

Enfoques y técnicas para el descubrimiento de procesos robotizables usando Process Mining. Una revisión bibliográfica.

Approaches and techniques for discovering robotizable processes using Process Mining: A literature review.

Jorge Nicasio Cáceres Zárate  

Universidad Nacional Mayor de San Marcos, Lima, Perú.

Recibido: 03/10/2024 Revisado: 20/01/2025 Aceptado: 15/05/2025 Publicado: 30/06/2025

RESUMEN

Las empresas han estado centradas en automatizar procesos para mejorar la eficiencia operativa, para ello necesitan conocer el detalle del proceso que se debe de robotizar, pero muchas no han tenido definido por diversos motivos como desconocimiento, complejidad, desinterés, etc. Este problema es solucionado con la minería de procesos que utiliza registros de la interfaz de usuario para identificar y analizar procesos empresariales, desempeñando un papel crucial al ayudar a identificar procesos adecuados para la automatización. Este artículo profundiza en revisar enfoques y técnicas para la minería de procesos orientada al descubrimiento de procesos que puedan ser robotizados. Se utiliza la metodología PRISMA con criterios de inclusión y exclusión, basándose en fuentes indexadas para responder preguntas de investigación específicas. Se concluye conociendo el formato de las tramas para los registros de eventos, herramientas comerciales, híbridas y propias para extracción de datos, técnicas de filtrado y clusterización, algoritmos y técnicas para el descubrimiento de procesos y los enfoques para determinar si un proceso es robotizable. Estos resultados respaldan la evolución hacia una mayor automatización en el ámbito empresarial.

PALABRAS CLAVE:

Automatización Robótica de Procesos, Minería de procesos, Descubrimiento de procesos, PRISMA, Registro de eventos.

ABSTRACT

Companies have been focused on automating processes to improve operational efficiency. To achieve this, they need to understand the details of the processes that should be robotized, but many have not defined them due to various reasons such as lack of knowledge, complexity, lack of interest, etc. This problem is addressed by process mining, which uses user interface logs to identify and analyze business processes, playing a crucial role in helping to identify processes suitable for automation. This article delves into reviewing approaches and techniques for process mining aimed at discovering processes that can be robotized. The PRISMA methodology is used with inclusion and exclusion criteria, based on indexed sources to answer specific research questions. It concludes by discussing event log formats, commercial, hybrid, and proprietary tools for data extraction, filtering and clustering techniques, algorithms and techniques for process discovery, and approaches to determine if a process is suitable for

automation. These findings support the evolution towards greater automation in the business environment.

KEYWORDS:

Robotic Process Automation, Process mining, Process discovery, PRISMA, Event logging.

Introducción

Las empresas en la actualidad tienen diversos flujos de negocio operativos y repetitivos que son realizado por sus trabajadores en el día a día que podrían realizarse de manera más eficiente por un robot conociendo el listado de pasos a ejecutar (Martínez-Rojas et al., 2022), es bajo esta necesidad que desde principios de siglo ha surgido un movimiento a nivel mundial llamado automatización de procesos robotizables (RPA), se ha evidenciado un aumento en el uso de esta técnica en las empresas por los beneficios competitivos mejorando la eficiencia, reduciendo costo y tiempo (Choi et al., 2022 ; Jimenez-Ramirez et al., 2019).

Leno et al., (2021) conceptualiza RPA como una técnica que permite a las empresas automatizar actividades repetitivas a través de la ejecución de scripts que codifican secuencias detalladas de interacciones con aplicaciones. Pero también surgen nuevos problemas debido que al aplicarse trae consigo la necesidad de poder identificar el proceso a robotizar de manera clara y sobre todo que actividades sean las adecuadas. (Šperka & Halaška, 2023)

Pasquadibisceglie et al., (2022) comenta que muchas empresas carecen de un modelo de proceso explícito y frecuentemente no poseen documentación detallada de sus procesos, los cuales existen principalmente como conceptos abstractos, lo que dificulta su automatización. Muchas veces para este descubrimiento de procesos las empresas necesitan de expertos humanos capacitados que identifiquen las rutinas del proceso de sus trabajadores de acuerdo a entrevistas y observación para poder generar el diagrama de negocio (Agostinelli, Lupia, et al., 2022). La falta de coincidencia del diagrama de proceso y el comportamiento real del proceso puede tener numerosas causas, entre ellas la inadecuación del modelo, la mala interpretación del proceso o el desarrollo de un modelo para condiciones ideales (Alizadeh & Norani, 2018; Agostinelli, Marrella, et al., 2022).

Es por eso que no es suficiente modelar los procesos de negocio con metodologías teóricas, sino que hay que complementarlo con el análisis del comportamiento de acciones reales que hacen los usuarios en los procesos para una mayor efectividad usando al registro de acciones o trazas que van dejando a través de los sistemas de información (Choi et al., 2021; Pishgar et al., 2022; Abb & Rehse, 2022; López-Carnicer et al., 2020). Es ahí donde surge un factor esencial para los algoritmos que descubren procesos dentro de la minería de procesos a partir de un registro de eventos que describe el mejor comportamiento de este proceso de negocio implementado en una organización (Effendi & Sarno, 2020a; Tang et al., 2022).

Kim et al., (2021) argumenta que la ventaja competitiva fundamental para las empresas con enfoque en procesos radica en la capacidad de la minería de procesos para gestionar y optimizar de manera continua todos los procesos implementados en las organizaciones. Sin embargo, es crucial contar con herramientas que puedan manejar las diversas situaciones que presentan los procesos empresariales.

Con la disciplina de la minería de procesos tendremos las ventajas para el descubrimiento y modelado de nuevos procesos que sean candidatos para ser robotizables, detectar los cuellos de botellas, información objetiva del ciclo de vida del proceso determinado por las acciones del usuario y sobre todo un flujo óptimo

eliminando ruido o eventos innecesarios que van a mejorar la eficiencia del proceso que se desea robotizar (Leno et al., 2022; Egger, 2020; Augusto et al., 2022).

Una de las preocupaciones importantes con respecto al descubrimiento del modelo de proceso es el problema del ruido o acciones irrelevantes que se genera al descubrir o encontrar un proceso que muchas veces no se evidencia al no compararse con un comportamiento real. El ruido se refiere a los datos que se repiten menos en un cierto número en comparación con los otros datos de entrada. Sin embargo, algunos datos válidos poco frecuentes pueden formar parte de los datos de entrada, que no deben confundirse (Effendi et al., 2021; Effendi & Sarno, 2020b).

Fantinato et al., (2023) observa que dentro de la minería de procesos necesita poder tener mejores algoritmos o técnicas que sirvan para la identificación de procesos con estructuras basada en acciones de usuarios que manejen el ruido y se validen su calidad y eficiencia.

Todas estas implicancias generan un problema que la minería de procesos puede analizar y proponer una solución eficiente para encontrar y descubrir el proceso que se debe de robotizar dentro de la empresa. La complejidad y rigidez de los entornos de las organizaciones hacen que se necesite robotizar procesos mediante herramientas que graben secuencia de acciones que simulen al usuario final y es ahí la importancia por la cual es necesario descubrir mediante acciones deterministas un proceso a robotizar dentro de la empresa (Pasquadibisceglie et al., 2022 ; Bosco et al., 2019 ; Agostinelli et al., 2021).

Es necesario poder plantear un enfoque de patrones de comportamiento de acuerdo a acciones de usuario para descubrir el proceso (Acheli et al., 2022) y conocer algoritmos o soluciones que sepan interpretar situaciones previas y posteriores a las actividades de procesos comerciales estructurados de manera coherente con las acciones de los usuarios de negocio (Al-Absi & R'bigui, 2023).

El presente artículo está dividido en secciones en donde se van a plantear de manera objetiva sustentada con los artículos encontrados los enfoques y técnicas para poder generar la minería de procesos para el descubrimiento de procesos candidatos a robotizar, esta revisión bibliográfica especifica los pasos que se han llevado a cabo para la determinación de los materiales y métodos, resultados y discusión y conclusiones de acuerdo a los artículos revisados en la investigación.

Materiales y métodos

Este estudio estableció un modelo jerárquico para la selección de artículos, utilizando criterios fundamentales con el fin de reducir sesgos y fusionar información existente. Esto permite sintetizar los descubrimientos y ofrecer recomendaciones relacionadas con la pregunta del proceso metodológico.

Se empleó la metodología de revisión sistemática PRISMA, ya que es útil para la valoración crítica de revisiones sistemáticas publicadas, ayuda a los autores en la presentación de informes de revisión sistemática y permite al lector poder revisar el artículo de manera sencilla.

Se ha planteado una búsqueda de artículos científicos tomando bases de datos indexadas de Scopus, Springer, MDPI, Sciencedirect y IEEE. La búsqueda incluye las palabras claves como "Mining for robotic process automation" y "RPA", "Candidate Tasks Selection Methodology for Automation", "UI interaction recorder", "methodology, techniques y RPA", "evento logs" y "algorithm".

Para el proceso de recolección de artículos se definieron criterios de inclusión resumidos en la siguiente **Tabla 1**.

Tabla 1. Criterios de inclusión.

Criterios	Detalle
Documento	Debe estar en un buscador indexado.
Tipos	Solo se aceptan los artículos (journal) y Conference Paper.
Publicación	Desde el 2018 hasta el 2023.
Base de datos	Las fuentes de Scopus, Springer, Sciencedirect, IEEE y MDPI.
Idiomas	Español o inglés
Metodología	Deben tener una técnica o enfoque para minería de datos para registro de IU o minería para el descubrimiento de procesos para robotizar.
Cuartil	1 ,2 y 3

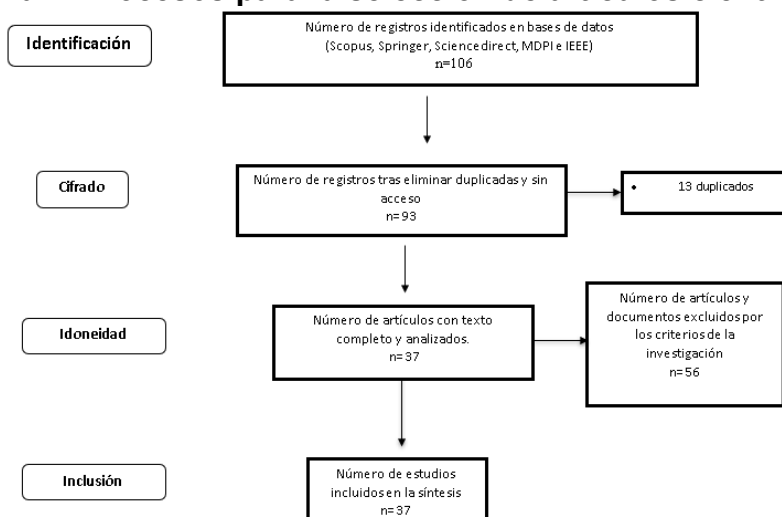
Como parte de los criterios de exclusión se definen en la siguiente **Tabla 2.**

Tabla 2. Criterios de exclusión.

Criterios	Detalle
Tipos	Tesis y artículos de revisiones bibliográficas.
Palabras clave	Artículos que no tengan relación con una palabra clave asociada a la investigación.
Idiomas	Artículos que tengan un idioma diferente a español o inglés.
Cuartil	Documentos con cuartil 4 o no tengan cuartil.

El protocolo de búsqueda y extracción de información de artículos científicos fue realizado de forma independiente.

Siguiendo los pasos que indica la metodología PRISMA en la búsqueda con palabras clave, se obtuvieron los artículos relevantes. Primero, se han extraído de las bases de datos de Scopus, Springer, Science direct, MDPI e IEEE un total de 106, Segundo, se procedieron a eliminar 13 artículos por duplicados; y por último se filtró excluyendo de acuerdo a los criterios de aceptación y exclusión con un total de 56 artículos, quedando finalmente 37 investigaciones para responder a las cinco preguntas de investigación. Se muestra el proceso de selección de acuerdo a la metodología PRISMA en la **Figura 1.**

Figura 1. Procesos para la selección de artículos científicos.

Las preguntas de investigación presentadas en la revisión de la literatura ayudarán a recuperar característica importante de los artículos que aportan información crucial para el descubrimiento de procesos candidatos a robotizar:

- R1: ¿Que enfoques o técnicas existen para poder registrar acciones de usuario en la minería de procesos?
- R2: ¿Que enfoques, técnicas y algoritmos existen para poder descubrir un proceso?
- R3: ¿Enfoques para identificar un proceso sea robotizable?
- R4: ¿Qué enfoques de clasificación existe para registrar acciones de usuario en la minería de procesos?

Resultados y discusión

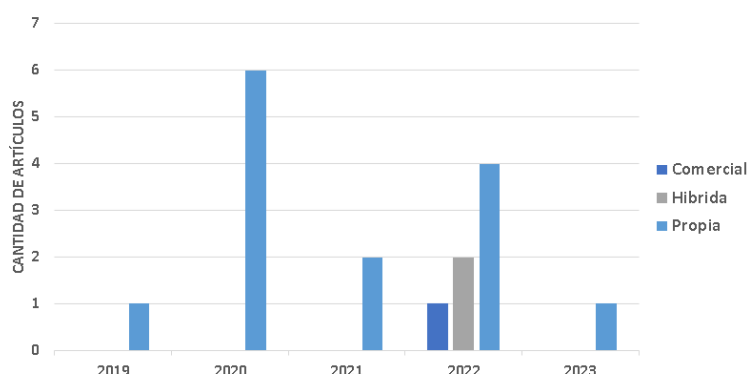
La revisión sistemática incluye 37 artículos sobre estrategias, técnicas y métodos eficientes para el descubrimiento de procesos candidatos a robotizar. Los artículos son de las cinco siguientes fuentes indexadas: 3 de IEEE (8 %), 3 de MDPI (8 %), 6 de Science Direct (16 %), 21 de Scopus (57 %) y 4 de Springer (11 %).

Respuesta a R1

Para poder responder en la primera pregunta hemos identificado 3 tipos de soluciones para poder realizar los registros de usuarios los cuales son herramientas comerciales, sistemas específicos e híbridos. Definimos herramienta comercial cuando la investigación solo usa herramientas de distribución comercial para la extracción de acciones de usuario como Leno et al., (2022) usando Action logger en su investigación, también hay herramientas híbridas cuando el investigador ha usado herramientas comerciales y herramientas propias desarrolladas para uso específico de su investigación (Agostinelli, Marrella, et al., 2022; Chen et al., 2022) y las herramientas propias o creadas a medida para la investigación (Choi et al., 2022; Jimenez-Ramirez et al., 2019; Leno et al., 2021; Šperka & Halaška, 2023; Agostinelli, Lupia, et al., 2022; Choi et al., 2021; Abb & Rehse, 2022; López-Carnicer et al., 2020; Effendi & Sarno, 2020a; Tang et al., 2022; Kim et al., 2021; Egger, 2020; Augusto et al., 2022; Effendi & Sarno, 2020b).

En la **Figura 2** vemos la tendencia para el registro de acciones de usuario mediante tipos de herramientas (comerciales, propias e híbridas) para la investigación. Notamos que la mayoría de herramientas que se usan son construidas específicamente.

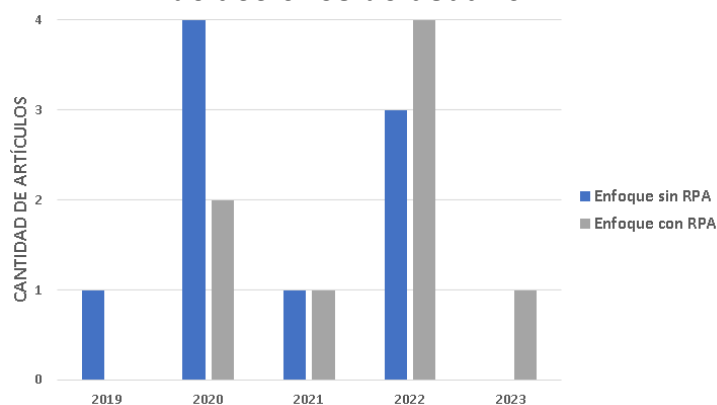
Figura 2. Tendencia de los tipos de herramientas en la literatura para registro de eventos de usuarios.



El enfoque para el registro de acciones de usuario usando RPA ha crecido en estos últimos 4 años teniendo 8 artículos con esta técnica. Los enfoques usados son de Agostinelli, Lupia, et al., (2022) los cuales construyen un “RPA

desatendido y registros basados en rutinas”, Choi et al., (2021) crean “Herramientas RPM (Robotic Process Mining Tools) para el registro de eventos Choi et al., (2022) un año después construyen un “User interface Interactions Recorder (Herramienta basada en RPA) como grabador de acciones en interfaz de usuario”, Abb & Rehse, (2022) et al. plantean un enfoque de “Marco unificado y estándar para identificar tipo de acción y objeto destino sobre el que se ejecuta la acción aplicable para registro de eventos.”, Agostinelli, Marrella, et al., (2022) generan un “Marco para registro de eventos en formato genérico y uso de RPA”, Leno et al., (2021) plantean una “Herramienta de registro de eventos por medio de técnicas de RPA”, Egger, (2020) plantea un enfoque que “Registra acciones de usuario con el Bot y los compara con el proceso comercial” y Šperka & Halaška, (2023) plantean un enfoque de “Marco de evaluación del desempeño de la implementación de RPA (PPAFR) basado en minería de procesos”. En la **Figura 3** vemos la tendencia que va en incremento del total de artículos encontrados para responder esta pregunta.

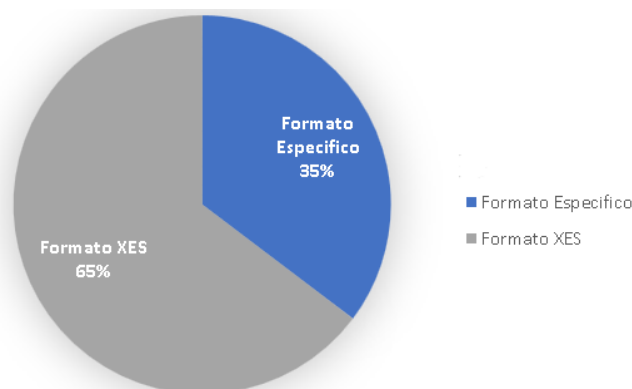
Figura 3. Tendencia de uso de enfoque orientado a RPA para registro de acciones de usuario.



Para realizar la definición de extracción de datos de acciones de usuario se identifica que el estándar más reconocido y usado por los autores es el formato XES (eXtensible Event Stream) (Choi et al., 2022; Jimenez-Ramirez et al., 2019; Šperka & Halaška, 2023; Agostinelli, Lupia, et al., 2022; Abb & Rehse, 2022; López-Carnicer et al., 2020; Tang et al., 2022; Kim et al., 2021; Egger, 2020; Augusto et al., 2022; Chen et al., 2022) el cual permite capturar acciones como texto, clics de mouse e incluso imágenes debido a su flexibilidad propiamente de su definición del estándar, los demás no usan un estándar sino una definición propia y específica para su investigación.

En la **Figura 4** vemos la cobertura de uso de acuerdo a los artículos que han servido para responder la pregunta de investigación.

Figura 4. Estadística del uso de XES de acuerdo a los documentos de la investigación.



El resumen de los 17 artículos para responder esta pregunta se muestra en la **Tabla 3.**

Tabla 3. Relación de enfoques, herramientas y técnicas.

Enfoques, herramientas o Técnicas	Artículos
Enfoque con RPA desatendido y registros basados en rutinas.	(Agostinelli, Lupia, et al., 2022)
User interface Interactions Recorder (UIIR), herramienta basada en RPA que funciona como grabador de acciones en interfaz de usuario.	(Choi et al., 2022)
Técnica usando RPM (Robotic Process Mining Tools) para el registro de eventos automatizado.	(Choi et al., 2021)
Action logger como herramienta comercial y genérica para grabar eventos.	(Leno et al., 2022)
Enfoque unificado y estándar para identificar el tipo de acción y objeto destino sobre el que se ejecuta la acción usada para el registro de eventos.	(Abb & Rehse, 2022)
Enfoque para el registro de eventos en formato genérico y uso de RPA	(Agostinelli, Marrella, et al., 2022)
Herramienta de registro de eventos basada en extracción de imágenes y actividades mediante KeyLogger	(Martínez-Rojas et al., 2022)
Herramienta de registro de eventos por medio de RPA	(Leno et al., 2021)
Enfoque de evaluación del desempeño de la implementación de RPA (PPAFR) basado en minería de procesos.	(Šperka & Halaška, 2023)
Enfoque para registrar acciones de usuario con RPA y comparar con el proceso comercial en un formato estándar	(Egger, 2020)
Herramienta registradora de acciones capaz de recopilar información de diferentes eventos y enviarla a un servidor	(López-Carnicer et al., 2020)
Herramienta Grabador específico de registro de eventos en formato propio del autor.	(Effendi & Sarno, 2020b)

Grabador específico de registro de eventos del autor en formato XES.	(Jimenez-Ramirez et al., 2019); (Effendi & Sarno, 2020a); (Tang et al., 2022); (Kim et al., 2021); (Augusto et al., 2022)
--	---

Respuesta a R2

Para esta segunda pregunta hemos podido recopilar 26 artículos que nos ayudan a responder estas preguntas. Encontramos primero un enfoque de solución es poder mejorar algoritmos que ya han sido desarrollado años atrás pero han aumentado su alcance, tenemos a Sun et al., (2021) que han mejorado el Algoritmo $\alpha+$ y minero heurístico, Alizadeh & Norani, (2018) aumentaron el alcance del algoritmo Competitivo Imperialista (ICA) , Kim et al., (2021) solucionó casuísticas de ruido y bucles del algoritmo σ , Effendi & Sarno, (2020a) han mejorado el bucle corto en el algoritmo Time-Based $\alpha++$ Miner , Effendi & Sarno, (2020b) mejoraron el alcance del algoritmo $\alpha+$ miner y nuevamente Effendi et al., (2021) añadió más alcance en los casos para bucles en el algoritmo fuzzy miner , a continuación en la **Tabla 4** se muestra el resumen de los 7 artículos que corresponde al 30% de artículos que responden esta pregunta.

Tabla 4. Artículos con algoritmos mejorados.

Artículo	Autor	Algoritmo mejorado
A process mining algorithm to mixed multiple-concurrency short-loop structures	(Sun et al., 2021)	Algoritmo $\alpha+$ y minero heurístico
ICMA: a new efficient algorithm for process model discovery	(Alizadeh & Norani, 2018)	Algoritmo Competitivo Imperialista (ICA)
ρ -Algorithm: A SICN- Oriented Process Mining Framework	(Kim et al., 2021)	Algoritmo σ
Parallel process discovery using a new Time- Based $\alpha++$ Miner	(Effendi & Sarno, 2020a)	Algoritmo Time-Based $\alpha++$ Miner
Time-based $\alpha+$ miner for modelling business processes using temporal pattern	(Effendi & Sarno, 2020b)	Algoritmo $\alpha+$ miner
Improved fuzzy miner algorithm for business process discovery	(Effendi et al., 2021)	algoritmo fuzzy miner

Con respecto a las soluciones planteadas y encontradas para poder descubrir procesos tenemos 8 enfoques los cuales son:

- Basado en Reglas: Planteado por un conjunto de reglas lógicas como el Minero inductivo de (Tax et al., 2019).
- Algorítmico general: Se han encontrado 8 algoritmos orientas a lógica propia generada por el autor para eliminar casos atípicos del proceso y depuración manual (Jimenez-Ramirez et al., 2019), algoritmos de detección de rutinas específicas de procesos (Bosco et al., 2019),

enfoque para descubrir patrones de comportamiento contextuales (Acheli et al., 2022), marco de minería de procesos algorítmicos caracterizado por el modelo de procesos matemáticos de redes de control de información estructurada (SICN) y el concepto de función impulsada por masas como criterio de toma de decisiones de los patrones de procesos estructurales (Kim et al., 2021), Effendi & Sarno en el 2020 (Effendi & Sarno, 2020a ;Effendi & Sarno, 2020b) plantea algoritmos “new Time-Based $\alpha++$ Miner” y “Time-based $\alpha+$ Miner (TBA+M)” orientado a mejorar el “ $\alpha++$ Miner” y “ $\alpha+$ Miner” respectivamente, marco algorítmico de descubrimiento de datos correlacionados desde dos perspectivas diferentes, desde el flujo de control (rutas del proceso) y perspectiva de atributos de calidad (atributos de la actividad) (Choueiri & Portela Santos, 2021), marco de evaluación de flujos complejos de procesos (Al-Absi & R'bigui, 2023).

- c. Lógica Difusa: Tenemos 3 enfoques orientado a identificar las tareas que son candidatas para la automatización con la automatización robótica de procesos(Choi et al., 2021), nuevo algoritmo minero difuso mejorado para representar la correlación binaria entre actividades (Effendi et al., 2021) y marco de evaluación de impacto de RPA con apromore (Šperka & Halaška, 2023).
- d. Secuencia: Se tiene un enfoque de minería de secuencias CloFast (Leno et al., 2022).
- e. Evolutivo: Encontramos 3 enfoques como modelo ICMA para descubrir modelos de procesos (Alizadeh & Norani, 2018), método de minería de servicios genéticos híbridos basado en la población de agrupamiento de trazas (HGSM) (Tang et al., 2022) y enfoque de tiempo para encontrar escenarios de procesos (Zhang et al., 2022).
- f. Genético: Se ha encontrado un enfoque basado en islas llamado X-Processes basado en algoritmos genéticos(Fantinato et al., 2023).
- g. Heurístico: Se encontró una solución AlphaMining para descubrir estructuras mixtas de ciclo corto de concurrencia múltiple(Sun et al., 2021).
- h. Otros enfoques: En esta última clasificación se registraron soluciones nuevas que no presentaban en un enfoque específico como uso de Síntesis reactiva para generar bots con Script RPA (Agostinelli, Lupia, et al., 2022), marco para analizar la variabilidad en las acciones humanas de forma automática. El marco aprovecha la información en la pantalla para detectar factores que influyen en las decisiones humanas con técnicas de ML (Martínez-Rojas et al., 2022), método de autómatas de estado finito acíclico determinista (DAFSA) (Průcha & Skrbek, 2022), generador de script de RPA para creación de nuevos procesos usando BPMN (Leno et al., 2021), modelo para predecir la siguiente actividad dentro de un flujo (Obodoekwe et al., 2022), método PROMISE+ que se basa en una abstracción que implica la minería de procesos predictivos para generar un resumen de registro de eventos (Pasquadibisceglie et al., 2022), marco de detección de múltiples vistas (Chen et al., 2022) y COLOSI, marco para respaldar la agrupación de registros de configuración (Ramos-Gutiérrez et al., 2021).

El resumen general de los 26 artículos se muestra en la **Tabla 5**.

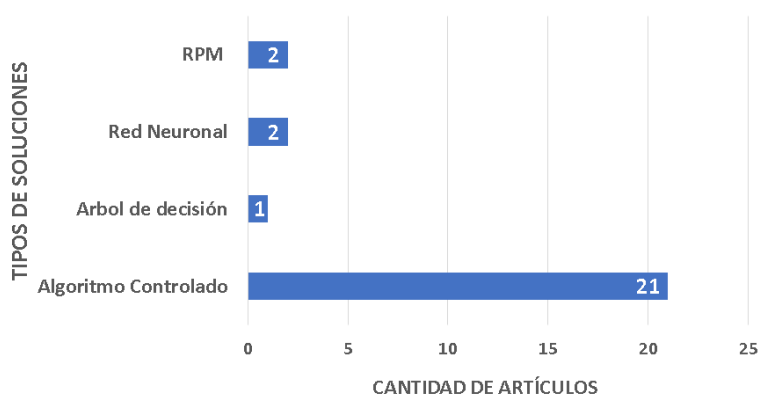
Tabla 5. Enfoques y algoritmos para el descubrimiento de procesos.

Enfoques	Algoritmo	Artículos
Basado en Reglas	Algoritmo Minero inductivo	(Tax et al., 2019)
Algorítmico General	Algoritmo de descubrimiento propio del autor y elimina casos atípicos del proceso.	(Jimenez-Ramirez et al., 2019)
	Algoritmo de detección de especificación de rutina	(Bosco et al., 2019)
	Algoritmo COBPAM y CCOBPAM	(Acheli et al., 2022)
	ρ Algoritmo	(Kim et al., 2021)
	Algoritmo new Time- Based $\alpha++$ Miner	(Effendi & Sarno, 2020a)
	Algoritmo Time-based $\alpha+$ Miner (TBA+M)	(Effendi & Sarno, 2020b)
	Algoritmo Apriori para encontrar relaciones de co-ocurrencia después de las interrelaciones a nivel de atributos de las actividades.	(Choueiri & Portela Santos, 2021)
	Algoritmo para la evaluación de construcciones complejas.	(Al-Absi & R'bigui, 2023)
Lógica Difusa	Algoritmo fuzzy miner	(Choi et al., 2021)
	Algoritmo Improved fuzzy miner	(Effendi et al., 2021)
	Marco de Minería difusa y Apromore	(Šperka & Halaška, 2023)
Secuencia	Algoritmo de minería de secuencias CloFast	(Leno et al., 2022)
Evolutivo	Algoritmo Minero Competitivo Imperialista (ICMA).	(Alizadeh & Norani, 2018)
	Algoritmo de optimización e integración del modelo de minería de servicios genéticos híbridos (HGSM)	(Tang et al., 2022)
	Algoritmo de extracción de conjunto de una transición (Propio) y Minero heurístico	(Zhang et al., 2022)
Genético	X-Processes basado en algoritmos genéticos (modelo basado en islas)	(Fantinato et al., 2023)
Heurístico	Algoritmo Alpha Mining	(Sun et al., 2021)
Otros enfoques	Algoritmo de Síntesis reactiva de robots SW.	(Agostinelli, Lupia, et al., 2022)
	Descubrimiento de modelo mediante árboles de decisión: CART, ID3, C4.5 y CHAID.	(Martínez-Rojas et al., 2022)
	Algoritmo FURIA y el método de autómata de estado finito acíclico determinista (DAFSA).	(Průcha & Skrbek, 2022)

Generador de script de RPA para creación de nuevos procesos	(Leno et al., 2021)
CNN - Based next activity prediction.	(Obodoekwe et al., 2022)
Red neuronal profunda con capas de memoria a corto plazo (LSTM) y algoritmo Backpropagation.	(Pasquadibisceglie et al., 2022)
Marco estadístico EMD con un modelo de PNL en similitud semántica	(Chen et al., 2022)
Marco COLOSSI para descubrimiento de registros de configuración.	(Ramos-Gutiérrez et al., 2021)

También de los artículos encontrados para responder el artículo hemos determinado 4 tipos de soluciones generales los cuales RPM o Robotic Process Mining tenemos 2 artículos orientados a robotizaciones, con redes neuronales hemos encontrado 2 soluciones y de árbol de decisión solo 1 enfoque, los datos propuestos en la tabla 6 muestran una tendencia de algoritmos controlados con 21 artículos que están orientado a mejorar algoritmos ya existentes o nuevos algoritmos con soluciones lógicas específicas, es importante indicar que la tendencia de las técnicas de RPM y redes neuronales son las que están empezando a aumentar en los últimos años. El resumen de la tendencia numérica se muestra en la **Figura 5**.

Figura 5. Tipos de solución para descubrir un proceso



El detalle de los aportes de cada artículo en el tipo de solución se muestra en la **Tabla 6** para una mayor especificación.

Tabla 6. Resumen de tipo de soluciones

Tipos	Artículos
RPM	(Leno et al., 2021);(Agostinelli, Lupia, et al., 2022)
Red Neuronal	(Pasquadibisceglie et al., 2022);(Obodoekwe et al., 2022)
Árbol de decisión	(Martínez-Rojas et al., 2022)
Algoritmo controlado	(Jimenez-Ramirez et al., 2019);(Šperka & Halaška, 2023);(Alizadeh & Norani, 2018);(Choi et al., 2021);(Effendi & Sarno, 2020a);(Tang et al., 2022);(Kim et al., 2021);(Leno et al., 2022);(Effendi et al., 2021);(Effendi & Sarno,

2020b); (Fantinato et al., 2023);(Bosco et al., 2019);(Acheli et al., 2022);(Al-Absi & R'bigui, 2023);(Chen et al., 2022);(Sun et al., 2021);(Tax et al., 2019);(Choueiri & Portela Santos, 2021);(Zhang et al., 2022);(Průcha & Skrbek, 2022);(Ramos-Gutiérrez et al., 2021)

El análisis del tipo salida de la solución para el descubrimiento de proceso se ha realizado en la **Figura 6**, en donde se ve una notable tendencia por el uso de red de Petri por encima de los árboles de decisión, BPMN, red de control, red de control de información y autómatas de estado finito.

Figura 6. Tipos de salidas para el descubrimiento de un proceso.



El detalle de artículos en los tipos de salidas de los algoritmos para descubrir procesos se muestra en la **Tabla 7**.

Tabla 7. Resumen de tipo de salidas

Tipos	Artículos
Red de Petri	(Jimenez-Ramirez et al., 2019);(Pasquadibisceglie et al., 2022);(Agostinelli, Lupia, et al., 2022);(Alizadeh & Norani, 2018);(Choi et al., 2021);(Effendi & Sarno, 2020a);(Leno et al., 2022);(Effendi et al., 2021);(Effendi & Sarno, 2020b);(Fantinato et al., 2023); (Bosco et al., 2019);(Al-Absi & R'bigui, 2023);(Sun et al., 2021);(Tax et al., 2019);(Choueiri & Portela Santos, 2021);(Zhang et al., 2022);(Obodoekwe et al., 2022);(Ramos-Gutiérrez et al., 2021)
Árbol de decisión	(Martínez-Rojas et al., 2022); (Tang et al., 2022), (Acheli et al., 2022)
BPMN	(Leno et al., 2021) ; (Šperka & Halaška, 2023)
Red de control	(Chen et al., 2022)
Red de control de información	(Kim et al., 2021)
Autómata de estado finito	(Průcha & Skrbek, 2022)

Respuesta a R3.

Para poder reconocer si un proceso se puede robotizar se han encontrado 8 artículos, los cuales se orientan en soluciones deterministas como el enfoque determinista para detectar actividades caóticas en registros de eventos y filtrarlas de esos registros de eventos realidad (Tax et al., 2019), enfoque para

descubrir rutinas automatizables a partir de registros de interacciones de usuarios con sistemas de TI y sintetizar especificaciones ejecutables para tales rutinas (Leno et al., 2022) y un método para descubrir rutinas totalmente deterministas (Bosco et al., 2019), procesos robotizables de Inteligencia Híbrida usando Machine Learning con enfoque determinista (Van der Aalst, 2021) y enfoque determinista FURIA (Průcha & Skrbek, 2022); Otro tipo de solución es el uso de herramientas de RPA como el uso de síntesis reactiva para generar bots con Script RPA (Agostinelli, Lupia, et al., 2022), selección de tareas candidatas para RPA basadas en frecuencia, periodicidad y duración (Choi et al., 2021) y scripts para rutinas que pueden ser automatizadas (Leno et al., 2021).

El resumen de soluciones encontradas en la literatura se detalla en la **Tabla 8** a continuación.

Tabla 8. Enfoques y algoritmos para reconocer que un proceso sea robotizable.

Enfoques o algoritmos	Artículos	Enfoque
Algoritmo de síntesis reactiva para crear procesos robotizables	(Agostinelli, Lupia, et al., 2022)	RPA
Algoritmo de selección de tareas candidatas para RPA basadas en frecuencia, periodicidad y duración.	(Choi et al., 2021)	
Generador de scripts para rutinas automatizables	(Leno et al., 2021)	
Enfoque determinista mediante validación de acciones	(Bosco et al., 2019)	
Enfoque determinista para eliminación de actividades poco frecuentes y ruido	(Tax et al., 2019)	Determinista
Enfoque determinista con algoritmo Routine automatability assessment	(Leno et al., 2022)	
Enfoque determinista FURIA	(Průcha & Skrbek, 2022)	
Enfoque determinista enfocado en actividades sin incertidumbre	(Van der Aalst, 2021)	

Respuesta a R4

En esta pregunta se han encontrado 19 soluciones que responden la pregunta de investigación los cuales se clasifican en:

- El Clustering se centra en la capacidad de agrupar o segmentar los datos para facilitar el análisis posterior, como la identificación de patrones secuenciales en términos de frecuencia, longitud y puntuación de cobertura.
- La mejora de la calidad de los datos se enfoca en agregar reglas para filtrar datos irrelevantes que no contribuyen al proceso.
- Enfoque Mixto en donde se usa Clustering y mejora de la calidad en donde se usan técnicas de ambos tipos de clasificación.

La clasificación de estas 19 investigaciones se muestra en la **Tabla 9**.

Tabla 9. Enfoques y soluciones para clasificar los registros de eventos en la minería de procesos.

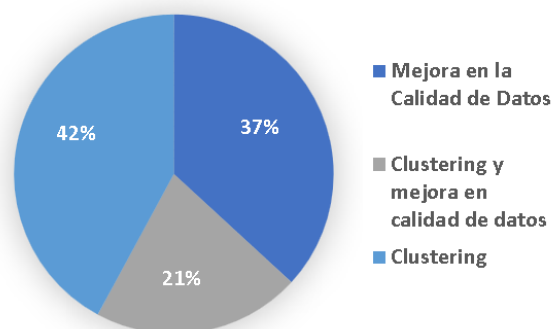
Enfoques	Solución	Artículos
----------	----------	-----------

Mejora en la Calidad de Datos	Algoritmo de filtrado de acciones redundantes (17 reglas)	(Choi et al., 2022)
	Reglas de transformación de Datos y algoritmo de filtrado de acciones irrelevantes	(Choi et al., 2021)
	Algoritmo con enfoque de filtrado de actividad indirectas y actividades basado en la entropía con suavizado de Laplace	(Tax et al., 2019)
	Algoritmo de similitud de imágenes para descartar acciones irrelevantes con atributos de configuración.	(Jimenez-Ramirez et al., 2019)
	Procesos de filtrado y conversión a un registro enriquecido de datos	(López-Carnicer et al., 2020)
	Algoritmo de detección de polígono plano, algoritmo de Detección de acciones y automatización mediante scripts RPA.	(Bosco et al., 2019)
	Algoritmo de probabilidad para eliminar la concurrencia y los bucles automáticos.	(Pishgar et al., 2022)
Clustering	Proceso de estructura de los patrones secuenciales como frecuencia, longitud, cobertura y puntuación de cohesión de patrón	(Leno et al., 2022)
	Algoritmo de clasificación de eventos con técnicas de Re- conocimiento Óptico de Caracteres (OCR) para entregar un modelo de decisión y generar extractores de ocurrencias con técnicas de procesamiento de imágenes (algoritmo de Canny)	(Martínez-Rojas et al., 2022)
	Método de segmentación mediante Carpetenter, RERII, COBBLER y Pattern-Fusion	(Ai et al., 2018)
	Algoritmo de optimización de ballenas Caóticas (CWOA) con la técnica de Clasificación Neuro- Fuzzy (NFC) para la clasificación de datos.	(Al Duhayyim et al., 2022)
	Enfoque de segmentación mediante Minero heurístico para identificar patrones con data grande y Algoritmo 1CT para añadir relaciones de interrelaciones de interdependencia y atributos para cada actividad.	(Choueiri & Portela Santos, 2021)
	Algoritmo de clúster de Markov (MCL)	(Tang et al., 2022)
	Algoritmo basado en densidad para segmentación.	(Zhang et al., 2022)

	Algoritmos de agrupación con codicioso, retrospectivo y jerárquicos	(Ramos-Gutierrez et al., 2021)
Clustering y Mejora en la Calidad de Datos	Algoritmo de filtrado de ruido y agrupación de eventos similares.	(Agostinelli, Lupia, et al., 2022); (Leno et al., 2021); (Van der Aalst, 2021)
	Técnica de identificación de patrones frecuentes en interfaz de usuario con alineación de seguimiento en Process Mining para agrupar las acciones del usuario.	(Agostinelli et al., 2021)

En la **Tabla 9** vemos enfoques de clasificación muy diferenciados entre clusterización, mejora en la calidad de datos e híbridos; el enfoque de clusterización es el que están siendo más usados con un 42% junto con la mejora de calidad en segundo puesto con 37% como se ve en la figura 7, dentro de este grupo tenemos un enfoque de limpieza de datos orientado a filtrado por imágenes (Jimenez-Ramirez et al., 2019) y Martínez-Rojas et al., (2022) en donde puede detectar acciones irrelevantes usando arboles de decisión ganando una mejor precisión. Es importante tener en cuenta que el filtrado de imágenes da una perspectiva adicional de mejora de calidad de datos porque lo valida mediante el reconocimiento y si es acorde a la acción realizada con el mouse o click generando mayor certeza en el flujo como lo indican los autores.

Figura 7. Tipos de clasificación para registro de eventos.



El análisis de las fuentes de datos usadas para esta revisión encontramos en la **Tabla 10** la distribución de los artículos sobre el tema en estudio teniendo más del 40% en el año 2022 y del cuartil 1.

Tabla 10. Porcentaje de artículos de acuerdo a Cuartiles y años de publicación.

Año	Q1	Q2	Q3	Total	Porcentaje
2018	0	1	1	2	5%
2019	0	1	2	3	8%
2020	2	1	4	7	19%
2021	4	2	1	7	19%
2022	8	5	2	15	41%
2023	1	2	0	3	8%
Total	15	12	10	37	100%

Conclusiones

Se han revisado diferentes investigaciones para poder obtener el objetivo de la investigación para encontrar enfoques y técnicas para el descubrimiento de procesos candidatos a robotizar entre los años 2018 y 2023 de los cuartiles 1, 2 y 3 para una mejor calidad de investigación, encontrando en las respuestas de las preguntas de la investigación soluciones basadas en la literatura que nos ayudan a encontrar el propósito de la revisión.

Como parte del flujo de descubrimiento del proceso a robotizar tenemos la primera pregunta de la investigación en donde se ve una tendencia en el uso de específicas para la extracción de la investigación, es decir los investigadores han creado su propia forma de extracción acorde a su necesidad, teniendo como tendencia el uso de herramientas de RPA para poder extraer los registros y datos específicos, pero usando en su mayoría un estándar de XES. Con estas tres tendencias podemos ver cómo se plantea la solución de extracción de registro de usuarios en la literatura en estos últimos años.

Luego de poder tener la data validamos como parte del descubrimiento de proceso que se debe tener como tendencia en la pregunta cuatro en donde la clusterización es la técnica que está siendo más usada en la literatura teniendo un enfoque de clasificación o segmentación en la mayoría de las investigaciones. El segundo enfoque es la técnica es mejorar la calidad de datos orientada a filtrar datos irrelevantes dentro de los registros de usuarios extraídos.

El siguiente paso del descubrimiento del proceso es la ejecución de la técnica del descubrimiento de procesos que respondemos en la pregunta dos en donde encontramos que hay un conjunto de investigadores en donde se enfocan en mejorar o cubrir nuevos escenarios de algoritmos existentes que son alrededor del 30% de soluciones encontradas, además que el uso de los algoritmos generales orientados a lógica propia del autor para poder solucionar un problema de manera controlada y acotada a escenarios particulares. La tendencia del diagrama de salida de las soluciones es la red de Petri por su estándar y facilidad de interpretación.

Como último paso para poder determinar si el proceso descubierto es robotizable el cual responde con la tercera pregunta se encuentra el enfoque determinista con reglas de frecuencia y duración usando algoritmos o redes neuronales para el aprendizaje, cubren más del 60% de investigaciones encontradas.

De todo lo anterior encontrado y descrito se resalta que se debe centrar en soluciones integrales que tengan todos los pasos del descubrimiento de procesos (proceso de extracción, procesos de filtrado de datos, proceso de descubrimiento y proceso de descubrimiento de RPA), se ve un aumento en el uso de técnicas de RPA para poder descubrir procesos robotizables ya sea por la flexibilidad, más fácil construcción, más información y más amigables con el usuario.

Referencias

- Abb, L., & Rehse, J.-R. (2022). A Reference Data Model for Process-Related User Interaction Logs. *Lecture Notes in Computer Science*, 13420. https://doi.org/10.1007/978-3-031-16103-2_7
- Agostinelli, S., Leotta, F., & Marrella, A. (2021). Interactive Segmentation of User Interface Logs. *Lecture Notes in Computer Science*, 13121, 65–80. https://doi.org/10.1007/978-3-030-91431-8_5
- Agostinelli, S., Lupia, M., Marrella, A., & Mecella, M. (2022). Reactive synthesis of software robots in RPA from user interface logs. *Computers in Industry*, 142. <https://doi.org/10.1016/j.compind.2022.103721>
- Agostinelli, S., Marrella, A., Abb, L., & Rehse, J. R. (2022). Mastering Robotic Process Automation with Process Mining. *Lecture Notes in Computer Science*, 13420, 47–53. https://doi.org/10.1007/978-3-031-16103-2_6
- Acheli, M., Grigori, D., & Weidlich, M. (2022). Discovering and Analyzing Contextual Behavioral Patterns From Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 34(12), 5708–5721. <https://doi.org/10.1109/TKDE.2021.3077653>
- Ai, D., Pan, H., Li, X., Gao, Y., & He, D. (2018). Association rule mining algorithms on high-dimensional datasets. *Artificial Life and Robotics*, 23(3), 420–427. <https://doi.org/10.1007/s10015-018-0437-y>
- Al Duhayyim, M., Marzouk, R., Al-Wesabi, F. N., Alrajhi, M., Hamza, M. A., & Zamani, A. S. (2022). An improved evolutionary algorithm for data mining and knowledge discovery. *Computers, Materials and Continua*, 71(1), 1233–1247. <https://doi.org/10.32604/cmc.2022.021652>
- Al-Absi, M. A., & R'bigui, H. (2023). Process Discovery Techniques Recommendation Framework. *Electronics (Switzerland)*, 12(14). <https://doi.org/10.3390/electronics12143108>
- Alizadeh, S., & Norani, A. (2018). ICMA: a new efficient algorithm for process model discovery. *Applied Intelligence*, 48(11), 4497–4514. <https://doi.org/10.1007/s10489-018-1213-3>
- Augusto, A., Mendling, J., Vidgof, M., & Wurm, B. (2022). The connection between process complexity of event sequences and models discovered by process mining. *Information Sciences*, 598, 196–215. <https://doi.org/10.1016/j.ins.2022.03.072>
- Bosco, A., Augusto, A., Dumas, M., Rosa, M. La, & Fortino, G. (2019). Discovering Automatable Routines From User Interaction Logs. *Lecture Notes in Business Information Processing*, 360, 144–162. https://doi.org/10.1007/978-3-030-26643-1_9
- Effendi, Y. A., & Sarno, R. (2020a). Parallel process discovery using a new Time-Based Alpha++ Miner. *IJUM Engineering Journal*, 21(1), 126–141. <https://doi.org/10.31436/ijumej.v21i1.1173>
- Effendi, Y. A., & Sarno, R. (2020b). Time-based α + miner for modelling business processes using temporal pattern. *Telkomnika*, 18(1), 114–123. <https://doi.org/10.12928/TELKOMNIKA.v18i1.12733>
- Effendi, Y. A., Sarno, R., & Marsha, D. V. (2021). Improved fuzzy miner algorithm for business process discovery. *Telkomnika*, 19(6), 1830–1839. <https://doi.org/10.12928/TELKOMNIKA.v19i6.19015>
- Egger, A. (2020). Bot Log Mining: Using Logs from Robotic Process Automation for Process Mining. *Lecture Notes in Computer Science*, 12400, 51–61. https://doi.org/10.1007/978-3-030-62522-1_4

- Fantinato, M., Peres, S. M., & Reijers, H. A. (2023). *X-Processes: Process Model Discovery with the Best Balance Among Fitness, Precision, Simplicity, and Generalization through a Genetic Algorithm*. <https://doi.org/10.1016/j.is.2023.102247>
- Chen, Q., Lu, Y., Tam, C. S., & Poon, S. K. (2022). A Multi-View Framework to Detect Redundant Activity Labels for More Representative Event Logs in Process Mining. *Future Internet*, 14(6). <https://doi.org/10.3390/fi14060181>
- Choi, D., R'Bigui, H., & Cho, C. (2021). Candidate digital tasks selection methodology for automation with robotic process automation. *Sustainability (Switzerland)*, 13(16). <https://doi.org/10.3390/su13168980>
- Choi, D., R'Bigui, H., & Cho, C. (2022). Enabling the Gap Between RPA and Process Mining: User Interface Interactions Recorder. *IEEE Access*, 10, 39604–39612. <https://doi.org/10.1109/ACCESS.2022.3165797>
- Choueiri, A. C., & Portela Santos, E. A. (2021). Discovery of path-attribute dependency in manufacturing environments: A process mining approach. *Journal of Manufacturing Systems*, 61, 54–65. <https://doi.org/10.1016/j.jmsy.2021.08.005>
- Jimenez-Ramirez, A., Reijers, H. A., Barba, I., & Del Valle, C. (2019). A Method to Improve the Early Stages of the Robotic Process Automation Lifecycle. *Lecture Notes in Computer Science*, 11483, 446–461. https://doi.org/10.1007/978-3-030-21290-2_28
- Kim, K. S., Pham, D. L., & Kim, K. P. (2021). p-Algorithm: A SICN-Oriented Process Mining Framework. *IEEE Access*, 9, 139852–139875. <https://doi.org/10.1109/ACCESS.2021.3119011>
- Leno, V., Augusto, A., Dumas, M., La Rosa, M., Maggi, F. M., & Polyvyanyy, A. (2022). Discovering data transfer routines from user interaction logs. *Information Systems*, 107. <https://doi.org/10.1016/j.is.2021.101916>
- Leno, V., Polyvyanyy, A., Dumas, M., La Rosa, M., & Maggi, F. M. (2021). Robotic Process Mining: Vision and Challenges. *Business and Information Systems Engineering*, 63(3), 301–314. <https://doi.org/10.1007/s12599-020-00641-4>
- López-Carnicer, J. M., del Valle, C., & Enríquez, J. G. (2020). Towards an OpenSource Logger for the Analysis of RPA Projects. *Lecture Notes in Business Information Processing*, 393, 176–184. https://doi.org/10.1007/978-3-030-58779-6_12
- Martínez-Rojas, A., Jiménez-Ramírez, A., Enríquez, J. G., & Reijers, H. A. (2022). Analyzing Variable Human Actions for Robotic Process Automation. *Lecture Notes in Computer Science*, 13420, 75–90. https://doi.org/10.1007/978-3-031-16103-2_8
- Obodoekwe, E., Fang, X., & Lu, K. (2022). Convolutional Neural Networks in Process Mining and Data Analytics for Prediction Accuracy. *Electronics (Switzerland)*, 11(14). <https://doi.org/10.3390/electronics11142128>
- Pasquadibisceglie, V., Appice, A., Castellano, G., & Van Der Aalst, W. (2022). PROMISE: Coupling Predictive Process Mining to Process Discovery. *Information Sciences*, 606, 250–271. <https://doi.org/10.1016/j.ins.2022.05.052>
- Pishgar, M., Razo, M., & Darabi, H. (2022). Improving Process Discovery Algorithms Using Event Concatenation. *IEEE Access*, 10, 69072–69090. <https://doi.org/10.1109/ACCESS.2022.3185235>
- Průcha, P., & Skrbek, J. (2022). Use of FURIA for Improving Task Mining. *Acta Informatica Pragensia*, 11(2), 241–253. <https://doi.org/10.18267/j.aip.183>

- Ramos-Gutiérrez, B., Jesus Varela-Vaca, Jose, Galindo, María, Gomez-Lopez, T., & Benavides, D. (2021). Discovering configuration workflows from existing logs using process mining. *Empir Software Eng*, 26, 11. <https://doi.org/10.1007/s10664-020-09911-x>
- Sun, H. W., Liu, W., Qi, L., Du, Y. Y., Ren, X., & Liu, X. Y. (2021). A process mining algorithm to mixed multiple-concurrency short-loop structures. *Information Sciences*, 542, 453–475. <https://doi.org/10.1016/j.ins.2020.07.003>
- Šperka, R., & Halaška, M. (2023). The performance assessment framework (PPAFR) for RPA implementation in a loan application process using process mining. *Information Systems and e-Business Management*, 21(2), 277–321. <https://doi.org/10.1007/s10257-022-00602-2>
- Tang, Y., Li, T., Zhu, R., Liu, C., & Zhang, S. (2022). A Hybrid Genetic Service Mining Method Based on Trace Clustering Population. *IEICE Transactions on Information and Systems*, E105D(8), 1143–1455. <https://doi.org/10.1587/transinf.2021EDP7190>
- Tax, N., Sidorova, N., & van der Aalst, W. M. P. (2019). Discovering more precise process models from event logs by filtering out chaotic activities. *Journal of Intelligent Information Systems*, 52(1), 107–139. <https://doi.org/10.1007/s10844-018-0507-6>
- Van der Aalst, W. M. P. (2021). Hybrid intelligence: to automate or not to automate, that is the question. *International Journal of Information Systems and Project Management*, 9(2), 5–20. <https://doi.org/10.12821/ijispm090201>
- Zhang, Z., Johnson, C., Venkatasubramanian, N., & Ren, S. (2022). Process scenario discovery from event logs based on activity and timing information. *Journal of Systems Architecture*, 125. <https://doi.org/10.1016/j.sysarc.2022.102435>