. A value of 1 indicates that there is a causal relationship between the row-activity to the column-activity. A value of -1 indicates that there is no causal relation. A value of 0 indicates that we cannot find any correlation (indefinite state). Any other positive values from this range indicate a connection [4,5]. For our cash loan process in Table 2, we present the causal activity matrix.

In Table 3, some main activity statistics in the event log are presented. We can analyze the activity total and the relative occurrences in the event log of the cash loan process. A vital activity attribute shown in Table 3 is an activity replay cost factor. Having a causal activity matrix, we may now determine the activity replay cost factor. The algorithm creates a replay cost factor for the given activity cluster array. The cost factor for an activity is correlated with the number of clusters that include this activity. The total replay factor value is 60 [2,7,28].

Table 2. The event log statistics.

	K	F	G	J	H	B	E	C	A	L	I	D
K	-0.78	0.90	-1.0	-0.89	0.83	0.8	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
F	-1.0	-0.81	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.91	-1.0	-1.0
G	-1.0	-1.0	-0.73	-1.0	0.66	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-0.7
J	0.73	0.5	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
H	-1.0	-1.0	-0.60	-1.0	-0.88	-1.0	0.5	-1.0	-1.0	-1.0	0.96	-1.0
B	-1.0	-1.0	0.85	-1.0	0.8	-1.0	-1.0	-0.22	-1.0	-1.0	-1.0	-0.64
E	0.5	-1.0	-1.0	1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
C	-1.0	-1.0	0.8	-1.0	-1.0	-0.30	-1.0	-0.84	-1.0	-1.0	0.5	-0.42
A	-1.0	-1.0	-1.0	-1.0	-1.0	0.8	-1.0	0.83	-0.66	-1.0	-1.0	0.88
L	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-0.78	-1.0	-1.0
I	-1.0	-1.0	0.5	-1.0	-1.0	-1.0	0.99	-1.0	-1.0	-1.0	-1.0	-1.0
D	-1.0	-1.0	-0.23	-1.0	0.5	-0.45	-1.0	-0.58	-1.0	-1.0	-1.0	-1.0

Table 3. The activity statistics.

Class	Occurrences (Absolute)	Occurrences (Relative)	Replay Cost Factor
credit scoring (H)	32	11.99%	12
offering (I)	28	10.49%	12
approval (K)	28	10.49%	20
gathering of documents (E)	28	10.49%	15
credit rating analysis (J)	28	10.49%	20
credit bureau information retrieval (G)	23	8.61%	10
financial data entry (B)	23	8.61%	12
identification data entry (C)	19	7.12%	10
personal data entry (D)	19	7.12%	10
loan request registration (A)	17	6.37%	60
money transfer (L)	11	4.12%	30
contract printing and signing (F)	11	4.12%	20

Creating useful and understandable visualizations of data is an extremely significant area for research purposes. To a large extent, it enables analysts to discover information from data using visual artifacts. BPM is committed to managing business processes using a variety of artifacts and the relationship between people and business processes plays an essential role in managing processes. Such a set of tools enables the effective discovery of social networks from the data of executed processes and can simplify the management of business processes to make them more productive and successful [2,3,29].

The social network in BPM is a concept that describes processes developing jointly and that are cyclically iterated. In literature, it is also known as "socially active processes." These socially supported processes emulate the way work is done from the end-users' perspective and how it is experienced from the user's perspective in order to harness the power of continuous cooperation. Social BPM is at the interface between business processes and joint activities. The combination of BPM and social media complements the interpersonal interactions at work by supporting social networks, cooperation, and communication. Several techniques can be used to analyze social networks, e.g., to identify patterns of interactivity, to assess the role of the individual in the organization, etc. [2,3,29]. In the process instance (case), the work is transferred from one resource (actor) to another. Therefore, knowledge about the structure of the process can be used to detect whether there is a causal relationship between the two actions. It is also possible not only to consider direct and indirect inheritance, to obtain which activity was first completed or not, but also to find hierarchical relations. There are also events such as changing the assignment of an action from one person to another (work delegation). This is an excellent subcontracting metric, which is obtained through social networks. The main idea of this metric is to aggregate the number of cases in which one person performed an action before another

activity performed by another person. It could be evidence that the work has been commissioned between these individuals [28,29].

Active communication between actors (resources) is one of the conditions of emergence. With the social network, we can judge the degree of correctness of communication in the data flow processes. Figure 8 shows a social network of interrelations between the main actors in our processes. The individual resources are placed in circles denoting the between ranking view and are connected with an arc indicating the actors' common correlation. The shape of the presented resources is correlated with the degree and size of the node and is correlated with the resource ranking. In addition, we can set the KPI of mutual cooperation [3,6,25,29]. Reciprocal relationships between actors were possible to obtain by generating the causal resource matrix.

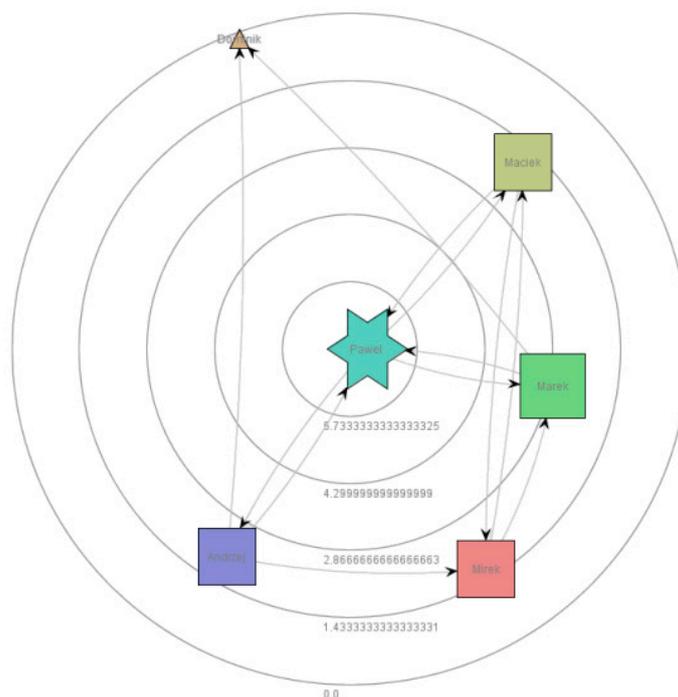


Figure 8. Subcontracting social network.

A significant benefit of social BPM is that it helps to eliminate the barrier between BPM decision-makers and the users affected by their decisions. Compared to conventional approaches to process modeling and management, social BPM engages a larger and more heterogeneous set of actors and aims to achieve a higher quantity, quality, variety, and timeliness of contributions [8,28]. Table 4 demonstrates the participation of each actor in the decision-making process. In addition, the relative frequency of each actor and the resource replay cost factors were determined (as in activity analysis). The total factor for the resources (actors) is 6.

Table 4. Resource statistics.

Class	Frequency (Absolute)	Frequency (Relative)	Replay Cost Factor
Pawel	53	19.85%	2
Mirek	49	18.35%	3
Maciek	48	17.98%	2
Andrzej	47	17.60%	3
Marek	42	15.73%	3
Dominik	28	10.49%	2

Details of the actors involved in the process of the cash loan are shown in Figure 9. Process discovery in BPM allows us to estimate the total commitment of each resource and service time. The identification and analysis of similarities and differences in a large amount of data have been addressed by involving RapidMiner with ProM extension [13,26]. This diagram has been used widely in different areas to support the identification and analysis of a large amount of data.

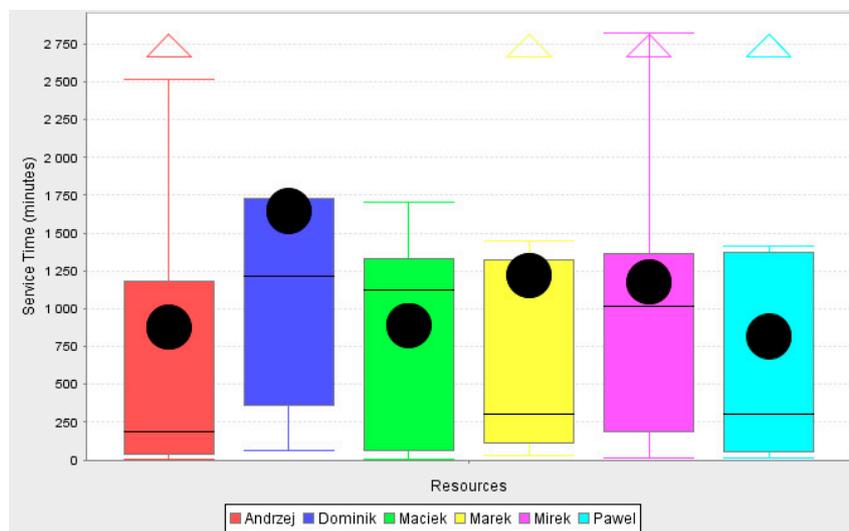


Figure 9. Resource service time statistics.

7. Conclusions

We have made a preliminary assessment of the feasibility of using our approach with a generated log file. We made a comparison of the graphical result of our approach to the cash loan sales process. In this article, we do not intend to evaluate the usefulness of the graphical results but instead focus on demonstrating the strength of our artifact, which can create visualizations that expose more aspects of BPM social networks and process-aware information systems (PAIS). Designing a proper sequence of use cases, as well as identifying rules to automate the flows between them, is a complex task that requires a lot of experience. The key to orchestrating use cases is the sequence in which they occur.

We have demonstrated that PAISs are systems dedicated to process management that function in many areas of human activity. There is no perfect instrument to monitor their work and analyze the processes they perform. To this end, one of the encouraging methods, process mining, can be used. Process mining is a discipline combining data mining and process modeling techniques. Process exploration offers automated techniques for detecting process models from event logs, verifying the suitability of process models and event logs (compatibility checking) and improving discovered processes with new data. The provided analysis of the discovery process can be used to validate the analysis in BPM and that the software fulfills the requirements, and an experiential study in an appropriate business environment could be used. It can also be used as a utility to give an uninterrupted view of the actors' behavior and allow for online improvement.

As our contribution, we consider the proposal for the construction of software systems based on the omnichannel business model (OBM) as well as a process mining engine (PME), a module which enables discovery of process models and provides their verification and expansion on the basis of data from event logs which describe real, not hypothetical, business processes. We showed an example of cash loan process optimization the functionality of PME component of the OBM architecture presented by us. The PME can be successfully used in software systems in case of having data from event logs which describe real, not hypothetical, business processes. Visualizations of the results show the different options for solving the problem clearly and legibly. The obtained results allow identification of bottlenecks and eliminate unnecessary operations that affect the efficiency of the process (e.g., cash

loan sales). These solutions and the example of implementation fill the existing gap between the rich literature on mining techniques and commercial systems and tools using discovery algorithms.

Further research will try to confirm experimentally correctness of assumptions taken, using big enough samples of business cases processed in the discussed system. Additionally, a significant problem to be faced in the future is the issue of the quality of business process event logs determined based on incorrect and missing values. Further research will focus on detecting abnormal and reconstructing missing values. We also plan to focus on how to represent procedural knowledge growing in the organization, making it useful for process automation.

Author Contributions: Conceptualization, P.D. and M.K.; methodology, P.D., M.K., and M.M.; software—formal analysis and investigation, P.D., M.K., and M.M.; resources, P.D., M.K., and M.M.; writing—original draft preparation, P.D., M.K., and M.M.; writing—review and editing, P.D. and M.M.; visualization, P.D., M.K., and M.M.; supervision, P. D.; project administration, P.D. and M.M.

Funding: This project is financed by the Minister of Science and Higher Education of the Republic of Poland within the "Regional Initiative of Excellence" program for years 2019 – 2022. Project number 027/RID/2018/19, amount granted 11 999 900 PLN.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meidan, A.; Garcia-Garcia, J.A.; Escalona, M.J.; Ramos, I. A survey on business processes management suites. *Comput. Stand. Interfaces* **2017**, *51*, 71–86. [[CrossRef](#)]
2. Wiśniewski, P.; Kluza, K.; Ligeza, A. An approach to participatory business process modeling: BPMN model generation using constraint programming and graph composition. *Appl. Sci.* **2018**, *8*, 1428. [[CrossRef](#)]
3. Wert, A.; Schulz, H.; Heger, C. AIM: Adaptable Instrumentation and Monitoring for automated software performance analysis. In Proceedings of the 10th International Workshop on Automation of Software Test, Florence, Italy, 16–24 May 2015; pp. 38–42.
4. Van der Aalst, W.M.P. *Process Mining: Data Science in Action*; Springer: Berlin/Heidelberg, Germany, 2016.
5. Tsakalidis, G.; Vergidis, K.; Kougka, G.; Gounaris, A. Eligibility of BPMN models for business process redesign. *Information* **2019**, *10*, 225. [[CrossRef](#)]
6. Hoang, H.H.; Jung, J.J.; Tran, C.P. Ontology-based approaches for cross-enterprise collaboration: A literature review on semantic business process management. *Enterp. Inf. Syst.* **2014**, *8*, 648–664. [[CrossRef](#)]
7. De Medeiros, A.K.A.; Guzzo, A.; Greco, G.; Van Der Aalst, W.M.; Weijters, A.J.M.M.; Van Dongen, B.F.; Saccà, D. Process mining based on clustering: A quest for precision. In *International Conference on Business Process Management*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 17–29.
8. Dymora, P.; Mazurek, M. Network anomaly detection based on the statistical self-similarity factor. In *Analysis and Simulation of Electrical and Computer Systems*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 324, pp. 271–287.
9. Motahari-Nezhad, H.R.; Swenson, K. Adaptive case management: overview and research challenges. In Proceedings of the IEEE International Conference on Business Informatics, Vienna, Austria, 15–18 July 2013.
10. Dos Santos Garcia, C.; Meincheim, A.; Junior, E.R.F.; Dallagassa, M.R.; Sato, D.M.V.; Carvalho, D.R.; Santos, E.A.P.; Scalabrin, E.E. Process mining techniques and applications—A systematic mapping study. *Expert Syst. Appl.* **2019**, *133*, 260–295. [[CrossRef](#)]
11. Maita, A.R.C.; Martins, L.C.; Lopez Paz, C.R.; Rafferty, L.; Hung, P.C.; Peres, S.M.; Fantinato, M. A systematic mapping study of process mining. *Enterp. Inf. Syst.* **2018**, *12*, 505–549. [[CrossRef](#)]
12. Van der Aalst, W.M.P. Process discovery from event data: Relating models and logs through abstractions. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1244. [[CrossRef](#)]
13. Koryl, M.; Mazur, M. Towards emergence phenomenon in business process management. *Arch. Control Sci.* **2017**, *27*, 263–277. [[CrossRef](#)]
14. Nguyen, H.T.C.; Lee, S.; Kim, J.; Ko, J.; Comuzzi, M. Autoencoders for improving quality of process event logs. *Expert Syst. Appl.* **2019**, *131*, 132–147. [[CrossRef](#)]
15. Hauder, M.; Pigat, S.; Matthes, F. Research challenges in adaptive case management: A literature review. In Proceedings of the 2014 IEEE 18th International Enterprise Distributed Object Computing Conference Workshops and Demonstrations, Ulm, Germany, 1–2 September 2014; pp. 98–107.

16. Motahari-Nezhad, H.R.; Bartolini, C.; Graupner, S.; Spence, S. Adaptive case management in the social enterprise. In *International Conference on Service-Oriented Computing*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 550–557.
17. Lantow, B. Adaptive case management—A review of method support. In *IFIP Working Conference on The Practice of Enterprise Modeling*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 157–171.
18. Moore, C. *The Process-Driven Business Of 2020*; Future Strategies Inc.: Lighthouse Point, FL, USA, 2012.
19. Abdi, N.; Zarei, B.; Vaisy, J.; Parvin, B. Innovation models and business process redesign. *Int. Bus. Manag.* **2011**, *3*, 147–152.
20. AbdEllatif, M.; Farhan, M.S.; Shehata, N.S. Overcoming business process reengineering obstacles using ontology-based knowledge map methodology. *Future Comput. Inform. J.* **2018**, *3*, 7–28. [[CrossRef](#)]
21. Halaska, M.; Sperka, R. Is there a need for agent-based modelling and simulation in business process management? *Organizacija* **2018**, *51*, 255–269. [[CrossRef](#)]
22. Liu, X.; Jia, W.; Wang, Y.; Guo, H.; Ren, Y.; Li, Z. Knowledge discovery and semantic learning in the framework of axiomatic fuzzy set theory. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1268. [[CrossRef](#)]
23. Banziger, R.B.; Basukoski, A.; Chausalet, T. Discovering business processes in CRM systems by leveraging unstructured text data. In *Proceedings of the IEEE 20th International Conference on High Performance Computing and Communications/IEEE 16th International Conference on Smart City/IEEE 4th International Conference on Data Science and Systems (HPCC/SMARTCITY/DSS)*, Exeter, UK, 28–30 June 2018; pp. 1571–1577.
24. Burattin, A. Heuristics Miner for Time Interval. In *Process Mining Techniques in Business Environments*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 207, pp. 85–95.
25. Dumas, M.; La Rosa, M.; Mendling, J.; Reijers, H. *Fundamentals of Business Process Management*; Springer: Berlin/Heidelberg, Germany, 2013.
26. Mans, R.S.; van der Aalst, W.M.P.; Verbeek, H.M.W. Supporting process mining workflows with RapidProM. In *BPM Demo Sessions 2014*; Limonad, L., Weber, B., Eds.; CEUR Workshop Proceedings: Eindhoven, The Netherlands, 2014; Volume 1295, pp. 56–60.
27. ProM Tool 6.6–6.9. Available online: <http://www.promtools.org> (accessed on 25 May 2019).
28. Holland, J.H. *Emergence: From Chaos to Order*; Perseus Publishing: New York, NY, USA, 1999.
29. Grigori, D.; Casati, F.; Dayal, U.; Shan, M. Improving business process quality through exception understanding, prediction, and prevention. In *VLDB*; Morgan Kaufmann: Burlington, MA, USA, 2001; pp. 159–168.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).