


Article

Candidate Digital Tasks Selection Methodology for Automation with Robotic Process Automation

Daehyoun Choi ¹, Hind R'bigui ^{2,*}  and Chiwoon Cho ^{1,*}¹ School of Industrial Engineering, University of Ulsan, Ulsan 44610, Korea; dhchoi@nsoft.co.kr² Digital Enterprise Department, Nsoft Co. Ltd, No.407, 14, Tekeunosaneop-ro 55beon-gil, Nam-gu, Ulsan 44776, Korea

* Correspondence: hind@nsoft.co.kr (H.R.); chiwoon6@mail.ulsan.ac.kr (C.C.)

Abstract: Today's business environments face rapid digital transformation, engendering the continuous emerging of new technologies. Robotic Process Automation (RPA) is one of the new technologies rapidly and increasingly grabbing the attention of businesses. RPA tools allow mimicking human tasks by providing a virtual workforce, or digital workers in the form of software bots, for automating manual, high-volume, repetitive, and routine tasks. The goal is to allow human workers to delegate their tedious routine tasks to a software bot, thus allowing them to focus on more difficult tasks. RPA tools are simple and very powerful, according to cost-saving and other performance metrics. However, the main challenge of RPA implementation is to effectively determine the business tasks suitable for automation. This paper provides a methodology for selecting candidate tasks for robotic process automation based on user interface logs and process mining techniques.

Keywords: RPA; robotic process automation; software robots; process mining; task mining; event log; user interface log; knowledge discovery; digital workforce; digital transformation



Citation: Choi, D.; R'bigui, H.; Cho, C. Candidate Digital Tasks Selection Methodology for Automation with Robotic Process Automation. *Sustainability* **2021**, *13*, 8980. <https://doi.org/10.3390/su13168980>

Academic Editors: Baihua Li and Fei Chao

Received: 24 June 2021

Accepted: 3 August 2021

Published: 11 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Technology is changing at a faster pace. Disruptive firms are one source of technological change, which have the potential to create and develop radical innovations that disrupt existing products and support industrial, economic, and social change [1–6]. Moreover, since the outbreak of the COVID-19 pandemic, digital transformation processes have significantly hastened and become the unchallenged leader of all initiatives launched by those companies and organizations that have been implementing initiatives related to digitization for a long time as well as those that have only transformed their activities in the areas of internal operations and work because of the pandemic [7]. In almost all industries, organizations have started conducting initiatives to explore new digital technologies and get their benefits. This includes transformations of crucial business operations, processes, products, organizational structures, and management aspects [8]. Digital transformation requires organizations to adapt their existing business models and enhance the automation of their business processes [9]. Robotic Process Automation (RPA) is one of the most recent developments to boost the automation level of business processes. RPA is a famous subject in the corporate world [10]. RPA uses many artificial intelligence and machine learning techniques: image recognition, Optical Character Recognition (OCR), etc. It is considered a new wave of digital technologies [11], which is increasingly drawing the attention of industries and administrations. RPA is a new technology that enables the automation of high-volume, manual, repeatable, routine, rule-based, and unmotivating human tasks [12]. This technology utilizes software robots to replace human actions for performing administrative activities [13]. Using bots to execute repetitive tasks saves organizations time and money and reduces errors. Moreover, software bots allow employees to focus more on higher-level work instead of on tedious tasks. Consequently, the return on investment is colossal. An example of clerical tasks is the as-is process performed by

a human depicted in Figure 1. The to-be process performed by a robot is illustrated in Figure 2. The to-be workflow is similar to the as-is workflow. However, some tedious tasks that were performed by a human user are now performed by a bot user. RPA has decreased the threshold for process automation. Repetitive activities performed by people can now be handed over to software bots. Software robots replace users by interacting directly with the user interface that was originally operated only by people and do not modify or replace any pre-existing information system in the organization. RPA is considered cheaper than traditional automation solutions. Therefore, it can be exploited to automate routine tasks that are broadly considered not cost-effective [14].

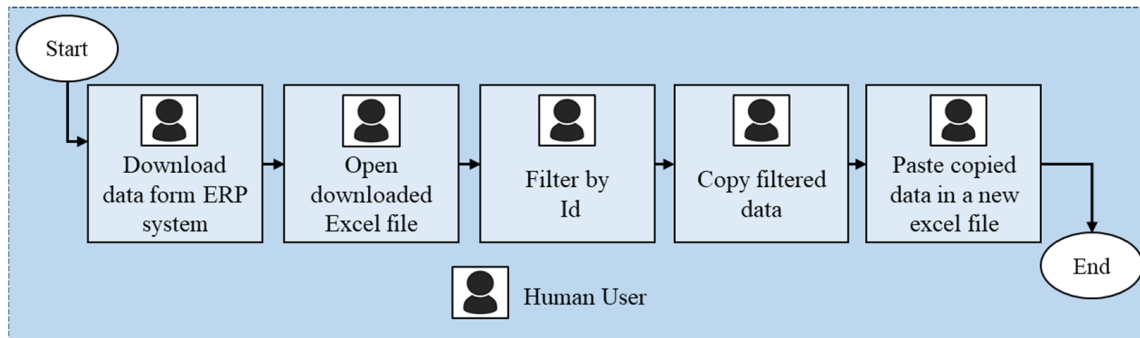


Figure 1. Example of routine tasks performed by a human.

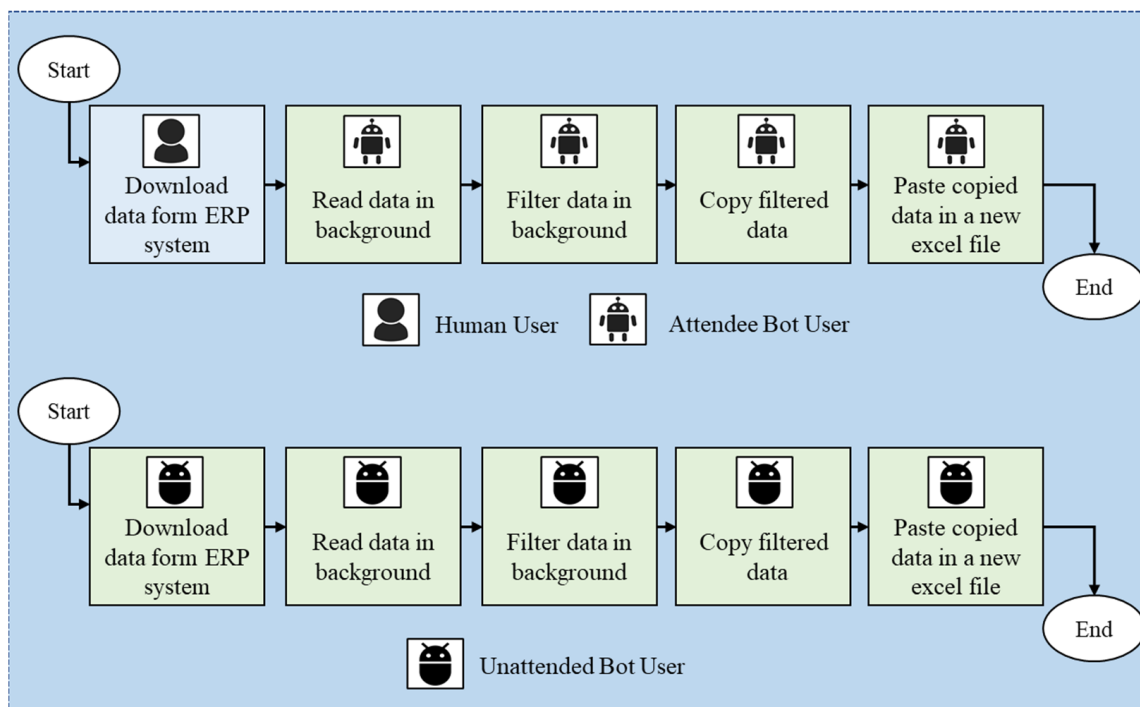


Figure 2. Example of routine tasks performed by a bot user.

Various benefits of implementing RPA within organizations have been reported (e.g., [13,15–18]) as it speeds up business growth by reducing a lot of manual and repetitive work. Nonetheless, as the technology is still new, the implementation of RPA faces some challenges [19]. As is known, RPA solutions automate repetitive manual tedious tasks. These repetitive, manual, and mundane tasks constitute the input for RPA tools. RPA solutions do not identify the tasks that can be or need to be automated. Thus, the question is: Which user work routines can favorably be automated with RPA? This is the main

challenge [12]. We need to identify beforehand the tasks that need to be automated to be able to use RPA tools. This work proposes a methodology to identify candidate digital tasks for automation with RPA tools. Digital tasks are tasks performed using a computer by interacting with the different user interfaces of different systems and applications. The proposed approach is based on user interface interaction logs and process mining techniques. Process mining [20] techniques extract knowledge from the event logs that record the execution of business processes, and which are stored in today's information systems such as Enterprise Resource Planning (ERP) systems, Workflow Management (WFM) systems, Supply Chain Management (SCM) systems, etc. One of the main techniques of process mining is process discovery. This technique takes an event log as input and automatically creates a process model which shows how the business process is behaving [21]. Process mining can play a primary role in deciding the tasks that can be automated. Since RPA operates on the user interface level; we can use process mining techniques to discover tasks from the user interface interaction logs. A process is a set of tasks, and a task consists of a sequence of steps. The goal of process mining is to discover business processes that are composed of a set of tasks or operations. However, here, the aim of using process mining techniques is for task discovery not for process discovery. Specifically speaking, the discovery of office tasks performed in a user interface. To perform a specific task (e.g., order handling), a user needs to perform several steps in a user interface. There are basically only two types of actions that can be performed in a user interface: mouse clicks and keyboard. Thus, steps are performed by switching screens with mouse clicks or/and entering different content with the keyboard. To do this, the actions performed in a user interface with the mouse and the keyboard need to be recorded to obtain the data. From the recorded data related to interactions with the user interface, a log can be generated. This log needs to be pre-processed so that it can be ready to be used by process mining techniques.

The contribution of this paper is to identify the tasks that are a candidate for automation with robotic process automation. This work shows how the tasks executed by an employee are discovered using process mining discovery techniques and user interface interaction logs. Then, once all tasks are discovered, the work will show how the manual and repetitive tasks will be selected.

This paper is organized as follows. The second section presents robotic process automation architecture and robot types, an overview of process mining, and related work. The proposed approach is explained in Section 3. A use case that applies the proposed approach is provided in Section 4. Finally, Section 5 concludes the work and highlights limitations.

2. Theoretical Framework

2.1. Robotic Process Automation Elements

The role of technology is emerging continuously at a rapid pace. Many businesses around the world are leveraging Robotic Process Automation (RPA) in a large variety of fields to grant a significant competitive advantage in business process automation. In a study, Gartner predicted that by 2020, companies investing in Robotic Process Automation software will reach \$1 billion, at a growth rate of 41% from 2015 through 2020 [22]. In another study, the global RPA market size has been estimated at USD 1.57 billion in 2020, and at USD 1.89 billion in 2021, and it has been predicted to grow at a rate of 32.8% from 2021 to 2028 [23].

Robotic process automation is composed of three components: bots, a studio, and an orchestrator. (1) RPA bots are the virtual workforce that will be executing repetitive and mundane human tasks. They are dedicated to handling an unmotivated mass of tasks so that employees can engage in valuable jobs and problem-solving. There are two categories of RPA bots, attended bots and unattended bots. Attended bots are bots configured to work side by side with a human user. The goal is to speed up the repetitive tasks where the tasks need to be triggered by the human user. This type of bot can be used in routine, manual, and rule-based tasks which require human intervention for decision points. Unattended bots are bots configured to work fully independently in the background. This type of bot is

dedicated to working without the intervention of a human user and can be scheduled to be started and executed automatically and triggered by a satisfying condition or a business event. Unattended bots can be used in routine, manual, and rule-based tasks which do not require any human intervention. (2) An RPA studio is responsible for configuring the workflow to be executed by the bots which will be mimicking the human behavior. It enables users to create, design, and automate the workflow to be executed by bots. Business users are enabled to configure the bots by record and screenplay capability and intuitive scenario design interface. (3) The RPA Orchestrator is responsible for scheduling, managing, monitoring, and auditing the bots. The bots can be used with third-party applications using Application Programming Interfaces (APIs). The elements of robotic process automation are illustrated in Figure 3.

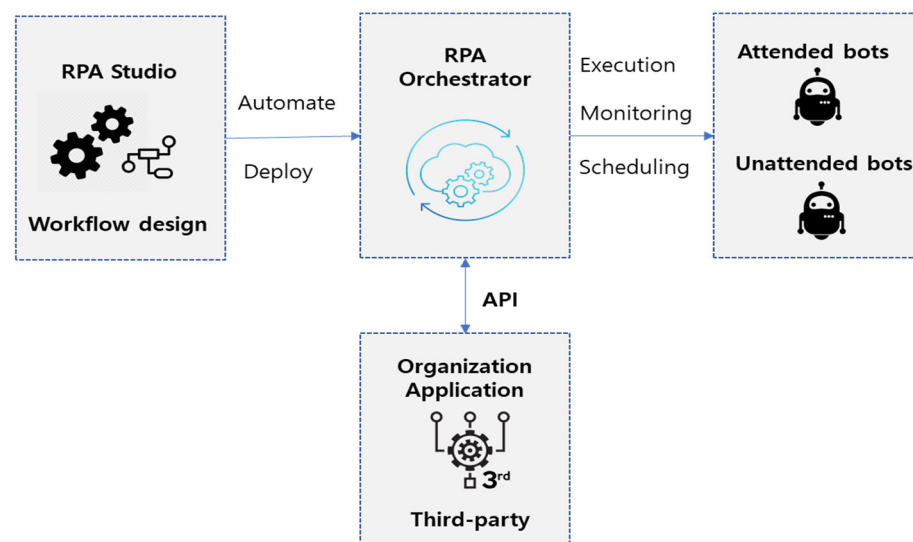


Figure 3. Robotic process automation components.

Integrating robotic process automation with Machine Learning and Artificial Intelligence can aid in the advancement of the capabilities of software bots beyond rule-based processes and in improving business insights and ameliorating data integrity [24]. Furthermore, integrating computer vision applications such as image processing, image detail extraction, text recognition in an image, occurrence detection in an image, etc., can largely help software bots in automating detailed tasks on a graphical user interface [25].

2.2. Process Mining

Process mining has been defined in [20] as a relatively young research discipline that falls between artificial intelligence and data mining on the one hand and process modeling and analysis on the other hand. Moreover, it was outlined in [26] as a recent set of techniques that provide a strong bridge between BI and BPM by combining both process models and event data forming a new form of process-driven analytics.

The main idea of business process mining is to extract the execution of business processes recorded in event logs available in today's information system to automatically discover the models of business processes, compare existing business process models with the new automatically constructed ones to identify deviations and bottlenecks, and to enhance the business processes. Process mining techniques exploit event data stored in today's information systems to show how people, machines, and organizations are really behaving. Using event data, process mining provides useful insights that can be utilized to improve processes of different domains.

There are basically four types of process mining [27]. The first type of process mining is process discovery. It is considered the most important type of process mining. Process discovery consists of learning process models from event data by taking only an event log as

input and producing a process model which shows what people, machines, organizations are really doing. There are three categories of process discovery perspectives based on the information available in the event data: the control-flow perspective, the organizational perspective, and the case perspective. These categories can be used to analyze processes from different perspectives. The second type of process mining is conformance checking. This type consists of checking whether the existing model is in conformance with the reality which is recorded in the event log. In other words, it checks whether what we think is happening is conform to what is really happening by identifying commonalities and differences between the existing process model and the recorded event log [20]. The third type of process mining is process reengineering. This type also takes an event log and a process model as input like the conformance checking. Nevertheless, the aim of this type is to change the existing process model not identifying the differences. The process model can be changed to better match reality. The fourth type of process mining is operational support. This type allows people to act at the time the process deviates by providing predictions (e.g., remaining time), warnings, or recommendations [28].

2.3. Related Work

Van der Aalst [29] defines RPA as an umbrella term for tools that operate on the user interface of other computer systems in the same way as humans. IEEE Corporate Advisory Group defines RPA as the use of a “preconfigured software instance that uses business rules and predefined activity choreography to complete the autonomous execution of a combination of processes, activities, transactions, and tasks in one or more unrelated software systems to deliver a result or service with human exception management” [30].

Robotic Process Automation is recently receiving increasing attention from industries and administrations due to the huge desire to implement digital transformation. There are three leading RPA vendors: (1) Blue Prism, (2) the most famous UiPath, and (3) Automation Anywhere. These tools have been demonstrated to be simple and very powerful in cost-saving and other performances. The main idea is that today’s information systems are not changed; only the tasks that have been performed by people by interacting with these systems are automated [14].

RPA is related to Workflow Management (WFM) [31], which has been available for several decades. However, workflow management was not that attainable since traditional automation is too expensive. Now Workflow Management might be achievable with robotic process automation.

Various benefits of implementing RPA have been reported (e.g., [18,32]) as it speeds up business growth by reducing a lot of manual and repetitive task-based works. However, the implementation of RPA is encountering some challenges. The main challenge is to properly determine first the candidate tasks that can be/need to be automated with RPA [11]. This challenge is not well researched. The identification of candidate tasks for automation via RPA tools is, so far, a largely unexplored problem [33]. A recent study proposed a methodology to analyze UI logs for the purpose to discover routines for RPA [34]. However, the presented approach focuses only on copying data from a spreadsheet or a form to another. The discovery of candidate tasks for RPA is related also to the field of a webform and automatic completion of tables. This latter consists for instance of detecting patterns from the values present in the cells of a sheet of an Excel file and then automating the completion of the table based on the detected patterns [35]. However, the focus of those approaches is tailored only toward partial automation. Another study proposed a supervised machine learning and natural language processing-based approach to automatically determine whether a task in a textual process description is a manual task, a user task where a human user interacts with an information system, or an automated task [9]. Nevertheless, the study does not identify candidate tasks to be automated. Process mining has been shown that it can be used to determine work done by people in [27]. Some other recent studies [36–38] show that RPA can benefit from process mining. Process mining enables process enhancement using event data. The starting of process mining

is an event log generated from today's information systems where each event refers to a task executed either by a person or a machine at a specific time and for a specific case. Therefore, we proposed a methodology to determine candidate tasks for automation using process mining techniques and user interface logs generated by recording the interactions with web and desktop applications.

3. Methodology

Robotic process automation tools can automate a lot of user routines in a business process. Thus, which user routines in an organization can be favorably automated with RPA? To answer this question, a class of tools called RPM (Robotic Process Mining Tools) tools has been envisioned in the work [33]. RPM has been defined as a class of methods and tools that will be used to analyze data obtained during the execution of user-driven tasks. The goal of RPM is to support the determination of candidate user processes that can be automated by RPA robots. In the same context, this section explains the approach we propose for selecting candidate routines/tasks to be automated with robotic process automation tools. In the next section, the application of the approach with a use case is explained. Since RPA is dedicated to automating manual and repetitive tasks operated on the user interface level, one can record the execution of all tasks and all the interactions between user interfaces. From the record, a UI log can be generated. The UI log shall contain user-driven tasks that involve interactions between a user and software applications. This latter can be used then as input by process mining techniques to discover routines. From the discovered model, candidate tasks for automation using RPA can be selected. Therefore, the proposed approach is composed of four steps: user interface (UI) log generation, UI log transformation into a log supported by process mining techniques, routines discovery with process mining, and candidate tasks selection based on specific criteria. An overview of the approach is depicted in Figure 4. Each step is explained in the upcoming subsections.

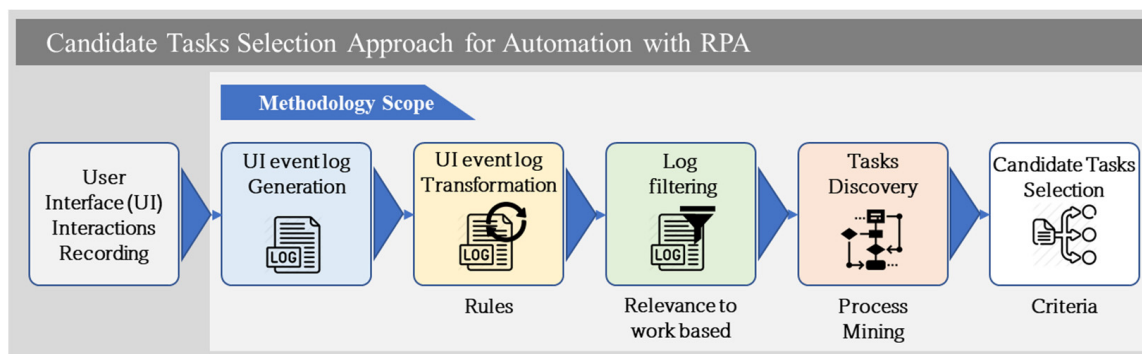


Figure 4. Proposed Approach.

3.1. User Interface Log

Generally, the input of process mining techniques is an event log that contains tasks belonging to a specific business process. Similarly, the input for RPM is a user interface log that also contains tasks. However, this time, the tasks are user-driven actions that involve interactions between a human user and software applications. A user interface log represents a sequence of actions performed in chronological order by a user when interacting with different applications (e.g., web, desktop, system, application) when performing an administrative task. We define an example of a user interface log in Table 1. Each row refers to a particular action and every action is characterized by timestamps, and other information showing where the action was performed and what objects were involved in performing the action. This work defines the user interface log such that it contains the following elements: timestamp, action type, source type, source name, content, field name, and the field value. The timestamp refers to the time when the action on the user interface was performed. Action type refers to the actions that have been performed on a user interface. Source type, source name, content, field

name, and field value are information related to the objects involved in performing the action. For instance, let us consider the first example.

Table 1. Example of a User Interface log.

Case ID	Timestamp	Action Type	Source Type	Source Name	Content	Field Name	Field Value
1	2020-08-20 9:29:10	Open	System Folder	Orders			
1	2020-08-20 9:29:16	Open	Excel	Order Qty			
1	2020-08-20 9:29:18	Go to Cell	Sheet	Sheet 1		Customer ID	
1	2020-08-20 9:29:23	Click Button	Sheet	Sheet 1	Filter button		
1	2020-08-20 9:29:27	Select Cell	Sheet	Sheet 1		Customer ID	Customer A
2	2020-08-20 9:33:27	Enter	Web	https://www.erp.kor , accessed on 22 July 2021		Order ID	"OR208201"

Example 1. *Opening an Excel with the name "Order Qty".*

Open is the action type. Excel is the source type of the object for which the action was performed and Order Qty is the name of the source, the name of the object for which the action was performed. Let us consider the second example of a task on a user interface.

Example 2. *Clicking on the filter button of sheet 1 of an Excel file.*

Click Button would be the action type. Sheet 1 would be the source name. The source type is an Excel sheet. Clicking a button with the mouse is an action, but there are many different buttons that can be clicked. To differentiate between them, the type of the clicked button can be recorded in the content column. Let us look at the third example.

Example 3. *Selecting the cell of sheet 1 which contains the value "Customer A" from the Customer ID column.*

Select Cell refers to the action type. The action-type "Select cell" is performed on the source name Sheet 1. The item or the cell that has been selected is "Customer A" which is considered as a field value. This value belongs to the Customer ID which is considered a field Name.

User Interface logs can be obtained using recording tools. Note that the UI event log recording is not in the scope of the methodology presented in this paper.

3.2. User Interface Log Transformation

In this study, we are concerned only with the control-flow perspective of process discovery techniques of process mining to discover a process model showing the interaction of a user with software applications. The control-flow perspective is a category of analysis that consists of discovering the sequences of tasks in a business process. By analyzing how activities are following each other in the event data, we can obtain a model showing the real behavior of the process [39]. The control-flow perspective of process discovery techniques uses the activity name and the corresponding event timestamps to discover a process model. Several action types such as *open*, *go to cell*, *click button*, *enter* are illustrated in the presented user interface log example (Table 1). If we consider the action type as a task and discover a process model using process mining techniques based only on the action type and the corresponding timestamps, we will not get a proper process model which shows the reality. In Table 1, for example, there are two actions of type "open". The first "open" is different from the second "open" action. The first one refers to the opening of a system folder with the name "Orders" and the second "open" refers to the opening of an Excel file with the name "Order Qty". The goal from using process discovery techniques of process mining for robotic process automation is to get insights not about how many times the user performed the action "open" or how many times he performed the "copy" action with the mouse but to get insight on how many times he performed the action

“open” on the same folder or file, or how many times he did the “copy” and “paste” of the same column of the same worksheet or the same section of a web or a system. Therefore, a transformation of the user event log is needed to obtain the actions name based on which the discovery techniques will be performed.

This work defines transformation rules for transforming the name of the original action which is the action type name illustrated in Table 1 into an action name having sufficient information for discovering the tasks model describing the sequence of actions performed on a user interface. For this, let:

- TA be the Transformed Action,
- OA be the Original Action (Action Type),
- ST be the Source Type,
- SN be the Source Name,
- C be the Content,
- FN be the Field Name, and
- FV be the Field Value.

The transformation rules are based on the information available in the user event log that we defined in Table 1. The rules differ based on the action type. We defined a set of action types and a set of transformation rules in Table 2. For instance, the action type “open” needs the source type and the source name to know what exactly has been opened. Thus, the action type which is the original action OA is transformed as follows: $TA = OA + ST + “SN”$. The source type and the source name have been attached to the original action, and it became “open folder system “Orders”” instead of just “open”. This way, we will be able to identify how many times the task “open folder system “Orders”” has been performed when discovering the tasks model. For instance, the action type “Click Button” needs the information in the Content section of the UI Event log to know which button has been clicked. Thus, after using the transformation rule $TA = OA + C$, the original action “Click Button” will be “Click Button Filter”.

Table 2. Interface Log Transformation Rules.

Action Category	Original Action	Transforming Rule
Open	Open	$TA = OA + ST + “SN”$
Close	Close	$TA = OA + ST + “SN”$
Go To	Go to Sheet	$TA = OA + ST + “SN”$
	Go to Cell	$TA = OA + “FN”$
	Go To URL	$TA = OA + “SN”$
	Go to Messenger	$TA = OA + “SN”$
	Go to Column	$TA = OA + “FN”$
	Go to chat	$TA = OA + “SN”$
Click	Click Button	$TA = OA + C$
	Click field	$TA = OA + “FN”$
	Click Item	$TA = OA + “C”$
Select	Select	$TA = OA + FN + “FV”$
Copy	Copy	$TA = OA + “FN”$
Paste	Paste	$TA = OA + “FN”$ Value
Enter	Enter FV	$TA = OA + FN + “FV”$
	Enter Message	$TA = OA$
	Enter Message	$TA = OA$
Log to	Log to	$TA = OA + “SN”$

TA, Transformed Action; OA, Original Action; ST, Source Type; SN, Source Name; C, Content; FN, Field Name; FV, Field Value.

After applying the transformation rules, the user interface log depicted in Table 1 will be transformed into the UI log shown in Table 3. The transformed log is composed of case,

action, and timestamp. Now, the action and all elements involved in performing the action are gathered in one column named Action or Step.

Table 3. Example of Transformed User Interface log.

Case ID	Action	Timestamp
1	Open System Folder "Orders"	2020-08-20 9:29:10
1	Open Excel "Order Qty"	2020-08-20 9:29:16
1	Go to Cell Sheet 1	2020-08-20 9:29:18
1	Click Button "Filter Button"	2020-08-20 9:29:23
2	Enter Web https://www.erp.kor , accessed on 22 July 2021	2020-08-20 9:33:27

After transformation, the UI log refers to a case, an action, and a timestamp. It can be seen as a collection of cases where a case can be seen as a sequence of events.

Definition 1 (Traces). S is the universe of steps or actions. A trace $t \in S^*$ is a sequence of steps. $T = S^*$ is the universe of traces. Trace $t = \langle \text{Open system folder, Open Excel Order Qty, Go to cell sheet 1, Click filter button} \rangle \in T$ refers to 4 events that belong to the same case (case 1 in Table 1). A UI log is a collection of cases. Each case is represented by a trace.

Definition 2 (UI Log). $L = B(T)$ refers to the universe of UI logs. A UI log $U_1 \in L$ represents a finite multiset of observed traces. A UI log represents a multiset of traces. For instance, a UI log $U_1 = [\langle \text{Open system folder, Open Excel Order Qty, Go to cell sheet 1, Click filter button} \rangle, \langle \text{Open URL kerp.com, login, click dashboard, \dots, close URL kerp.com} \rangle, \langle \text{open excel orders, copy column customer, past column customer, \dots, close excel orders} \rangle]$ represents 3 cases (i.e., $|U_1| = 3$).

3.3. Relevance to Work-Based Filtering

Office employees are performing their work on a computer. However, while using a computer, one can do other actions on the user interface that are not related to work too. For instance, sending messages on SNS applications. Thus, irrelevant actions to work need to be filtered. This can be done by defining first the list of software applications, systems, and URLs related to work and then performing filtering which removes the cases containing information on actions performed on user interfaces that are not included in the list.

3.4. Tasks Discovery from the Transformed UI log

Process discovery techniques allow, specifically, the control-flow perspective, i.e., the identification of how tasks belonging to a business process are following each other based on the recorded events, and thus provide as result the process model, which shows the full behavior and allows us to have a full understanding of the process behavior. Even though process mining is considered as recent research, there are plenty of process discovery algorithms that have been developed today which discover the sequential behavior of a process and have been successfully applied to different domains [26,40,41]. For example, the α algorithm [42], the fuzzy miner [43], the heuristic miner [44], the genetic miner [45], the heuristic rule-based algorithm [46], and others [40,47], etc. All these techniques use different methodologies to arrive at the same result which is a model that illustrates the transitions between activities, and they take an event log as input to discover the process model. An example of a discovered process model is depicted in Figure 5.

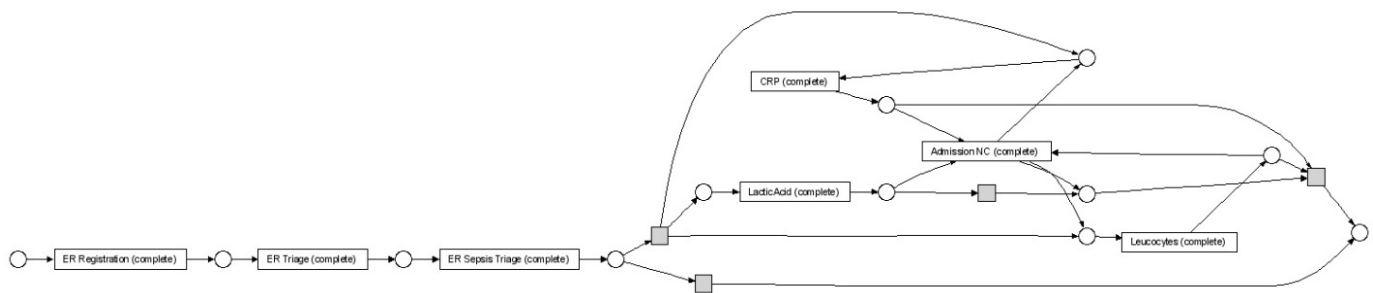


Figure 5. Example of a process model discovered by heuristic-rule-based algorithm [19].

In this study, we would like to use process discovery algorithms but this time by taking a user interface log as input to discover the tasks model which shows the sequence and the transitions between the actions that a user has performed on interfaces of systems and software applications. The approach of this study consists of applying process discovery techniques on the transformed user interface log. An example of a transformed UI log is depicted in Table 3.

3.5. Candidate Tasks Selection Criteria

After applying a process discovery technique on the transformed UI log, one model will be constructed. This model may contain tremendous cases. For example, (1) performing many operations in sequence on an excel file such as downloading an excel file, sorting data, filtering data, copy and paste data, etc., (2) Opening and closing a folder, (3) sending messages in SNS applications, etc. One model will contain all the actions that were performed in the user interface. A selection of the relevant and candidate cases needs to be performed to reduce the number of discovered cases. Based on the candidate routines, a business manager can decide those are relevant for automation. For this purpose, we defined three criteria: Frequency, Periodicity, and Duration.

Frequency. The aim of RPA is to automate repetitive routines. Routines refer to the same tasks that are performed many times. In other words, performing specific tasks frequently. Using extension techniques of process mining, we can enrich the discovered model with frequency. By applying the case frequency technique, we can display a model showing the frequency of each task and the frequency of each transition from one task to another. Based on this, frequent cases can be selected.

Periodicity. The goal of using periodicity selection criteria is to identify and select periodic cases. Periodic cases are performed frequently but periodically. They are performed every time at the same time (i.e., every Friday).

Unattended bots can be applied on periodic cases if they do not need an intervention of a human. The bots can be scheduled to perform them.

Duration. Identifying cases and tasks that are taking a long time can play a crucial role in the decision of the routines that need to be automated. Routines that take hours by an employee can be performed by RPA bots in milliseconds.

To calculate the duration of a discovered task, the duration of all cases referring to the corresponding task need to be calculated first. Then, the mean duration of the cases related to each specific task is calculated. We defined the formula for calculating the duration of a case in Equation (1) and the formula for calculating the mean duration of a task in Equation (2).

$$T_{Case} = \sum_{i=0}^n t_{s_i} + t_{s_i s_{i+1}} \quad (1)$$

T_{Case} : case duration, the time spent to perform a case,
 t_{s_i} : step duration, the time spent to perform a step,

$t_{s_i s_{i+1}}$: the waiting time between the previous and the next step.

$$Mean_Duration(task_a) = \frac{1}{n} \sum_{i=0}^n Tc_i^{task_a} \quad (2)$$

$task_a$: a given task,

$Tc_i^{task_a}$: the time spent to perform a case related to a specific task $task_a$

n : the total number of cases referring to $task_a$

4. Results and Discussion

In this section, we present a use case to illustrate the methodology presented in this study. This section shows how candidate tasks can be selected for automation with RPA.

■ User Interface Log Generation and Pre-processing

The starting point of robotic process mining is the preparation of the data. This step consists of the identification of actions related to the interaction of the user with the user interfaces and their attributes such as timestamps and case id, etc. To obtain a log, we have recorded the user interface interaction for 8 days, every day for two hours morning. Then we generated a user interface log containing the following columns: timestamps, action type, source type, source name, context, field name, and the field value. After that, we transformed the UI log into a log supported by process discovery techniques using the transformation rules presented in Table 2. The transformed UI log contains actions and their timestamps. Finally, we removed from the transformed log any action containing applications or URLs related to sending messages and chatting considered not relevant to work. After finishing the preprocessing of the user interface log, we imported it into the Disco tool [48] and visualized it. The imported user interface log consists of 50 different activities, 11 cases with 113 events.

■ User Interface interactions discovery

Process discovery techniques take an event log as input and automatically construct a process model. The main idea behind the process discovery is to find the as-is process model. This phase aims to discover using process discovery techniques the cases or the interaction that have been performed in a user interface. In this study, we used the famous fuzzy miner as an algorithm for the discovery. We applied the fuzzy miner [31] using the Disco tool [48] on the transformed and filtered UI log. The discovered model of user interactions with the user interface is illustrated in Figure 6. The cases shown in the discovered model are highlighted separately in Figures 7 and 8 to see clearly process cases. The discovered model shows five tasks consisting of a sequence of steps (action) derived.

■ Candidate Tasks Selection for RPA

In this section, we assessed the derived model based on three criteria: frequency, periodicity, and duration for selecting candidate tasks for automation via RPA tools.

■ Frequency Based Selection

The numbers, the thickness of transitions, and the coloring in the discovered model shown in Figure 6 depict how frequently each step (action) and path are executed. The frequency of actions is shown inside the box representing a step of a task. The frequency of transition from one action to another is specified on the edge connecting two actions. The darker the color is the more frequent the action and the transition are. Accordingly, we can see that Figure 7 shows a frequent case while Figure 8 shows infrequent cases.

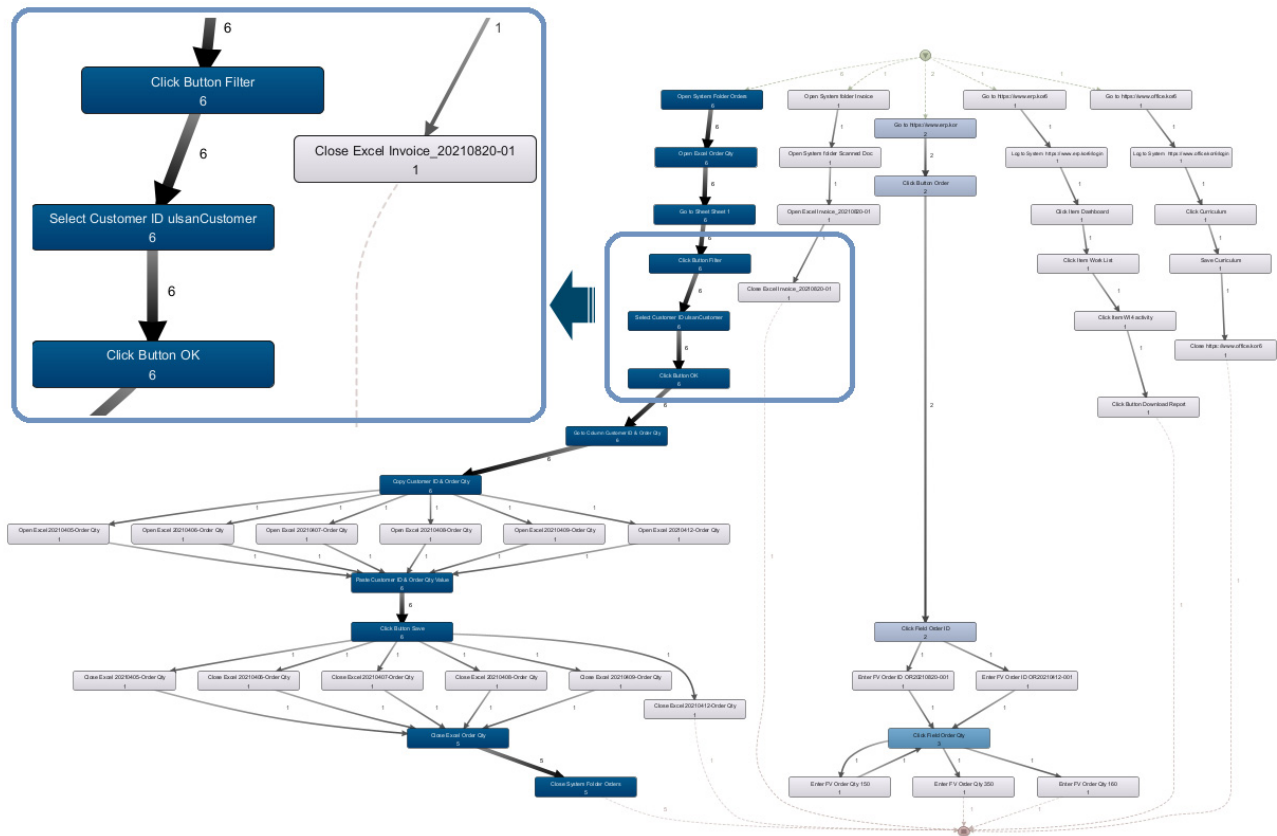


Figure 6. User interface interactions discovered with fuzzy miner using Disco tool.

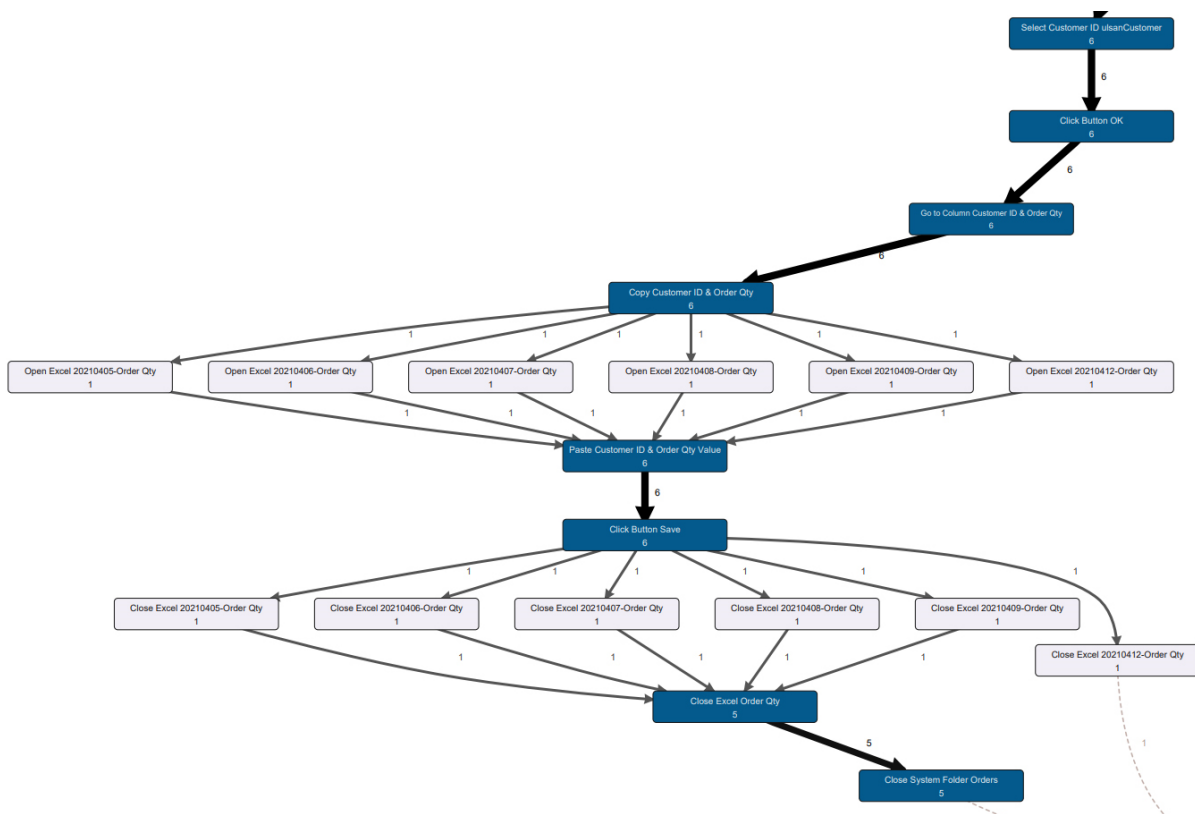


Figure 7. Sequence of Frequent steps.

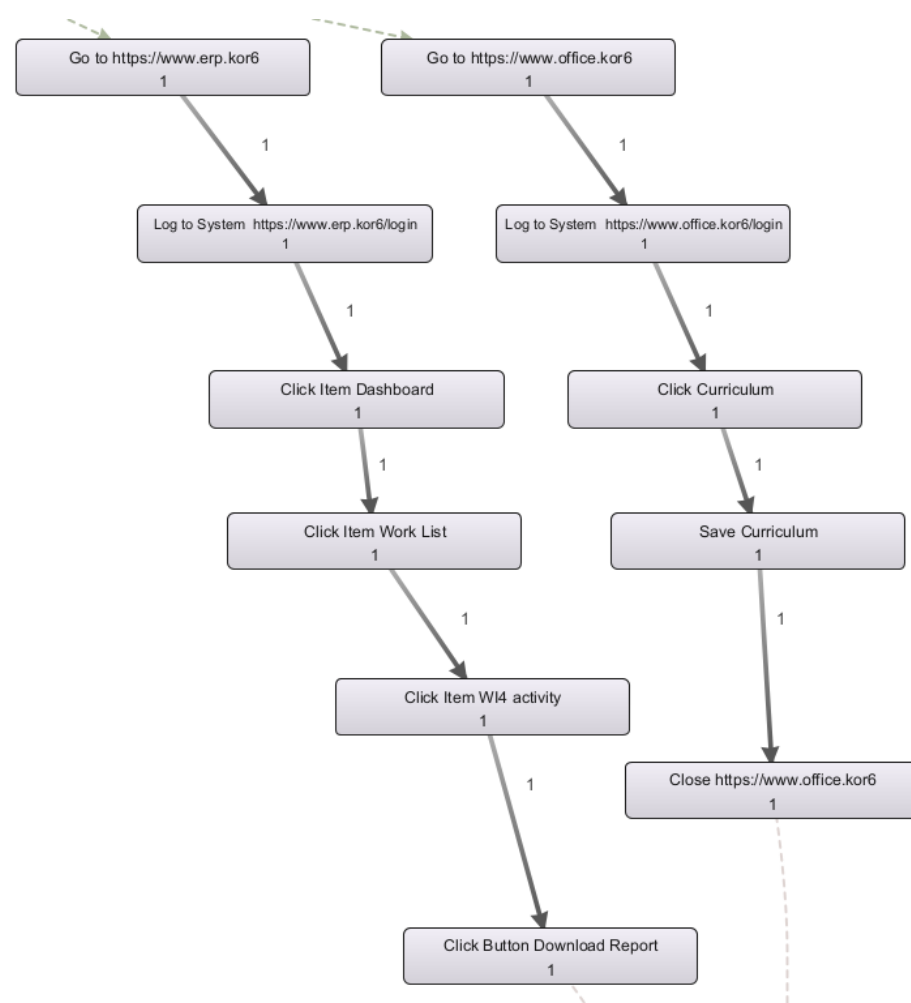


Figure 8. Sequence of Infrequent steps.

Figure 7 shows one discovered task consisting of a sequence of steps represented with a darker color. This indicates that those steps are frequently performed. Figure 7 also shows some steps that are not frequent. If we look at the steps that are not frequent, we see that they are all executed right after a frequently executed step. By taking a careful look at the frequently executed steps in Figure 7—which are Go to URL, Click Button Order, Click Field Order, Click Order Qty, Click Field Total, and at the infrequent tasks which are “Open Excel 20210405-Order Qty”, “Open Excel 20210406-Order Qty”, “Open Excel 20210407-Order Qty”, etc.—we can see that the infrequent steps are the same steps but the entered value is different. They are performed directly after the same frequent step every time. Therefore, this case which consists of a sequence of frequently executed steps and of infrequent steps executed directly after the frequent ones can be classified as a frequently executed case and can be considered as a routine and can be considered a candidate for automation. In contrast, Figure 8 shows a discovered task consisting of a sequence of steps with frequencies equal to 1, which means that the performed task is not usually performed. Thus, this type of task cannot be considered as a candidate for automation with RPA.

Note. Figures 6–8 are presented in detail in the Supplementary File.

Table 4 depicts the number of events and the frequency of each sequence of steps discovered in the model shown in Figure 6. We call each sequence of steps a task. The most frequently executed task is Task #1 with a frequency equal to 6. This was followed by Task #2 with a frequency equal to 2. The infrequent tasks were Task #3, Task #4, and Task #5. Their frequency is equal to 1. The most frequent task is Task #1 contains 15 events (i.e., 15 steps) that are executed 6 times. Since the user interface is recorded for one week from

5 April 2021 to 12 April 2021 excluding the weekend, and the frequency is equal to 6, we can conclude that Task #1 is executed every day. Accordingly, Task #1 can be considered as a candidate for automation with RPA. Task #2 is executed through six steps and was not executed every day. At the same time, it cannot be considered infrequent because it was performed twice within eight days. This task needs further investigation which will be outlined in the next section. Based on frequency, it is clear Task #3, Task #4, and Task #5 cannot be a candidate for automation with RPA. These tasks are performed only once. They are irregular tasks and irregular tasks are not the target of automation with RPA.

Table 4. Discovered Tasks Frequency.

Tasks	Related Cases	Number of Events	Frequency
Task #1	Case #2, Case #5, Case #6, Case #9, Case #11	14	6
Task #2	Case #3, Case #7, Case #8	6	2
Task #3	Case#1	4	1
Task #4	Case #10	5	1
Task #5	Case #4	7	1

■ Periodicity Based Selection

A periodic task is a task that repeats itself after a fixed time interval. Identifying periodic tasks and their periodicity is very important for identifying how and when to execute the robot in charge of executing the corresponding tasks. If a task detected to be a candidate for automation with RPA through “frequency”, is identified to be periodic, the robot can be configured such that its execution is triggered by scheduling the time of execution based on the detected periodicity. Table 5 shows the periodicity of the tasks discovered in the model shown in Figure 6. Task #1 has been detected to be frequent in the previous section with a frequency equal to 6. This frequency was equal to the number of days from which the data was obtained. Since it has been executed every day, we can consider the periodicity parameter of task #1 to be “every day”. The frequency of performing Task #2 was identified to be equal to two. By investigating the data result, we found that this task is executed every Monday based on the period of the retrieved data. The rest of the tasks are not periodic since they were executed only once during the period of the extracted data. In conclusion, Task #1 and Task #2 can be considered a candidate for automation with RPA. The execution of Task #2 can be scheduled to be executed every Monday by an unattended robot in case it does not require human intervention.

Table 5. Discovered Tasks Periodicity.

Table	Related Cases	Periodic	Periodicity
Task #1	Case #2, Case #5, Case #6, Case #9, Case #11	Yes	Every day
Task #2	Case #7, Case #8, Case #3	Yes	Every Monday
Task #3	Case#1	No	-
Task #4	Case #10	No	-
Task #5	Case #4	No	-

■ Duration Based Selection

The Mean duration of tasks can play a primordial role in choosing between candidate tasks for automation. Automating a frequent task that is taking time to be fulfilled can be more valuable than automating a task that can be performed in just seconds or minutes. The aim is to provide much information that can help in deciding between candidate tasks for automation. For our use case, the duration of all cases has been calculated. After that, the mean duration has been calculated for each task. Figure 9 depicts the duration of tasks.

As can be seen, Task #1 takes the longest time. This task has been already detected to be the most frequent in the previous section. Thus, automating this task can save a lot of time.

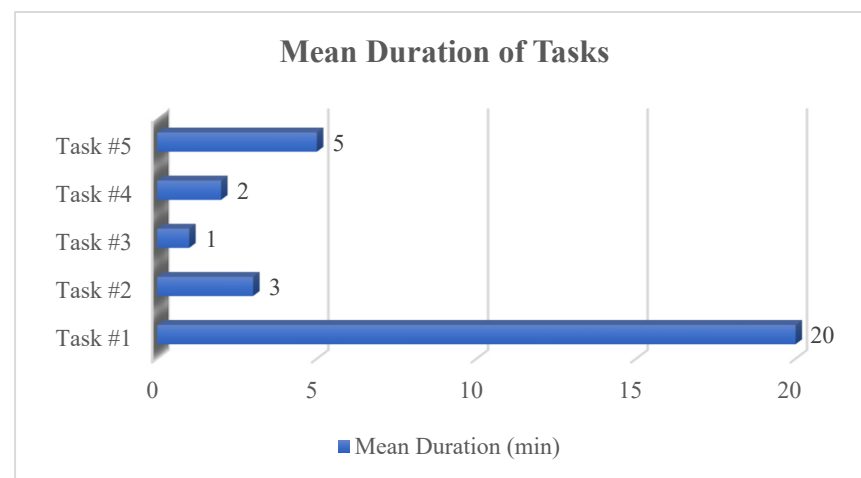


Figure 9. Mean duration of tasks.

5. Conclusions and Limitations

Robotic process automation (RPA) is a new wave of new digital technologies that are increasingly capturing the attention of administrations and industries. RPA aims to automate manual, repetitive, rule-based, and un motivating activities performed on a computer. RPA uses software robots to replace specific human administrative tasks. The goal is to allow human workers to delegate their tedious routine tasks to a software bot to allow them to focus more on difficult tasks. RPA tools have been demonstrated as simple and very powerful in cost-saving and other performances. However, to be able to implement RPA, we need first to effectively identify the tasks that are suitable for automation before applying RPA. Therefore, this work introduced an approach for selecting candidate tasks to be automated with RPA. The proposed methodology is based; using process mining techniques; on discovering tasks consisting of a sequence of steps performed in a user interface from a UI log generated from recording the performed actions while interacting with the user interface. This work has shown (1) what information a log should contain to be able to derive tasks, (2) how the log should be transformed to be ready to be used by process mining techniques, (3) how we can discover digital administrative tasks, and how candidate tasks can be selected for automation from the discovered tasks. As a real use case implementation is required for validation, we will consider real event data to validate the proposed framework in future work. As research of using process mining with RPA is so recent, there are some challenges that need to be addressed to be able to use process mining techniques properly to identify candidate tasks for automation with RPA. This study outlines some encountered challenges that need to be tackled in future works.

5.1. Challenge 1: Generating Event Log from Recorded Interactions with the User Interface

An event log is a necessary input for process mining techniques. User interface logs cannot be generated by today's existing systems. The only way is to record the interactions with the user interface and then generate a user interface log. The interaction with a user interface is based on clicks with the mouse and content entered with the keyboard. For instance, for us, an action or a step would be opening a google URL. This action is recorded with mouse clicks and URL content entered with the keyboard. When entering for instance "google.com" in the URL section of the browser with the keyboard, it is recorded one letter by letter, and each entered letter is recorded in a separate row as follows "g", "o", "o", "g", "l", "e", ".", "c", "o", "m". Accordingly, we will have 10 data rows created just for entering "google.com" in the URL section of the browser. An action that we perform in a

user interface that seems for a person as one action, is in fact performed by several actions by the system. There is a need to translate or transform the several actions (for the system) that were recorded into one action for us. This is just one example from many cases. Hence, one of the main challenges is to identify how to transform the recording of user interface interaction based on mouse and clicks into a user interface log that can be used by process mining techniques.

5.2. Challenge 2: Case Identification

In general, an event log refers to a case, an activity, and a timestamp. It can be seen as a collection of cases where a case can be seen as a sequence of events. To be able to analyze processes with process mining, there is a need to first identify what the appropriate case ID is for your process. Table 6 shows an example of an event log generated from an ERP system to be analyzed with process mining. Each row corresponds to an event. In this example, the Order number will be the appropriate case ID to discover the corresponding process model. A case ID usually is unique with which different activities are associated. The performed tasks that are having the same Order number will be in one case. However, the situation of a user interface log is different. A user interface log is generated from recording interactions of a user with user interfaces which are based on clicks with the mouse and entering with the keyboard. The user interface log is different from the general event log generated from today's information systems (e.g., ERP, BPM, etc.). We defined in Table 1 the data that should be in the log to be able to discover tasks from it using process mining. When generating a user interface log, we might not have a unique ID with which many actions or steps performed in a user interface are associated. Therefore, the main challenge or the main question is how to define or identify an appropriate case ID of a user interface log. This challenge needs to be addressed in future works to be able to use process mining for selecting candidate tasks for automation with RPA.

Table 6. Example of an event log generated from ERP system to be analyzed with process mining.

Order Number	Task	Timestamp	User	Product	Quantity
8801	Register order	21-05-01 9.15	Soukaina	Fan78	15
8802	Register Order	21-05-02 9.18	Soukaina	Heater100	20
8801	Check Stock	21-05-01 9.25	Younes	Fan78	15
8803	Register Order	21-05-05 9.22	Soukaina	Frotter05	10
8802	Check Stock	21-05-02 9.30	Younes	Heater100	15
8801	Pack Order	21-05-01 10.30	Mohammed	Fan78	15
8802	Pack Order	21-05-02 11.10	Mohammed	Heater100	20
8801	Ship Order	21-05-01 15.20	Safae	Fan78	15

5.3. Challenge 3: Case Duration

A manual office task that is taking a long time can be performed in just a few seconds if automated with RPA. This will allow saving a lot of time. Hence, calculating the total time spent in performing a specific case can play a huge role in identifying the office tasks that need to be automated. In a user interface, any case can be started with opening an object (e.g., URL, system, etc.) and completed with closing the same object. The duration can be calculated by calculating the time between opening the object until closing it. However, one can open an object to start performing a task and finish performing the task without close the object immediately. In this case, the time calculated from opening the object until closing it will not reflect the real duration spent in performing the task. The challenge here is how to calculate the duration of tasks by taking into consideration real-life situations.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/su13168980/s1>, Figure 6: User interface interactions discovered with fuzzy miner using Disco tool, Figure 7: Sequence of Frequent steps, Figure 8: Sequence of Infrequent steps.

Author Contributions: H.R. and D.C. contributed to the main idea and the methodology of the research. H.R. and D.C. designed the experiment, performed the simulations, and wrote the original manuscript. H.R. contributed significantly to improving the technical and grammatical contents of the manuscript. H.R. and C.C. reviewed the manuscript and provided valuable suggestions to further refine the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

- Coccia, M. Asymmetry of the technological cycle of disruptive innovations. *Technol. Anal. Strat. Manag.* **2020**, *32*, 1462–1477. [CrossRef]
- Coccia, M. Disruptive firms and technological change. *Quad. IRCrES-CNR* **2018**, *3*, 3–18. [CrossRef]
- Coccia, M. Disruptive firms and industrial change. *J. Econ. Soc. Thought* **2017**, *4*, 437–450. [CrossRef]
- Coccia, M.; Watts, J. A theory of the evolution of technology: Technological parasitism and the implications for innovation management. *J. Eng. Technol. Manag.* **2020**, *55*, 101552. [CrossRef]
- Coccia, M. The theory of technological parasitism for the measurement of the evolution of technology and technological forecasting. *Technol. Forecast. Soc. Chang.* **2019**, *141*, 289–304. [CrossRef]
- Coccia, M. A theory of classification and evolution of technologies within a Generalised Darwinism. *Technol. Anal. Strat. Manag.* **2018**, *31*, 517–531. [CrossRef]
- Siderska, J. The Adoption of Robotic Process Automation Technology to Ensure Business Processes during the COVID-19 Pandemic. *Sustainability* **2021**, *13*, 8020. [CrossRef]
- Matt, C.; Hess, T.; Benlian, A. Digital Transformation Strategies. *Bus. Inf. Syst. Eng.* **2015**, *57*, 339–343. [CrossRef]
- Leopold, H.; Van Der Aalst, H.; Reijers, H.A. Identifying Candidate Tasks for Robotic Process Automation in Textual Process Descriptions. *Bus. Inf. Syst.* **2018**, *318*, 67–81. [CrossRef]
- Balakrishnan, S.; Hameed, M.S.; Venkatesan, K.; Aswin, G. An Exploration of Robotic Process Automation in all Spans of Corporate Considerations. In Proceedings of the 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 19–28 March 2021; Volume 1, pp. 1881–1884.
- Madakam, S.; Holmukhe, R.M.; Jaiswal, D.K. The Future Digital Work Force: Robotic Process Automation (RPA). *J. Inf. Syst. Technol. Manag.* **2019**, *16*. [CrossRef]
- Choi, D.; R'bigui, H.; Cho, C. Robotic Process Automation Implementation Challenges. In Proceedings of the International Conference on Smart Computing and Cyber Security—Lecture Notes in Networks and Systems, Kyungdong University, Global Campus, Korea, 7–8 July 2020; pp. 297–304. [CrossRef]
- Nawaz, N. Robotic Process Automation for Recruitment Process. *Int. J. Adv. Res. Eng. Technol.* **2019**, *10*, 608–611. [CrossRef]
- Van der Aalst, W.M.P. On the Pareto Principle in Process Mining, Task Mining, and Robotic Process Automation. *DATA* **2020**, 5–12. [CrossRef]
- Leshob, A.; Bourgoquin, A.; Renard, L. Towards a Process Analysis Approach to Adopt Robotic Process Automation. In Proceedings of the 2018 IEEE 15th International Conference on E-Business Engineering, Xi'an, China, 12–14 October 2018; pp. 46–53.
- Gejke, C. A new season in the risk landscape: Connecting the advancement in technology with changes in customer behaviour to enhance the way risk is measured and managed. *J. Risk Manag. Financ. Inst.* **2018**, *11*, 148–155.
- Choubey, A.; Sharma, M. Implementation of robotics and its impact on sustainable banking: A futuristic study. *J. Phys. Conf. Ser.* **2021**, *1911*, 012013. [CrossRef]
- Aguirre, S.; Rodriguez, A. Automation of a Business Process Using Robotic Process Automation (RPA): A Case Study. *Commun. Comput. Inf. Sci.* **2017**, *742*, 65–71. [CrossRef]
- Pokharkar, A. Robotic Process Automation: Concept, Benefits, Challenges in Banking Industry IIBM's. *J. Manag.* **2019**, *4*, 17–25. [CrossRef]
- Van der Aalst, W.M.P. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2011.
- R'bigui, H.; Al-Absi, M.A.; Cho, C. Detecting complex control-flow constructs for choosing process discovery techniques. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *9*, 1389–1393.
- Kommerer, V. Robotic Process Automation. *Am. J. Intell. Syst.* **2019**, *9*, 49–53. [CrossRef]
- Robotic Process Automation Market Size, Share & Trends Analysis Report by Type, by Service, by Application, by Deployment, by Organization, by Region, and Segment Forecasts, 2021–2028; Report Id: GVR-1-68038-145-0. April 2021. Available online: <https://www.grandviewresearch.com/industry-analysis/robotic-process-automation-rpa-market> (accessed on 22 June 2021).
- Zhang, X.; Wen, Z. Thoughts on the development of artificial intelligence combined with RPA. *J. Phys. Conf. Ser.* **2021**, *1883*. [CrossRef]

25. Gavrilovskaya, N.V.; Kuvaldin, V.P.; Zlobina, I.S.; Lomakin, D.; Suchkova, E. Developing a robot with computer vision for automating business processes of the industrial complex. *J. Phys. Conf. Ser.* **2021**, *1889*, 022024. [[CrossRef](#)]
26. R'Bigui, H.; Cho, C. Customer Order Fulfillment Process Analysis with Process Mining: An Industrial Application in a Heavy Manufacturing Company. In Proceedings of the ACM International Conference on Computer Science and Artificial Intelligence, Jakarta, Indonesia, 5–7 December 2017; pp. 247–252.
27. Van der Aalst, W.M.P. *Process Mining: Data Science in Action*; Springer: Berlin/Heidelberg, Germany, 2016.
28. Van der Aalst, W.M.P. Object-Centric Process Mining: Dealing with Divergence and Convergence in Event Data. In *Software Engineering and Formal Methods; SEFM 2019—Lecture Notes in Computer Science*; Ölveczky, P., Salaün, G., Eds.; Springer: Cham, Switzerland, 2019; Volume 11724. [[CrossRef](#)]
29. Van der Aalst, W.M.P.; Bichler, M.; Heinzl, A. Robotic Process Automation. *Bus. Inf. Syst. Eng.* **2018**, *60*, 269–272. [[CrossRef](#)]
30. IEEE Corporate Advisory Group. *IEEE Guide for Terms and Concepts in Intelligent Process Automation*; IEEE: Piscataway, NJ, USA, 2017.
31. Van der Aalst, W.M.P.; Hee, K. *Workflow Management: Models, Methods, and Systems*; MIT Press: Cambridge, MA, USA, 2004.
32. Lacity, M.; Willcocks, L.P. Robotic process automation at telefonica O2. *MIS Q. Exec.* **2016**, *15*, 1.
33. Leno, V.; Polyvyanyy, A.; Dumas, M.; La Rosa, M.; Maggi, F.M. Robotic Process Mining: Vision and Challenges. *Bus. Inf. Syst. Eng.* **2021**, *63*, 301–314. [[CrossRef](#)]
34. Leno, V.; Dumas, M.; Rosa, M.; Maggi, F.; Polyvyanyy, A. Automated Discovery of Data Transformations for Robotic Process Automation. *arXiv* **2020**, arXiv:2001.01007.
35. Gulwani, S. Automating string processing in spreadsheets using input-output examples. *ACM SIGPLAN Not.* **2011**, *46*, 317–330. [[CrossRef](#)]
36. Geyer-Klingeberg, J.; Nakladal, J.; Baldauf, F.; Veit, F. Process mining and Robotic process automation: A perfect match. *CEUR Workshop Proc.* **2018**, *2196*, 124–131.
37. Jimenez-Ramirez, A.; Reijers, H.A.; Barba, I.; Del Valle, C. A Method to Improve the Early Stages of the Robotic Process Automation Lifecycle. In Proceedings of the International Conference on Advanced Information Systems Engineering (CAiSE), Rome, Italy, 3–7 June 2019; pp. 446–461.
38. Ketkar, Y.; Gawade, S. Effectiveness of Robotic Process Automation for data mining using UiPath. In Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; pp. 864–867.
39. Ferreira, D.R. Control-Flow Perspective. In *A Primer on Process Mining*; Springer: Cham, Switzerland, 2020. [[CrossRef](#)]
40. R'Bigui, H.; Cho, C. Purchasing Process Analysis with Process Mining of a Heavy Manufacturing Industry. In Proceedings of the 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, 17–19 October 2018; pp. 495–498.
41. Werner, M.; Wiese, M.; Maas, A. Embedding process mining into financial statement audits. *Int. J. Account. Inf. Syst.* **2021**, *41*, 100514. [[CrossRef](#)]
42. Van der Aalst, W.; Weijters, A.T.; Maruster, L.L. Workflow mining: Discovering process models from event logs. *IEEE Trans. Knowl. Data Eng.* **2004**, *16*, 1128–1142. [[CrossRef](#)]
43. Günther, C.W.; Van der Aalst, W.M.P. Fuzzy Mining—Adaptive Process Simplification Based on Multi-Perspective Metrics. In *Business Process Management (Lecture Notes in Computer Science)*; Alonso, G., Dadam, P., Rosemann, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 328–343. [[CrossRef](#)]
44. Weijters, A.; Ribeiro, J.T.S. Flexible heuristics miner (FHM). In Proceedings of the 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Paris, France, 11–15 April 2011; pp. 310–317.
45. Van der Aalst, W.M.P.; de Medeiros, A.K.A.; Weijters, A.J.M.M. Genetic Process Mining. In *Applications and Theory of Petri Nets 2005; ICATPN 2005—Lecture Notes in Computer Science*; Ciardo, G., Darondeau, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; Volume 3536. [[CrossRef](#)]
46. R'Bigui, H.; Cho, C. Heuristic rule-based process discovery approach from events data. *Int. J. Technol. Policy Manag.* **2019**, *19*, 352. [[CrossRef](#)]
47. Sun, H.; Liu, W.; Liang, Q.; YuYue, D.; Xiaojun, R.; XinYing, L. A process mining algorithm to mixed multiple-concurrency short-loop structures. *Inf. Sci.* **2021**, *542*, 453–475. [[CrossRef](#)]
48. Gunther, C.; Rozinat, A. Disco: Discover your processes. *Proc. CEUR Workshop* **2012**, *940*, 40–44.