Mining Self-Defined Business Process in Electronic Administration

Zineb Lamghari, LRIT, CNRST (URAC 29), Rabat IT Center, Faculty of Sciences, Mohammed V University, Morocco*

iD https://orcid.org/0000-0003-0916-4178

Rajaa Saidi, SI2M Laboratory, National Institute of Statistics and Applied Economics, Morocco

iD https://orcid.org/0000-0002-8292-3188

Maryam Radgui, National Institute of Statistics and Applied Economics, Morocco

iD https://orcid.org/0000-0002-8447-3686

Moulay Driss Rahmani, Faculty of Sciences, Mohammed V University, Morocco

ABSTRACT

The information retrieval system is a set of resources and tools that allow users to search for information in a given domain. This system permits users to perform their research according to their objectives in diverse ways producing different behaviors. Even users with the same objective may follow different paths and stand different sub-processes, which are introduced as self-defined business processes that vary in terms of structure, objective, and result. This puts forward the difficulty of obtaining and studying these user's behaviors. This paper targets the problem of representing and managing self-defined business process variability. A special interest is given to the use of process mining to deal with this challenge. In this regard, a case study about citizens in interaction with the electronic administration is presented to discover and manage variability of this process type. The main result is a set of recommendations to end users.

KEYWORDS

Configurable Process Model, Decision Miner, Fuzzy Miner, Inductive Miner, Information Retrieval System, Process Discovery, Process Mining, Semantic Reasoning, Variability

INTRODUCTION

Information Retrieval (IR) System (Naouar et al., 2017a; Naouar et al., 2017b) is a set of resources and tools that allow users to search for information in a given domain. These systems permit users to accomplish their research according to their objectives. In this sense, processes provided by the IR systems allow users often to determine their own procedure. In such context, the user's manner to perform a task (purchasing a product from an e-commerce website, searching for a document in a digital library, etc.) represents a "self-defined process" (Luengo & Sepúlveda, 2011). Among these systems that produce the self-defined BP type, there is the case of digital libraries (Shiri, 2018), e-commerce websites (Laudon & Traver, 2016), cyber-physical systems (Seiger et al., 2019), Electronic administration services (Kasprzyk, 2018), and others.

Self-defined BPs are considered as a special category of BPs, with high variability level (Cole, 2015; Dinh & Tamine, 2015). This category of BPs represents user's behaviors that may achieve one objective in diverse ways to perform a research. Indeed, users in information-seeking situations adopt behaviors that depend on many factors (Naouar et al., 2017b; Ruso et al., 2013; Laudon et al., 2016), that can change user's processes. These factors are: 1. User's objective. 2. User's Requirement and 3. Engine's Knowledge used to search the information.

Therefore, the difficulty of representing self-defined BP emerges from its variability (Athukorala et al., 2015; Luengo & Sepúlveda, 2011). The later changes according to different contexts and requirement. Even though managing process variability is a non-trivial task because it requires specific standards, methods, and technologies, it still involves many parameters that are not always formally defined. For example, designing the reference process model, which represent the commonalities from the process family, is a challenge, as well as the necessary adjustments to configure a specific process variant.

To overcome these challenges, it would be useful to represent users' behavior (information-seeking processes), i.e., to define the generic process model, in order to study the self-defined BP variability and recommend the suitable path to each user. Also, it is useful to manage the process variants through ontologies based on semantic reasoning and Configurable Process Model (CPM) (Gottschalk et al., 2007), i.e., to select the appropriate process variant according to the combination between different self-defined BP ontologies.

To achieve these objectives, this study uses process mining (Van der Aalst, 2016) algorithms to mine user self-defined BPs. Seven process mining algorithms are applied on two real datasets of users' traces interactions with the e-administration. The first step is to select the most preferment algorithm, based on process model quality criteria (Buijs et al., 2012; Günther & Van der Aalst, 2007), to discover the generic process model. The second step aims at managing existing process variants to recommend the suitable path according to user's objective, requirements, and engine knowledge. This is done by using decision miner algorithm to obtain the CPM (process model with variants details) and by employing related semantic reasoning of the self-defined BP ontologies.

Mainly, the contribution presented in this paper is about mining self-defined BPs in terms of representing users' behavior in their information-seeking processes and capturing variants paths for managing its variability. The novelty, here, consists of using techniques to treat the whole entire information retrieval process (a complete process with a beginning and an end activity). The techniques differ from others employed in Nouvellet et al. (2017), which have analyzed data related to user sessions. It is limited to the deduction of a model based on a Markov chain of order 1. Nevertheless, the question of the capacity of these process mining techniques to analyze self-defined processes and to determine their structure and variations (Luengo & Sepúlveda, 2011) is not previously tackled. In addition, in the literature, there is a lack of research works, in self-defined BP context, treating scenarios in which process mining techniques could be combined with ontologies based on semantic reasoning, for providing decision choice analysis.

Therefore, the paper is organized in order to evaluate the performance of process mining algorithms in representing self-defined processes and their ability to generate user's behaviors and identify variation points, alternatives, and rules. Besides, it illustrates the applicability of semantic reasoning as a decision task that can be combined with process mining, to recommend one path instead another, in the case of interactions between users and e-administration services.

Concerning section 2, it provides the required background knowledge. Section 3 introduces the methodology to mine and manage variability of self-defined BP. In Section 4, a concrete case study is depicted to simulate the proposed approach. The case study is about interactions between users and electronic services of e-administration with self-provided processes, where citizens' service requests can proceed through self-defined processes. Section 5 discusses the obtained results, the related works, and the implication of practice examples. Section 6 summarizes the paper and introduces future research.

Theoretical Background

The contribution presented in this paper is about mining self-defined BPs. Indeed, a suitable technique for representing users' behavior in their information seeking processes is required. In this sense, process mining seems to be an interesting option to deal with. Also, managing self-defined BP variability is essential (CPM). Furthermore, it is important to handle semantic content (Ontologies based on semantic reasoning) to recommend one process in the case of variation points.

Process Mining

Process mining is a scientific discipline that falls between data mining and business process analysis (Van der Aalst, 2016). The main idea of process mining is to use the execution BPs event logs that are recorded in the information system, to automatically generate, check and enhance process models. An event log is a collection of traces, which represent process instances that have been executed. Each event in an event log is assigned to an activity executed for a singular process instance (one trace). For each trace, all events belonging to that case are ordered in a chronological style.

Furthermore, process mining consists of three types (Cf. Figure 1), which are: discovery, conformance, and enhancement. Discovery: An automatic process modeling methodology that takes event logs as input and produces a BP model as output. Conformance: compares the newly discovered process model with the existing process model. The purpose is to detect deviations and identify bottlenecks. Enhancement: focuses on improving or extending the existing process model using the information recorded in event logs.

Figure 1. *Overview of process mining and its three types of techniques (Van der Aalst, 2016*
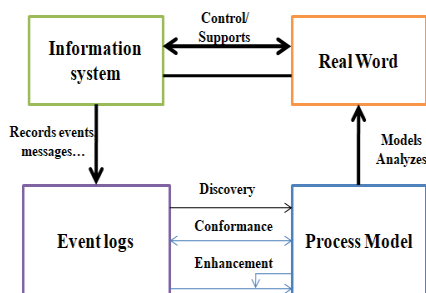
Configurable Process Model (CPM)

CPM (Gottschalk et al., 2007) appeared with the objective of integrating different process variants into one model. Thus, the CPM enables extracting a process variant, which is a process model different from the original one, but that fits better in the application environment. This approach enables to represent the commonalities of the process variants. By sharing the particularities among multiple variants, this approach also promotes the model reuse (Ayora et al., 2012; Radgui et al., 2013).

Several aspects related to the BP variability have been discussed, such as: management, (re) design (Kumar & Yao, 2012), modeling (La Rosa et al., 2009), and configuration (Ghoula et al., 2007). Furthermore, most of the proposed approaches present a low level of automation. Indeed, the configuration of the process variant requires the verification of a syntactical and semantically levels of resulted models, where existing approaches do not differ between planned execution and real process execution, i.e., what happens during the process execution may be not planned to happen. Therefore,

the use of process mining techniques is mandatory because its enable the extraction of information from event logs (Gottschalk et al., 2007). Thus, by analyzing the generated process model, process variants can be discovered, and problems can be corrected.

For this purpose, a process mining technique called, Decision Miner, is selected to analyze decision points that enable detailing variation points, alternatives and rules.

The benefits provided by the semantic enrichment of the BP include the improvement of its representation and understanding; the automation of tasks related to the modeling, configuration, evolution, and the adaptability of the BP according to different requirements. Therefore, it is possible to analyze the CPM in a semantic manner.

Ontologies and semantic reasoning

The ontology enables to capture, represent, re (use), share and exchange common understanding in a domain (Bogarin et al., 2018). The ontology is composed by commonly agreed terms, thus describing the domain of interest. However, knowledge shared and reused among applications and agents are only possible through the semantic annotation.

Semantic annotation enables to reasoning over the ontology, thus ensuring the quality of the ontology by deducting new knowledge (Liao et al., 2015; Staab & Studer, 2013).

The semantic enrichment of the BP was proposed to increase the level of BPM lifecycle (Hepp & Roman, 2007) and to compliance checking (Szabó & Varga, 2014). Regarding to the CPM, semantic technologies have been applied for semantic enrichment (El Faquih et al., 2014) and for semantic validation (Fei & Meskens, 2008).

Methodology

Self-defined BPs are characterized by variability. This puts forward the difficulty of representing this BP type. Therefore, the authors propose an approach for representing and managing self-defined BP. Indeed, the applicability of process mining and semantic ontologies, to represent users' behavior in their information-seeking process is required. In this context, the mining self-defined BP approach consists of three main steps (Cf. Figure 2).

Figure 2. Approach for mining self-defined BP (Authors' work)
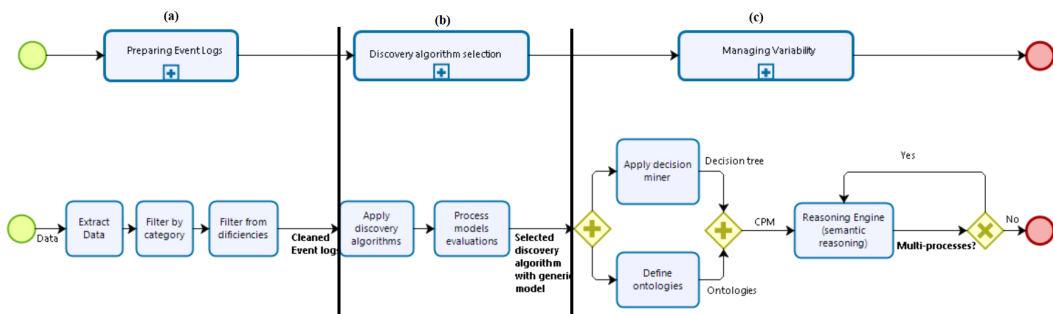
The self-defined BP approach starts by the event logs preparation step (part **a** of figure 2). In this step, it is important to pass through the extraction and the filtering operations, to obtain cleaned event data. In this sense, the authors focus on two main filters: The category filtering (Rafiei, 2018; Lamghari et al., 2019) and the deficiencies filtering (Cheng & Kumar, 2015; Tax et al., 2019). The

first ignores data with forbidden access, while the second ignores noise (ignores data incorrectly logged) and chaotic (ignores data logged by the system even though it is not part of the main process flow) behaviors. Accordingly, the authors can proceed to the discovery algorithm selection step (part **b** of figure 2). This step aims at evaluating the capacity of each process discovery algorithm for representing self-defined processes. This evaluation can be released based on process models quality criteria. This phase's output can be used in the managing variability step (part **c** of figure 2). To do so, the authors define ontologies and apply the decision miner algorithm. These two operations are required for semantic reasoning. At this stage, it is crucial to detail paths variants (CPM) using events gathering knowledge and ontologies, to determine the cluster of paths (processes with the same user objectives and requirements) on which the refinement activity can be released to choose one path instead another. This is done by the reasoning engine activity that can combine the resulted CPM with the level of knowledge ontology (semantic reasoning). This recommends a unique process for each user according to the inputs' data. Thus, the unicity of the resulted process is checked. In the case of multi-processes result, the reasoning step restarts. All these approach steps will be detailed in the following sections.

Discovery Algorithms and Process Model Quality Criteria

A process discovery algorithm constructs a generic process model based on event logs. Indeed, the generic model is an abstracted and general representation of real event logs. Several discovery algorithms are described with basic representation of process models, like alpha algorithm. Other algorithms are representing different abstraction levels combined with clustering and classification techniques, to model processes from unstructured and complex events. In this regard, the authors conduct a comparative study to nominate the most preferment discovery algorithm to represent self-defined processes. In this regard, the authors are inspired by (Augusto et al., 2018; Mans et al., 2008; Pegoraro & Van der Aalst, 2019) to list the following process discovery algorithms: Alpha ++ algorithm, Heuristic Miner (HM), Inductive Miner (IM), Genetic Miner (GM), Fuzzy Miner (FM), State Based Regions (SBR), Language Based Regions (LBR).

On one hand, the Alpha ++ detects non-free choice relation by describing activities of the selected relation that depends on other activities (Wen et al., 2007). It cannot detect invisible tasks. Therefore, this algorithm gives unsound results. In this sense, an extended version of the alpha algorithm has been created, to take into consideration the patterns' frequency. Indeed, the HM algorithm (Weijters & van der Aalst, 2003) can discover main behaviors and abstract exceptional and noisy ones leaving out less important activities. This later cannot group traces with sub-logs representation. Accordingly, the IM algorithm has been developed to treat events by grouping them into sub-logs. For each sub-log, a sub-process is generated. Then, a combination between the resulted sub-processes are released to obtain the generic process model. In this respect, the IM algorithm produces sound models (Bogarin et al., 2018), i.e., less none-conformities detected and it fits with the majority of present logs. Besides, it cannot identify complex and non-local process control patterns.

On the other hand, new algorithms have been developed to treat event logs in their uncertainty, for example the GM algorithm. This algorithm uses the genetic concept in creating process models from logs. This is done randomly. For each process, the precision metric is calculated. Then, sound models are combined based on the mutation operation. The main limitation of this approach is their complexity in discovering and representing process models from real data sets (Vanden Broucke & Weerdt, 2017). From the same complexity standpoint, the FM deals with unstructured processes (Günther & Van der Aalst, 2007). In this sense, FM simplifies unstructured processes by preserving significant behavior, while less significant but highly correlated behaviors are aggregated into clusters, and less significant or less correlated behaviors are abstracted.

Furthermore, the SBR algorithm generates a Petri net from a Transitions System (ST) based on specific abstractions, such as: Set, Multi-Set, Sequence and other types of abstractions, in which each state of the ST can be represented by a complete or partial trace. This algorithm ensures the fitness

metric, as well as the identification of complex control structures. Besides, SBR is unable to process incomplete and noisy logs (Van der Werf et al., 2008), while the LBR algorithm can find process model places based on the language process. Indeed, the LBR algorithm uses properties derived from logs (causal relationships), to determine the final model by describing different places. Unfortunately, this algorithm is unable to process incomplete and noisy logs (Van der Aalst et al., 2010).

In summary, obtaining a quality model is the main goal of a process discovery algorithm. There are various metrics and approaches for estimating process model quality criteria (Günther & Van der Aalst, 2007; Buijs et al., 2012), which are:

- **Fitness**: This metric quantifies how much the observed behavior captured by the model. i.e., it quantifies the ðt of a log in a model.
- **Generalization**: The model should generalize the present behavior in the log. This metric quantifies how well the model explains unobserved behaviors. The main difficulty with the generalization metric is the need of unobserved behaviors treatment.
- **Precision**: The model ignores unrelated behavior that is stored in the log. This metric quantifies how much behavior exists in the model that are not observed. A high level of generalization model could represent much more behavior than once presented in the log (underfitting model vs overfitting model).
- **Simplicity**: The model should be represented in a simplified structure. This metric quantifies the model complexity, and it is not treated in this paper.

For the fuzzy miner algorithm, the output model is a fuzzy model. To evaluate the fuzzy model, two metrics are available: Node detail and conformance.

- The **Node Detail** describes activities displayed in the Fuzzy model, related to the aggregated or deleted activities. Nodes of visible activities are called explicit nodes, while nodes corresponding to an activity are denoted as implicit ones.
- **Compliance** is a measure that describes the alignment between the Fuzzy model F and the logs T. Each activity in the logs that does not exist in the Fuzzy model will be counted as a deviation.

In table 1, the authors classify process mining discovery algorithms, according to quality criteria. This is based on the logic of the final representation, i.e., Algorithms producing Petri net models are suitable with fitness, generalization, and precision metrics, while algorithms producing fuzzy models are adequately evaluated using the node detail and the conformance metrics.

In this regard, this sub-section paves the way for a comparative study that can demonstrate the process discovery algorithms efficiencies and limitations, related to self-defined BP, where process model quality values can determine the suitable algorithm to represent self-defined processes. This is can be achieved by comparing the quality values of each resulted process model.

Managing Variability

To overcome the variability challenge, this work manages the process variants through ontologies based on semantic reasoning (Cf. Figure 3). Th authors propose to select the appropriate process variant according to the user's objective, the user's requirements, and the engine level of knowledge. To this end, the semantic annotation aims at reasoning over ontologies, to recommend final decision to each user.

Analyzing event logs by process mining techniques provides all process instances. The instances properties can be used in the process model generation. At this stage, it is important to detail the generic model into a CPM emerged with the objective of integrating different process variants into one model.

Table 1. Process model quality criteria according to discovery algorithms (Authors' work)

Produced Models	Quality Metric	Formulas	Definition
HM, IM, GM, LBR, SBR and Alpha++	Fitness	$$\text{Fitness}\left(L, M_2\right) = 1 - \frac{\delta\left(\lambda_{\text{opt}}^{M_2}\left(L\right)\right)}{\delta\left(\lambda_{\text{worst}}^{M_2}\left(L\right)\right)}$$ With δ is the cost function, $\lambda_{\text{opt}}^{M_2}\left(L\right)$ is the worst case where there is no possible synchronization between the trace and the model. $\lambda_{\text{opt}}^{M_2}\left(L\right)$ represents the obtained costs for each optimal alignment.	Quantifies how much behavior captures the model.
HM, IM, GM, LBR, SBR and Alpha++	Precision	$$\text{Precision}\left(L, M\right) = \frac{1}{|E|} \sum_{e \epsilon E} \frac{\text{en}_T\left(e\right)}{\text{en}_M\left(e\right)}$$ With E is the set of events in the T logs, A is the set of activities, $\text{en}_T\left(e\right) \subseteq A$ is the set of activities presented in the traces and in $M(e) \subseteq A$ the set of activities presented in the model.	Quantifies how much behavior exists in the model that was not observed.
HM, IM, GM, LBR, SBR and Alpha++	Generalization		Quantifies how well the model explains observed system behavior.
HM, IM, GM, LBR, SBR, Alpha++	Simplicity		Verifies the simple structure of the process model
Fuzzy Miner	Node detail	$$\text{Node Detail} = \frac{\sum v \in V^{s(v)}}{\sum n \in V^{s(n)}}$$ Where N corresponds to all the primitive nodes (present, aggregated or deleted activities) of the Fuzzy model F, $V \subseteq N$ is the subset of all visible nodes and s is a function that associates each primitive node F with a measure of importance.	Quantifies activities displayed in the Fuzzy model, relatively to the aggregated or deleted activities
Fuzzy Miner	Conformance	$$\text{Conformance} = \frac{M\left(T\right) - d + 1}{M\left(T\right) + 1}$$ With M the total number of activities present in the logs T and d is the number of deviations identified.	Describes the alignment between the Fuzzy model F and the logs T

The CPM can be obtained by applying the decision miner algorithm. In this sense, decision point analysis allows identifying variation points (parts of the model that are subjects to variation), alternatives (available for the variation points), and rules enabling to choose one path instead another (Ayora, 2012). By identifying these aspects, the process variants can be extracted.

From a technical standpoint, the user provides information related to daily objectives and requirements. Indeed, the questionnaire-model approach (La Rosa et al., 2009) is applied to guide the configuration process. Here, each variation point is associated to a question, whose alternatives determine the path selection. Thus, by selecting an alternative related to a question, the system configures a process variant. In the proposed approach, the questionnaire is developed using the decision point analysis.

After discovering process variants, it is required to combine them with ontologies. Here, a focus is given on self-defined BP factors, to define ontologies, which are: 1. Objective ontology, 2. Requirements ontology and 3. Level of knowledge ontology. Respectively, variation points, available alternatives and rules related to citizen and e-administration interaction must be formalized. Therefore, process mining can enrich ontologies definition by extracting knowledge from event logs in order to recommend one variant to each user. For this purpose, the authors illustrate the step of combining CPM with ontologies through semantic annotation in figure 3.

In this context, figure 3 illustrates the proposed approach for managing variability into self-defined BP. This figure presents two aspects: the front-end aspect and the back-end aspect. The first aspect concerns the user's inputs, which are the user's objective and requirements. These two points will lead, later, to obtain the final decision process. The second aspect encompasses the analysis and the engine parts. The analysis part details the generic model and the CPM, based on historic data, ontologies, and the decision miner algorithm (Decision tree). Consequently, the self-defined BP ontologies, guidelines and process variants attributes are emerged. In this context, the use of semantic reasoning (Detro et al., 2017) is required, to recommend the suitable process that the user must be achieved, based on the engine knowledge, and learned from the user's objective and requirements.

In this way, when the user (Front-end) provides an information related to the user's objective and requirements, a cluster of activities is performed. In addition, the knowledge of the reasoning engine will be used to recommend the unique process to achieve.

For example, when the user asks for a document, usually the first operation is the evaluation of the document type. Thus, some required information must be performed, such as, identity, reason of the request, etc. The objective and requirements presented by the user determine multi-processes. Their refinement must be released with ontologies based on semantic reasoning. In this regard, the unique result must be enriched by the level of knowledge ontology (Reasoning Engine), according to these steps (a), (b), (c), (d) and (e) of figure 3.

Figure 3. Approach for managing self-defined business process variability (Authors' work)
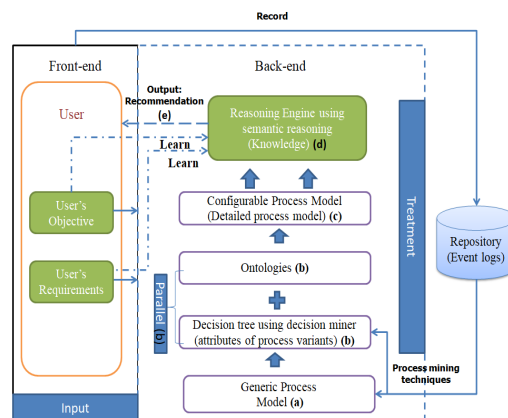
Case Study: Analysis and Results

In this section, a case study about citizens in interaction with the Electronic Administration is presented, to illustrate the applicability of the mining self-defined BP approach (Cf. Figure 2).

Preparing Event Logs

The definition of e-administration or electronic administration refers to any of the mechanisms, which are meant to transform what in a traditional office, based on paper processes, into a paperless office based on electronic ones. This is an ICT (Information and Communications Technology) tool intended to improve productivity and performance.

Administration processes are characterized by the fact that several organizational units (community, municipalities, etc.) can be involved in the process treatment of citizens (Ruso et al., 2013). These organizational units often have their own specific IT applications; it becomes clear that getting data, which is related to e-administration processes, is not an easy task because the latter are characterized by variability. Indeed, this sub-section aims at extracting and filtering events, in order to prepare data (part **a** of figure 2) for next steps.

The process between citizen and government is a particular example of electronic administration. In this sense, figure 4 illustrates interaction process between the citizen and the government that can be partially predefined since this interaction between them depends on specific objective, level of knowledge and requirements. This can vary the process and produce different user behaviors that cannot be predefined. These processes need advanced analysis to understand user's behavior in their interaction with the e-administration services.

As shown in figure 4, the e-administration is a system that allows navigating through large volumes of administrative documents created in interaction with other parties. These documents are often heterogeneous. The information search activity carried out by users is a process (Cole et al., 2015) that relies mainly on search engine queries, filters and consulted documents. These elements depend on users' requirements, types of information handling strategies, as well as the way how data are indexed. Indeed, the use of a user-guided system, in relation to electronic administration, produces many processes with different variations and self-identification. These two characteristics describe self-defined BPs.

Generally, as mentioned in (Hai, 2007), electronic services related to e-administration are defined in three categories: public, voluntary and private. The private events are with forbidden access (accessible only by the data owner) like services between the citizen and the police entity. The public

Figure 4. Interactions between e-administration and citizens (example of e-government) (Authors' work)
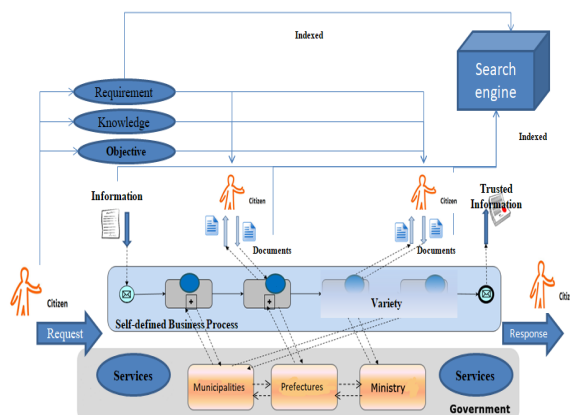
and voluntary services are accessible to the data analyst. Therefore, this case study focuses on two event logs categories: public and voluntary.

According to the data volume, some filters must be applied to obtain clean and specific data (Cf. Table 2). Indeed, the authors use two filters: The first filter aims at ignoring forbidden (Rafiei, 2018; Lamghari et al., 2019) noise and chaotic data. The second filter aims at filtering events on specific period. A focus is given to the period defined by behaviors with high level of variability. This level distinguishes self-defined BPs structure (Cf. Figure 5).

Table 2. An example of event logs

Case_ID	User	Date	Activity	Category
1	Citizen1	2018-05-17|14:23:25	Trustprofil	Public
2	Citizen2	2018-05-17|14:24:25	Trustprofil	Public
1	Citizen1	2018-05-17|14:24:26	Scroll	Public
1	Citizen1	2018-05-17|14:24:28	ResourceAccess	Public
3	Citizen3	2018-05-17|14:26:26	request	Voluntary

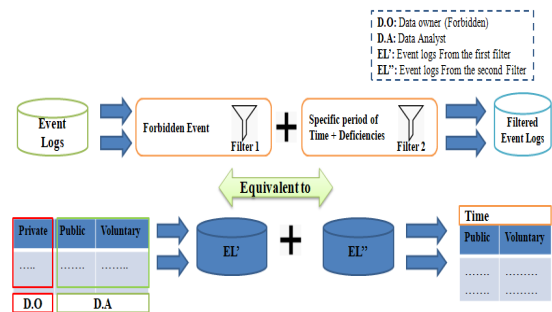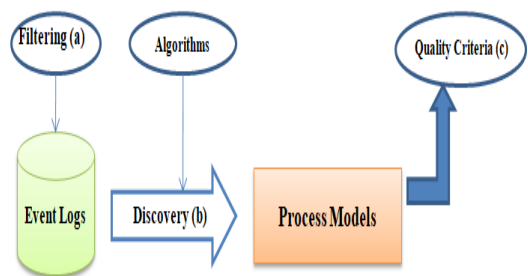**Figure 5. Event logs preparation (Source: Authors' work)**

Figure 6. The Processing steps (Source: Authors)

In this work, the authors use raw data collected by the IR of electronic services. The data contain information about a group of interaction between citizens and the e-administration (municipalities, prefectures, and ministries) treated in 2018, for which all diagnostic and treatment activities have been recorded. The data distinguish two events logs categories that encompass a set of activities for both public and voluntary categories. The public category contains 110 traces, 1755 events, while the voluntary category contains 39 traces, 1570 events.

Indeed, the logs in the public category contain 110 process instances (traces) of 80 variants. 10 traces appear more than once while 70 are unique. In the logs of the voluntary category, there are 39 unique traces. This is an important property of self-defined processes, where recurrent traces are rare and not very redundant. Process discovery algorithms have the advantage of finding generic models from non-redundant traces. Another important aspect of user self-defined process is the repetition of an activity within the same trace. Therefore, long traces can be generated.

## Discovery Algorithm Selection and Process Evaluation

After the data preparation step (part **a** of figure 6), the authors proceed to the process discovery algorithms application (part **b** of figure 6), in order to select the suitable algorithm for the self-defined BP representation. Therefore, in this sub-section, the authors compare and evaluate each resulted process model with the aforementioned quality criteria (part **c** of Figure 6), to deduct the most generic, representative, and performing one, by necessity defining the suitable discovery algorithm related to the e-administration citizen's behaviors (part **b** of Figure 2).

Table 3. Fitness, Precision and Generalization for all models

	Public			Voluntary		
	Fitness	Precision	Generalization	Fitness	Precision	Generalization
HM	0.00	0.00	0.00	0.00	0.00	0.00
IM	0.98	0.24	0.99	0.93	0.15	0.99
GM	0.99	0.18	0.99	0.62	0.80	0.99
LBR	0.62	0.38	0.97	0.78	0.19	0.96
SBR	0.90	0.42	0.99	0.96	0.29	0.99
Alpha++	0.00	0.00	0.00	0.00	0.00	0.00
FM						

Source: Authors

Figure 7. Heuristic Miner algorithm applied on the voluntary category (Source: Authors' work)
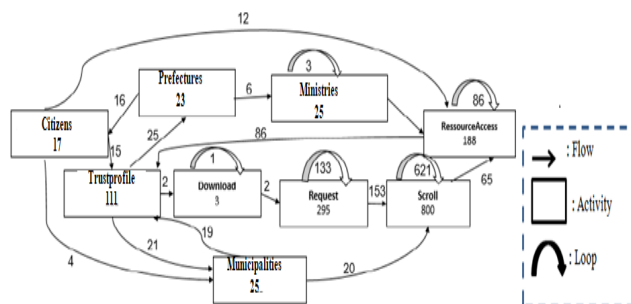
To discover the generic process model and to evaluate its quality, the authors use the ProM tool (Van der Aalst et al. 2009). In this sense, the fitness is calculated using the "PNetReplayer" package, while the Precision and the Generalization metrics are calculated using the "PNetAlignementAnalysis" package.

All discovered models, using HM, IM, GM, LBR, SBR and Alpha++ algorithms, must be transformed into a Petri net form, to check model's conformance. The simplicity of these process models is not checked in this paper. To this end, Petri net representations are evaluated with three process model quality metrics: fitness, precision, and generalization, while the fuzzy miner representation based on two metrics: Node detail and conformance.

The Petri net discovered with the Alpha++ algorithm is an unsound model because it contains deadlocks and insignificant traces (few traces are generated and none of them correspond to those presented in the logs). The same limitations are observed for the models generated by the HM algorithm

Figure 8. Inductive Miner algorithm applied on the public category (Source: Authors' work)
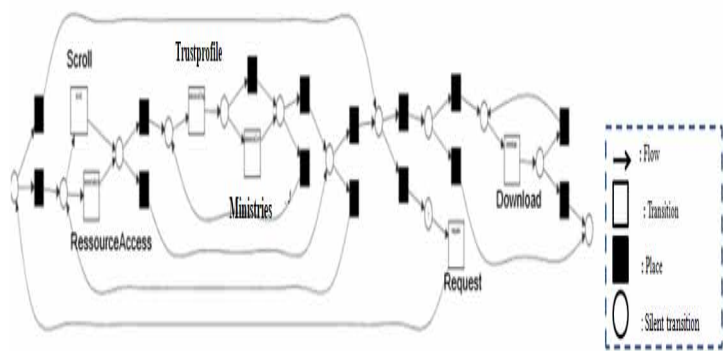

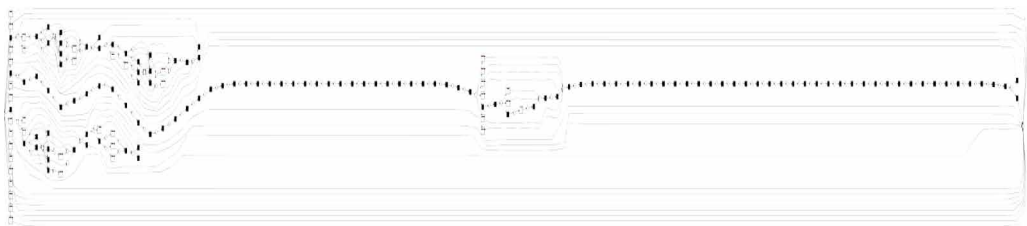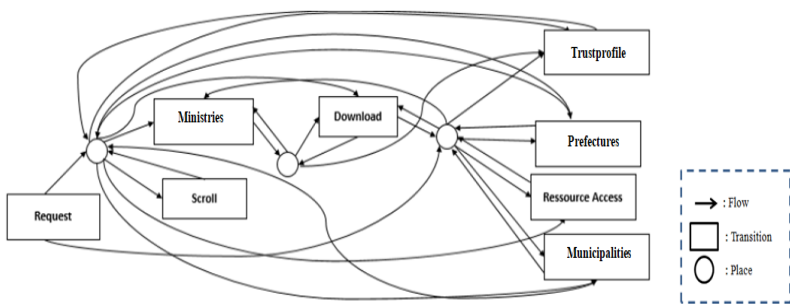Figure 9. Generic Miner algorithm applied on the public category (Source: Authors' work)

Figure 10. LBR algorithm applied on the public category (Source: Authors' work)

(Cf. Figure 7 and Cf. Table 3). The heuristic model has no initial and final marking, i.e., the model has no departure or arrival points, but it is readable. However, it helps understanding links between different stages of the studied process.

Figure 8 illustrates the process model discovered from the public logs, using the inductive miner algorithm. In this respect, the fitness value is excellent on both public and voluntary logs (more than 0.90). In addition, the obtained models have a high Generalization and a low precision (Cf. Table 3). These models can detect the process start and end points. They also ignore unsignificant arcs. These arcs show different users processes to accomplish their information-seeking process. This inductive algorithm gives a very clear representation of the process that is followed by citizens using e-administration services.

The genetic miner algorithm discovers two models, with high level of generalization. The public process model, in Figure 9, covers more logs than the voluntary process model. On the other hand, the fitness for public logs reaches 0.99, while it is equal to 0.62 for the voluntary category. Consequently, the model discovered from the public logs are underfitting (precision=0.18) and equal to 0.80 (for the voluntary logs). Therefore, the genetic model has a high level of precision and generalization. The inconvenient of these models is the complex representation. They do not allow determining the most important links between different activities. Thus, different user's processes, to achieve their objectives, cannot be determined. The resulted model is unreadable.

The LBR models are similar to the inductive models, in terms of the absence of frequency and different choices made by users. Also, these models are similar to the flower form i.e. all activities are accessible from other states. Moreover, it does not represent initial marking. For both categories, the LBR algorithm produces models with high generalization and fitness rates (both cases equal to 0.7). Therefore, the LBR model (Cf. Figure 10) provides an interesting generalization. The different activities are well represented and links between them are very detailed.

The obtained models with the SBR algorithm are discovered based on the Multiset abstraction option and the choice of partial traces. The two discovered Petri nets have initial marking. The SBR

Figure 11. Fuzzy Miner algorithm applied on the voluntary and the public categories (Source: Authors' work)

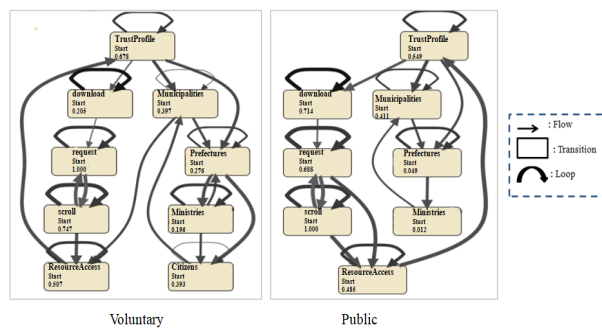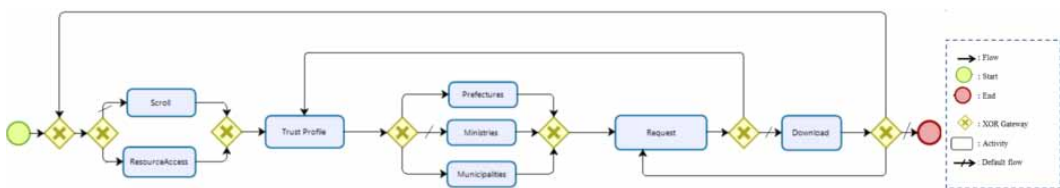Table 4. Measures for the fuzzy miner algorithm

	Node details	Conformance
Public	1.00	0.80
Voluntary	1.00	0.84

(Source: Authors' work)

algorithm also produces two models with a high level of generalization. In addition, replaying logs on the obtained models produces sound results for both public and voluntary events (Fitness=0.9). The SBR models allow visualizing, very precisely, all different users' choices, in order to navigate from one activity to another. However, the SBR models are failed to represent the suitable generalization metric.

The Fuzzy algorithm is applied on the public and voluntary event logs categories, using the Fuzzy Miner package. The Table 4 shows that the visible nodes, found in the two graphs, are important and significant. Moreover, replaying logs on obtained models gives sound results (conformance 0.8). Therefore, these models represent a clear and interesting generalization (Cf. Figure 11). The frequency of users' choices is well presented and provides useful overview of the studied processes.

Figure 12. The generic process model with BPMN representation (Source: Authors' work)

To conclude, the suitable process discovery algorithm to represent self-defined BPs is the Inductive Miner. Also, the Fuzzy Miner algorithm gives sound results. For more simplification, the generic model (public and voluntary categories) is converted to the Business Process Model Notation (BPMN) form. Thus, the process illustrates 8 activities. The user has two possible ways to consult the e-administration services: via scroll mode (search for information) or via the resource access mode (search for resource). In addition, the user can be connected using the trust profile (officially generated for each citizen). According to the trust profile (citizen ID and objective), the citizen can directly contact municipalities, prefectures, or ministries. Then, the citizen can apply for administrative certifications. If the request requires the contact of other resources, citizen must go back to the trust profile activity (the citizen's profile against the requested document: owner, participant, intermediate, etc.), to get a final decision. Next, the citizen can download the paper for a limited period. Last, the citizen can log out.

In this abstracted representation, four XOR gateways are observed, while the deep abstraction level generates more process variants (Cf. Figure 9). This puts forward the difficulty of managing self-defined BP variability.

Managing Variability Application

After defining the generic self-defined BP, it is important to meet the challenge of the BP variability. So in this sub-section, an application of the proposed approach to manage self-defined BP variability (part **c** of Figure 2) is presented.

In this context, process variant configurations through ontologies have been proposed by some authors, such as Huang et al. (2013). However, the approach proposed, in this research, enables to identify process variants, its characteristics and ontologies from event logs. In this way, the process model can be correctly individualized by meeting the requirements of the context application. Moreover, event logs reflect what is happening during the citizen and e-administration interaction and enable the process variants improvement.

To this end, figure 13 illustrates how the management approach could treat this variability. Indeed, a focus is given on the fragment of three XOR gateways, as decisions points' example of the generic process model (Cf. Figure 12).

First, the decision miner algorithm is applied on the generic model to obtain the decision tree of these gateways and to define user's objectives and requirements ontologies. Based on these two elements, process variants are obtained. For instance, a citizen request could be a demand for simple or complex documents. The simple request is immediately accepted, while the complex one requires the trust profile status verification (owner, intermediate or participant).

According to this status, e-administration services (ministries, prefectures, or municipalities) could respond. Side by side, objectives and requirements ontologies examples are defined. The objective ontology contains trust profile class (citizen ID), status (owner or participant or intermediate) and information as sub-classes (personal information). The requirements ontology encompasses Requests, Resources as classes and documents as sub-class.

Figure 13. selection of the appropriate process variant according to the user's objective, the user's requirements, and the engine level of knowledge
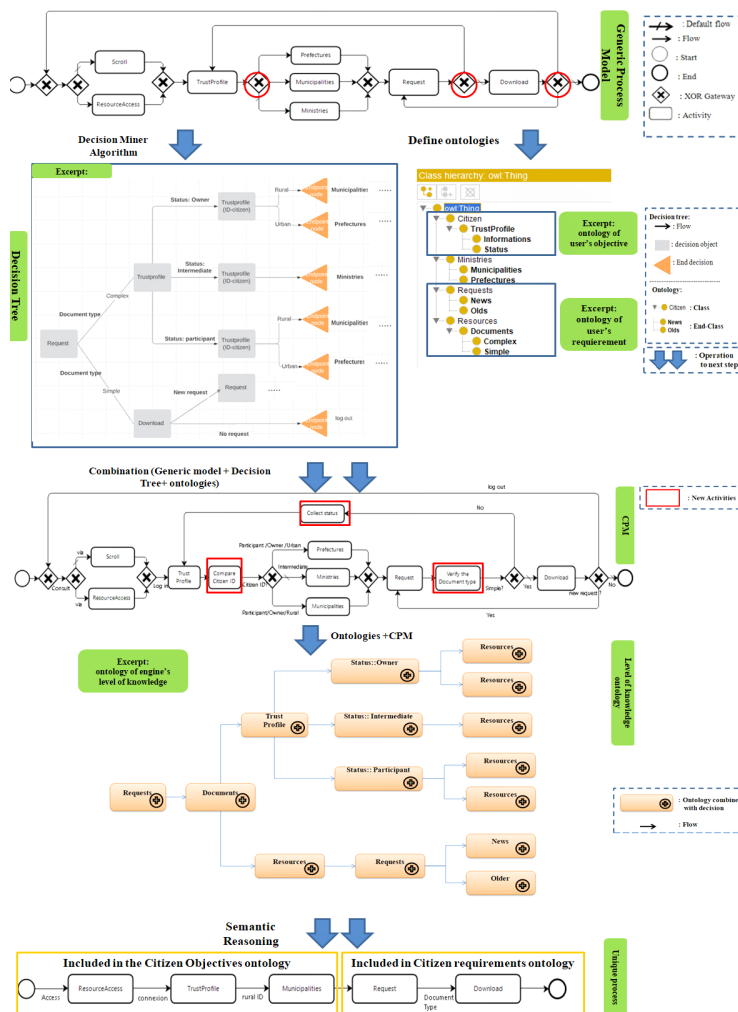
In this order, the CPM is obtained. This will add three additional activities: Activity of comparing the citizen-ID (verify if citizen origins; rural or urban area in order to contact the suitable area to its situation) Activity of collecting status (take into consideration the status verification) and the activity of verifying the document type (classify the request).

Next, the CPM is combined with the existing ontologies. This allows the emergence of an advanced ontology entitled: level of engine knowledge. Therefore, the correspondence between decisions points and different classes and sub-classes of ontologies is observed. In this order, the request class calls the document sub-class. This later has two choices called the trust profile sub-class or the resource sub-class, etc.

Last, the semantic reasoning based on the CPM and the advanced ontology is applied. This operation takes into consideration the inputs information (request of citizen), to recommend unique process. For instance, the citizen consults the website via the ResourceAccess option. He is connecting via trust profile with a rural ID. The citizen requests a simple document. The citizen can download the requested document.

Therefore, the recommended path is: ResourceAccessè Trust Profileè Municipalitiesè Requestè Download.

DISCUSSION

In this section, the authors discuss the results of the mining self-defined BP approach application, its related works, and its practice implication in two examples.

Results Analysis

In this work, the approach for mining self-defined BPs hands over three main results. These results provide a decision-making support for the user and enable to individualize a process model that respects the user's requirements and the internal or external regulations.

The first result represents the suitable process discovery algorithm to generate self-defined BP. This phase is similar to the studies of Mans et al. (2008), Augusto et al. (2018) and Jouck et al. (2018); the authors applied several process discovery algorithms in order to decide between their effectiveness and their limits. In this approach, the authors looked at self-defined processes in the context of electronic services, to be more specific, in e-administration. Based on the comparative study, the inductive miner achieves the best performance in terms of fitness and precision, while other algorithms cannot guarantee soundness. In case of complex events, it is necessary to use a filtering method prior to applying existing automated process discovery algorithms.

The second result is about detailing the variation points of the generic process model. In the literature, the application of decision miner algorithm is used with variable (Radgui et al., 2013), flexible (**Dustdar & Hoffmann**, 2005) and dynamic (Vasilecas et al., 2016) BPs as the Ad-hoc BPs example (**Duma & Aringhieri**, 2018). The novelty, here, is to introduce decision miner with self-defined BP.

The third result gives rise to the ontologies definition, to apply semantic reasoning and recommend suitable path to users. The authors extract knowledge from event logs to define ontologies, then combine them in semantic reasoning based on training samples. The added value over existing studies (**Detro et al.**, 2017; De Toledo et al., 2019) is the application in self-defined BP context, respecting its different properties, such as: variability, unique paths, different ontologies, etc.

Related Works

In reality, there is one field of research which addresses a comparable problematic, and it is user activity analysis. In this field, human computer-interaction is recorded to evaluate the user with respect to a specific research interest. The approach is used in several areas such e-commerce and online social networks research to create services like recommendation systems (**Plumbaum et al.,**

2009) or to analyze social behavior (**Dang et al.,** 2016). There is no case of user activity analysis related to electronic administration specifications, using process mining techniques in combination with semantic reasoning. Available approaches are used in the academic digital environment as e-learning (**Jadrić et al.,** 2020), digital libraries (Shiri, 2018) and in the healthcare domain. These scientific studies have been developed between 2008 and 2020.

Firstly, the work of (Diamantini et al., 2016) treats event logs in the domain of health. This paper introduces the Behavioral Process Mining approach, which is used to identify significant sub-processes from unstructured ones. This approach is based on the use of hierarchical clustering algorithms.

Second, the work in (Liu et al., 2017) uses extended methods, such as analysis sequence, to identify student processes in online courses. The objective of this work is to identify the most significant students' behaviors, to help professors, to adapt their teaching strategies to different student populations. Likewise, in (Jadrić et al., 2020) demonstrates how to leverage from process mining techniques in obtaining smart mobility and higher education. Third, the authors (Liu et al., 2017; Song & Günther, 2008) present approaches based on classification techniques, in order to compare process discovery algorithms. Fourth, the authors, in (Pérez-Alfonso et al., 2013), demonstrate the ability of process mining techniques to examine challenges that none confidential data poses against process discovery and conformance checking techniques.

These studies focus mainly on analyzing recorded traces in the context of representing users' behaviors, where the execution represents many repeated or semi-similar traces. While self-defined BP is a very particular category of unstructured processes, with complex and unrepeated behaviors, that depends on users' requirements, types of information, handling strategies, as well as the way how data are indexed. This demonstrates the originality of this mining self-defined BP approach that tackles precise process model representation and decision variability management. Indeed, advanced analysis is applied, and more metrics are taken into consideration.

Implication Examples

The mining self-defined BP approach can be applied in systems that require recommendations with semantic reasoning. In this regard, the authors present two concrete examples (e-learning and crowdfunding), where the implication of practice is clearly explained.

Example 1 (E-learning)

In the context of e-learning, the challenge for users, especially for students, is to be able to easily discover and exploit stored digitized resources. The accessibility to these resources usually starts by queries and is followed by various interactions with the search engine and the documents themselves. The user's interactions with digital resources presents unstructured processes where the execution characterized by high level of variability as unique or unsimilar behaviors. Indeed, they strongly depend on users' requirements, their levels of knowledge, types of information handling strategies, as well as the way how data are indexed.

In addition to the basic idea of process mining in detecting, monitoring, and improving real processes based on the extracted knowledge from event logs, the mining self-defined BP approach can be applied on the problems of education. The main goals in this direction is the personalization of educational processes via the recommendation of the best course units or learning paths to students, by defining different ontologies: profiles, preferences, or target skills, on which the semantic reasoning can be applied. Therefore, the use of the mining self-defined BP approach can improve the quality of teaching (Jadrić et al., 2020), i.e., the enhancement of educational process models with performance indicators: execution time, bottlenecks, decision point, etc.

Example 2 (Crowdfunding platforms)

Recent trends in funding takes into consideration the dominant role of crowdfunding (Brem et al., 2019) platforms to offer financial step-up to individual, small, or big companies. This is done by raising small amounts of money from a large number of people, typically via the Internet. Indeed, people are grouped into matrix nominated as "Crowdfunding success matrix", where each sponsor collect donations to finance their project. In this context, the applicability of the mining self-defined BP approach can recommend to each sponsor possible future participants. It lies in the possibility of proposing all the people who can be invited to join the matrix. It depends on the sponsor behavior, which is described by its project subject, location, and date of registration in the platform. Clearly, process mining can generate detailed sponsor behavior and define the ontologies that can be combined with the reasoning engine to recommend the list of suitable participants to each sponsor behavior. This allows increasing the matrix construction according to the time metric, by necessity the project can be quickly well financed in a very generous way.

Conclusion

This paper provides an approach for treating related challenges of self-defined BPs, in terms of process model's representation and variability management.

The authors study the applicability of process mining algorithms, to model the generic self-defined process model of user's behaviors in interaction with the e-administration domain (services provided by ministers, municipalities/local communities, and prefectures/states.). In these systems, users can have diverse ways to perform their research according to their objectives. In this context, users apply self-defined processes that may vary in terms of significance, structure, and results.

Besides, the quality of process discovery algorithms evaluation (Alpha++, FM, HM, GM, IM, LBR and SBR) is applied with fitness, precision and generalization measures; except for the fuzzy miner algorithm the node details and the conformance measures are used.

At this stage, the resulted self-defined process model requires variation point elaboration, to define possible choices related to the execution process. To do so, the use of decision miner algorithm is required. This algorithm aims at detailing all sub-processes of the generic self-defined process model for defining self-defined BP ontologies, which are: user objective, user requirement and engine knowledge level. Hence, the configurable process model can be obtained.

Last, the combination between the semantic reasoning through ontologies and the CPM can be released, to manage self-defined BP variability. For this purpose, this work aims at selecting the suitable process variant according to the user's objective, requirements, and knowledge level of the used engine (training data). To do so, the orchestration between back-end and front-end aspects details the steps that must be followed to manage self-defined BPs.

The proposed approach seems to be interested in the way that it takes advantages from the combination between process mining and ontologies using semantic reasoning to analyze self-defined BPs. This brings new knowledge to the research field in terms of producing a complete scenario of the orchestration of existing and new techniques through a defined approach, from the extraction of a process model based on event data to recommendations at a later stage. Moreover, the use of the chaotic activities' filters (Tax et al., 2019) positively impacts the quality of discovered process models. Therefore, more precise representation is sufficiently training, as trusted data, the ontology's definition step.

However, some limitations of this approach should be noted. First, the challenge of gathering self-defined events at runtime must be treated. It could be difficult in the big data context (Bernardi et al., 2018). However, from a technical standpoint, it still be possible, such is not an easy task because the consequences of failing to properly collect data include the inability to answer to research questions, inability to validate the results, distorted findings, wasted resources, misleading recommendations and decisions, and harm to participants. Hence, intelligent information retrieval systems must be included (Croft, 2019). Second, the approach uses supervised learning techniques that focus on labelled samples. It would be fair to say that it reflects an ambiguity in the case of no ground truth knowledge about the business process, especially in the definition of ontologies (which ontologies are exactly fit). Therefore,

an extended approach must be developed using unsupervised learning algorithms. For instance, the use of unsupervised clustering (De Souza & Queiroz, 2020) can be profitable in ignoring abstractly labeled samples.

As further work, it is important to treat the ability of process mining techniques, in modelling uncertain behaviors of self-defined processes, related to the information retrieval systems. This is in the objective of achieving the business process maturity (Milanović, 2020) and measuring how effectively and efficiently the self-defined BP is working.

ACKNOWLEDGMENT

REFERENCES

Athukorala, K., Głowacka, D., Jacucci, G., Oulasvirta, A., & Vreeken, J. (2016). Is exploratory search different? A comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*, *67*(11), 2635–2651. doi:10.1002/asi.23617

Augusto, A., Conforti, R., Dumas, M., La Rosa, M., Maggi, F. M., & Marrella, A. et al.. (2018). Automated discovery of process models from event logs: Review and benchmark. *IEEE Transactions on Knowledge and Data Engineering*, *31*(4), 686–705. doi:10.1109/TKDE.2018.2841877

Ayora, C., Torres, V., Reichert, M., Weber, B., & Pelechano, V. (2012). Towards run-time flexibility for process families: open issues and research challenges. In *International Conference on Business Process Management*. Springer.

Bernardi, M. L., Cimitile, M., & Mercaldo, F. (2018). Cross-organisational process mining in cloud environments. *Journal of Information & Knowledge Management*, *17*(02), 1850014. doi:10.1142/S0219649218500144

Bogarín Vega, A., Cerezo Menéndez, R., & Romero, C. (2018). Discovering learning processes using inductive miner: A case study with learning management systems (LMSs). *Psicothema*, *30*(3), 322–329. PMID:30009756

Brem, A., Bilgram, V., & Marchuk, A. (2019). How crowdfunding platforms change the nature of user innovation–from problem solving to entrepreneurship. *Technological Forecasting and Social Change*, *144*, 348–360. doi:10.1016/j.techfore.2017.11.020

Buijs, J. C., Van Dongen, B. F., & van Der Aalst, W. M. (2012). On the role of fitness, precision, generalization, and simplicity in process discovery. In *On the Move to Meaningful Internet Systems Confederated International Conferences*. Springer. doi:10.1007/978-3-642-33606-5_19

Cheng, H. J., & Kumar, A. (2015). Process mining on noisy logs-Can log sanitization help to improve performance? *Decision Support Systems*, *79*, 138–149. doi:10.1016/j.dss.2015.08.003

Cole, M. J., Hendahewa, C., Belkin, N. J., & Shah, C. (2015). User activity patterns during information search. *ACM Transactions on Information Systems*, *33*(1), 1–39. doi:10.1145/2699656

Croft, W. B. (2019). The Importance of Interaction for Information Retrieval. *Conference on Research and Development in Information Retrieval*, *19*, 1-2. doi:10.1145/3331184.3331185

Dang, A., Moh'd, A., Milios, E., & Minghim, R. (2016). What is in a rumour: Combined visual analysis of rumour flow and user activity. *Proceedings of the 33rd Computer Graphics International*, 17-20. doi:10.1145/2949035.2949040

Detro, S. P., Portela, E., Rocha, E. L., Panetto, H., & Lezoche, M. (2017). Configuring process variants through semantic reasoning in systems engineering. *International Council on Systems Engineering*, *20*(4), 36–39. doi:10.1002/inst.12179

De Souza Oliveira, M., & Queiroz, S. (2020). Unsupervised Feature Selection Methodology for Clustering in High Dimensionality Datasets. *Revista de Informática Teórica e Aplicada*, *27*(2), 30–41. doi:10.22456/2175-2745.96081

De Toledo, P., Joppien, C., Sesmero, M. P., & Drews, P. (2019). Mining Disease Courses across Organizations: A Methodology Based on Process Mining of Diagnosis Events Datasets. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 354-357. doi:10.1109/EMBC.2019.8857149

Diamantini, C., Genga, L., & Potena, D. (2016). Behavioral process mining for unstructured processes. *Journal of Intelligent Information Systems*, *47*(1), 5–32. doi:10.1007/s10844-016-0394-7

Duma, D., & Aringhieri, R. (2020). An ad hoc process mining approach to discover patient paths of an Emergency Department. *Flexible Services and Manufacturing Journal*, *32*(1), 6–34. doi:10.1007/s10696-018-9330-1

Dustdar, S., Hoffmann, T., & Van der Aalst, W. (2005). Mining of ad-hoc business processes with TeamLog. *Data & Knowledge Engineering*, *55*(2), 129–158. doi:10.1016/j.datak.2005.02.002

El Faquih, L., Sbai, H., & Fredj, M. (2014). Semantic variability modeling in business processes: A comparative study. In *The 9th International Conference for Internet Technology and Secured Transactions (ICITST-2014)*. IEEE.

Fei, H., & Meskens, N. (2010). Discovering patient care process models from event logs. In *Proc. 8th Int. Conf. Modeling Simulation (MOSIM)* (pp. 10-12). Academic Press.

Gottschalk, F., Van der Aalst, W. M., & Jansen-Vullers, M. H. (2007). *Configurable process models-a foundational approach.* Springer.

Ghoula, N., Khelif, K., & Dieng-Kuntz, R. (2007, November). Supporting patent mining by using ontology-based semantic annotations. In *International Conference on Web Intelligence (WI'07)*. IEEE. doi:10.1109/WI.2007.45

Günther, C. W., & Van Der Aalst, W. M. (2007). Fuzzy mining–adaptive process simplification based on multi-perspective metrics. In *International conference on business process management*. Springer. doi:10.1007/978-3-540-75183-0_24

Hai, C., & Jeong, I. (2007). *Fundamental of development administration*. Scholar.

Hepp, M., & Roman, D. (2007). An ontology framework for semantic business process management. In *International Conference on Business Information Systems*. Springer.

Huang, Y., Feng, Z., He, K., & Huang, Y. (2013). Ontology-based configuration for service-based business process model. In *2013 IEEE International Conference on Services Computing*. IEEE. doi:10.1109/SCC.2013.59

Islam, M. S., & Scupola, A. (2011). E-service research trends in the domain of e-Government: A Contemporary Study. *International Journal of E-Services and Mobile Applications*, *3*(1), 39–56. doi:10.4018/jesma.2011010103

Jadrić, M., Ninčević Pašalić, I., & Ćukušić, M. (2020). Process Mining Contributions to Discrete-event Simulation Modelling. Business Systems Research. *International Journal of the Society for Advancing Innovation and Research in Economy*, *11*(2), 51-72.

Jouck, T., Bolt, A., Depaire, B., de Leoni, M., & van der Aalst, W. M. (2018). *An Integrated Framework for Process Discovery Algorithm Evaluation.* preprint arXiv:1806.07222.

Kasprzyk, B. (2018). E-administration digital services in Poland. *Nierówności społeczne a wzrost gospodarczy*, *53*, 308-319.

King, S. F., & Johnson, O. A. (2006). VBP: An approach to modelling process variety and best practice. *Information and Software Technology*, *48*(11), 1104–1114. doi:10.1016/j.infsof.2006.02.003

Kumar, A., & Yao, W. (2012). Design and management of flexible process variants using templates and rules. *Computers in Industry*, *63*(2), 112–130. doi:10.1016/j.compind.2011.12.002

La Rosa, M., van der Aalst, W. M., Dumas, M., & Ter Hofstede, A. H. (2009). Questionnaire-based variability modeling for system configuration. *Software & Systems Modeling*, *8*(2), 251–274. doi:10.1007/s10270-008-0090-3

Lamghari, Z., Radgui, M., Saidi, R., & Rahmani, M. D. (2019). A Framework Supporting Supply Chain Complexity and Confidentiality Using Process Mining and Auto Identification Technology. *International Conference Europe Middle East & North Africa Information Systems and Technologies to Support Learning*, 352-361.

Lanza, B. B. B., & Cunha, M. A. (2013). Relations among Actors in Governmental Projects: The Case of Paraná mGov. *International Journal of E-Services and Mobile Applications*, *5*(3), 25–42. doi:10.4018/jesma.2013070102

Laudon, K. C., & Traver, C. G. (2016). *E-commerce: business, technology, society*. Addison-Wesley Pub (Sd).

Liao, Y., Lezoche, M., Panetto, H., Boudjlida, N., & Loures, E. R. (2015). Semantic annotation for knowledge explicitation in a product lifecycle management context: A survey. *Computers in Industry*, *71*, 24–34. doi:10.1016/j.compind.2015.03.005

Liu, S., Hu, Z., Peng, X., Liu, Z., Cheng, H. N., & Sun, J. (2017). Mining learning behavioral patterns of students by sequence analysis in cloud classroom. *International Journal of Distance Education Technologies*, *15*(1), 15–27. doi:10.4018/IJDET.2017010102

Luengo, D., & Sepúlveda, M. (2011). Applying clustering in process mining to find different versions of a business process that changes over time. In *International Conference on Business Process Management*. Springer.

Mans, R. S., Schonenberg, M. H., Song, M., van der Aalst, W. M., & Bakker, P. J. (2008). Application of process mining in healthcare–a case study in a dutch hospital. In *International joint conference on biomedical engineering systems and technologies*. Springer. doi:10.1007/978-3-540-92219-3_32

Milanović Glavan, L. (2020). An Investigation of Business Process Maturity: Report on Croatian Companies. *Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy*, *11*(2), 159-165.

Naouar, F., Hlaoua, L., & Omri, M. N. (2017a). Collaborative Information Retrieval Model based on Fuzzy Clustering. In *2017 International Conference on High Performance Computing & Simulation (HPCS)*. IEEE. doi:10.1109/HPCS.2017.80

Naouar, F., Hlaoua, L., & Omri, M. N. (2017b). Information retrieval model using uncertain confidence's network. *International Journal of Information Retrieval Research*, *7*(2), 34–50. doi:10.4018/IJIRR.2017040103

Nouvellet, A., Beaudouin, V., D'Alché-Buc, F., Prieur, C., & Roueff, F. (2017). *Analyse des traces d'usage de Gallica*. Academic Press.

Pegoraro, M., & van der Aalst, W. M. (2019). Mining uncertain event data in process mining. In *2019 International Conference on Process Mining (ICPM)*. IEEE. doi:10.1109/ICPM.2019.00023

Pérez-Alfonso, D., Yzquierdo-Herrera, R., & Lazo-Cortés, M. (2013). Recommendation of process discovery algorithms: A classification problem. *Research in Computing Science*, *61*, 33–42.

Plumbaum, T., Stelter, T., & Korth, A. (2009). Semantic web usage mining: Using semantics to understand user intentions. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 391-396). Process mining group. Math & Computing department, Eindhoven University of Technology. http://www.processmining.org/prom/decisionmining?s%5b%5d=decision&s%5b%5d=mining

Radgui, M., Saidi, R., & Mouline, S. (2013). Design for reuse in business process: Method and experiments. *International Journal of Enterprise Information Systems*, *9*(4), 12–27. doi:10.4018/ijeis.2013100102

Rafiei, M., von Waldthausen, L., & van der Aalst, W. M. (2018). Ensuring Confidentiality in Process Mining. In *The 7th international symposium on data-driven process discovery and analysis*. Springer.

Ruso, J., Krsmanovic, M., Trajkovic, A., & Rakicevic, Z. (2013). Quality Management in Public e-Administration. *International Journal of Management Science and Engineering*, *7*(10), 550–554.

Sani, M. F. (2020). Preprocessing Event Data in Process Mining. In *CAiSE* (pp. 1–10). Doctoral Consortium.

Seiger, R., Huber, S., Heisig, P., & Assmann, U. (2019). *A framework for self-adaptive workflows in cyber-physical systems. In Software Engineering and Software Management*. Springer.

Shiri, A. (2018). Methodological Considerations in Developing Cultural Heritage Digital Libraries: A Community-driven Framework. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. ACM. doi:10.1145/3197026.3203893

Song, M., Günther, C. W., & Van der Aalst, W. M. (2008). Trace clustering in process mining. In *International conference on business process management*. Springer.

Staab, S., & Studer, R. (2010). *Handbook on ontologies*. Springer Science & Business Media.

Szabó, I., & Varga, K. (2014). Knowledge-based compliance checking of business processes. In *On the Move to Meaningful Internet Systems Confederated International Conferences*. Springer.

Tax, N., Sidorova, N., & van der Aalst, W. M. (2019). Discovering more precise process models from event logs by filtering out chaotic activities. *Journal of Intelligent Information Systems*, *52*(1), 107–139. doi:10.1007/s10844-018-0507-6

Vanden Broucke, S. K., & De Weerdt, J. (2017). Fodina: A robust and flexible heuristic process discovery technique. *Decision Support Systems*, *100*, 109–118. doi:10.1016/j.dss.2017.04.005

Van der Werf, J. M. E., van Dongen, B. F., Hurkens, C. A., & Serebrenik, A. (2008). Process discovery using integer linear programming. In *International conference on applications and theory of petri nets*. Springer. doi:10.1007/978-3-540-68746-7_24

Van der Aalst, W. M., van Dongen, B. F., Günther, C. W., Rozinat, A., Verbeek, E., & Weijters, T. (2009). ProM: The process mining toolkit. *BPM (Demos)*, *489*(31), 2.

Van der Aalst, W. M., Rubin, V., Verbeek, H. M. W., van Dongen, B. F., Kindler, E., & Günther, C. W. (2010). Process mining: A two-step approach to balance between underfitting and overfitting. *Software & Systems Modeling*, *9*(1), 87–111. doi:10.1007/s10270-008-0106-z

Van der Aalst, W. M. P. (2016). *Process mining: data science in action*. Springer. doi:10.1007/978-3-662-49851-4

Vasilecas, O., Rusinaite, T., & Kalibatiene, D. (2016). Dynamic Business Processes and Their Simulation: A Survey. Data base and information Systems, 9, 155-166.

Weijters, A. J., & Van der Aalst, W. M. (2003). Rediscovering workflow models from event-based data using little thumb. *Integrated Computer-Aided Engineering*, *10*(2), 151–162. doi:10.3233/ICA-2003-10205

Wen, L., Van Der Aalst, W. M., Wang, J., & Sun, J. (2007). Mining process models with non-free-choice constructs. *Data Mining and Knowledge Discovery*, *15*(2), 145–180. doi:10.1007/s10618-007-0065-y

Zineb Lamghari received her technical university degree in software engineering at the Higher School of Technical Education, Mohammed V- Souissi University, Morocco, and her licence degree in information systems management from polydisciplinary faculty then her master's degree in computer system and network from Faculty of Technical Sciences, Abdelmalek Essaadi University, Morocco. She prepares her PhD in business process improvement in the Computer Science and Telecommunication Research Laboratory (LRIT) at the faculty of sciences, Mohammed V University. Her interests cover mainly business process management, software engineering and process mining techniques.

Rajaa Saidi is an Associate Professor at the National Institute of Statistics and Applied Economics – Rabat (INSEA-Morocco); she is a member of the Information Systems, Intelligent Systems and Mathematical Modelling Laboratory (SI2M) of (INSEA). She holds a PhD degree in Information Systems and Software Engineering from Mohammed V University of Rabat, Morocco and the Grenoble Institute of Technology (INPG-France). She is also a member of the Computer Science and Telecommunication Research Laboratory (LRIT) in Mohammed V University. Her research areas include information systems, business process management, ubiquitous computing, context-aware information systems, service-oriented architectures and component-based engineering.

Maryam Radgui is an Assistant Professor at the National Institute of Statistics and Applied Economics (INSEA) in Rabat. She is a member of the Information Systems, Intelligent Systems and Mathematical Modelling Laboratory (SI2M) of INSEA and she is also a member of the Computer Science and Telecommunication Research Laboratory (LRIT) in Mohammed V University in Rabat. She received her PhD in Software Engineering from the Mohammed V University in Rabat, Morocco in 2015. Her research interests are mainly focused in information systems, software development, business process, process mining, reuse and service-based development methods. She has published several papers in international journals, conferences, and workshops.

Moulay Driss Rahmani is a Professor of Higher Education at the Faculty of Sciences Rabat, Morocco. He received his PhD in Surfaces, Interfaces and Devices Studies Laboratory, Montpellier II University, France. His interests are mainly focused on human-computer interaction, scientific computing, and urbanism. He has published several papers in international conferences, workshop, and journals on these topics.