

Recibido 06 FEB. 2024

ReCIBE, Año 12 No.2, NOV. 2023

Aceptado 12 FEB. 2024

Revisión de literatura de elementos de minería de procesos para la minería de procesos de software

Literature review of process mining elements for software process mining

Silvia Jaqueline Urrea Contreras¹

Brenda L. Flores Rios¹

brenda.flores@uabc.edu.mx

Angélica Astorga Vargas²

Jorge E. Ibarra-Esquer²

Félix F. González-Navarro¹

Iván A. García Pacheco³

¹Instituto de Ingeniería,

²Facultad de Ingeniería. Universidad Autónoma de Baja California, Baja California, México.

³División de Estudios de Posgrado, Universidad Tecnológica de la Mixteca, Oaxaca, México.

RESUMEN

Para la aplicación de la Minería de Procesos, es fundamental la identificación de elementos clave del proceso, tales como eventos, casos, trazas. Estos elementos han sido formalmente definidos a través de notaciones matemáticas que incluyen las relaciones entre ellos y se requiere un conocimiento especializado para su aplicación en proyectos. Por medio de una semi-Revisión Sistemática de Literatura se analizaron las representaciones formales de los elementos y perspectivas de la Minería de Procesos. Como resultado se presenta un marco referencial que incluye definiciones, propiedades y aplicaciones, con el objetivo de brindar a los investigadores, profesionales y practicantes de la Ingeniería de Software conocimiento tópico de Minería de Procesos que sea de utilidad en proyectos de Minería de Procesos de Software.

PALABRAS CLAVE

Minería de Procesos. Ingeniería de software. Semi-Revisión Sistemática de Literatura. Conocimiento tópico.

ABSTRACT

For the application of Process Mining, the identification of key elements of the process, such as events, cases, traces, is essential. These elements have been formally defined through mathematical notations that include the relationships between them, requiring specialized knowledge for their application in projects. Through a semi-Systematic Literature Review, the formal representations of the elements and perspectives of Process Mining were analyzed. As a result, a reference framework is presented that includes definitions, properties, and applications, with the objective of providing researchers, professionals, and practitioners of Software Engineering with topical knowledge of Process Mining that is useful in Software Process Mining projects.

KEYWORDS

Process Mining, Software Engineering, Semi-Systematic Literature Review, Topic Knowledge.

1. INTRODUCCIÓN

La Minería de Procesos establece un puente entre el conocimiento del dominio de las tecnologías de la información y las Ciencias Administrativas, contribuyendo significativamente tanto a la comunidad de Ciencias de la Computación como al Modelado de Procesos de Negocio (Marin-Castro & Tello-Leal, 2021). Por otro lado, la Minería de Procesos es una disciplina que aprovecha el volumen y la disponibilidad de los datos, así como la importancia de extraer conocimiento (Van der Aalst, 2016), por medio de diversos algoritmos (*Alpha*, heurístico, genético, difuso, entre otros) y técnicas (*Handover of work*, *Token replay*, Alineaciones), lo que resulta en mejoras y optimización de los procesos (Urrea-Contreras et al., 2021). Los algoritmos adoptan una formalización matemática en sus entradas (Leemans, van Zelst & Lu, 2023), lo que permite la definición de métricas y medidas de rendimiento para evaluar y comparar los procesos. Las técnicas pueden aplicarse a una variedad de procesos, incluyendo los procesos administrativos, médicos, de logística, productivos o incluso al proceso de desarrollo de software (Keith & Vega, 2016).

Actualmente, en las empresas de desarrollo de software se aplica la Minería de Procesos con el propósito de extraer el conocimiento almacenado en sus sistemas de gestión a partir de los datos generados durante la ejecución de las actividades. A esta aplicación se le denomina Minería de Procesos de Software (Rubin, Lomazova & Van der Aalst, 2014; Liu et al., 2016; Erdem, Demirörs & Rabhi, 2018). En este contexto, se están desarrollando y aplicando diversas técnicas, algoritmos y herramientas, tanto específicas como genéricas, para la extracción de registros de eventos (Dakic, 2020). El analizar los registros de eventos creados a partir de los datos del proceso de desarrollo de software permite a las empresas de desarrollo de software poseer una visión holística del ciclo de vida del proceso y brindarles información relacionada a las áreas de mejora en cada etapa para la optimización del flujo de trabajo completo.

Para la aplicación de la Minería de Procesos, es esencial identificar elementos clave del proceso, tales como eventos, casos, trazas, entre otros. Estos elementos han sido formalmente definidos a través de notaciones matemáticas que incluyen las relaciones entre ellos, como la relación entre eventos, casos específicos y actividades (Van der Aalst, 2016). Las notaciones establecen un lenguaje formal y estructurado, ofreciendo un conocimiento explícito para representar las actividades y sus flujos de trabajo facilitando la interpretación, el análisis y modelado de procesos.

Bajo una perspectiva de conocimiento, la principal aportación de este artículo es implementar la metodología Semi-Revisión Sistemática de Literatura (semi-SLR por sus siglas en inglés) para sintetizar la información y el conocimiento tópicamente existente sobre las definiciones formales de los elementos de la Minería de Procesos: evento, caso, traza, registros de eventos, red Petri y perspectivas. Por medio de esta síntesis, se desea establecer un marco referencial con el conocimiento semántico para su socialización e interiorización entre los roles que participan en la identificación de patrones, desviaciones y oportunidades de mejora, especialmente en los roles como analistas de negocio, responsable de mejora de procesos, gestor de proyectos pero que no cuentan con conocimientos de Ciencia de Datos, Ciencia de Procesos, Minería de Procesos, programación para el pre-procesamiento de datos, entre otro conocimiento especializado (Urrea-Contreras et al., 2022).

El documento se estructura de la siguiente manera: la sección 2 presenta un contexto de la Minería de Procesos de Software; la sección 3 especifica la metodología semi-SLR; en la sección 4 se interpretan y analizan las definiciones formales e importancia de los elementos de Minería de Procesos para contribuir en los procesos de software. Finalmente, la sección 5 muestra las discusiones y la sección 6 las conclusiones y trabajo futuro.

2. MINERÍA DE PROCESOS DE SOFTWARE

La rápida evolución de los sistemas de gestión está permitiendo controlar procesos empresariales, organizacionales, componentes, considerando recursos tecnológicos y humanos. En consecuencia, estos sistemas registran datos, por medio de mensajes, transacciones, registro de actividades, seguimiento de tareas, entre otras propiedades. Estos datos se extraen para construir registros de eventos y realizar tres tipos básicos de minería: el descubrimiento, la conformidad y la mejora de procesos (Dakic, 2020). De esta forma, los registros de eventos en la Minería de Procesos representan los datos de ejecución histórica de las actividades definidas por modelos de procesos almacenados como una tabla (Van der Aalst, 2016; Rivas & Bayona-Oré, 2019). No obstante, uno de los desafíos más significativos reside en la extracción e integración de los registros de eventos procedentes de múltiples fuentes de datos o repositorios.

En el trabajo relacionado se usa el término Minería de Procesos de Software (Rubin, Lomazova, & Van der Aalst, 2014; Liu et al., 2016; Erdem, Demirörs, & Rabhi, 2018) para referirse a la posibilidad de llevar a cabo diversas acciones. Por ejemplo: 1) Utilizar técnicas de descubrimiento de procesos para analizar el comportamiento real del proceso de desarrollo de software, 2) Aplicar técnicas de verificación de conformidad para detectar desviaciones en los procesos, 3) Identificar discrepancias en los procesos de software, obteniendo datos que pueden ser utilizados para planificar correcciones y mejoras futuras (Keith & Vega, 2016) y 4) Realizar predicciones en los problemas relacionados con el rendimiento o la conformidad a partir de la alineación de los modelos y el comportamiento real (Urrea-Contreras et al., 2022).

Se han documentado diversos ejemplos de la aplicación exitosa de la Minería de Procesos de Software, que proporcionan una visión más profunda de los desafíos y soluciones en este campo (Keith & Vega, 2016; Van der Aalst, 2015; Caldeira & e Abreu, 2016; Dong et al., 2017; Štolfa, 2018; Marques, da Silva, & Ferreira, 2018; Kneuper, 2018). Estos estudios muestran cómo las empresas de software enfrentan problemas concretos, como errores en el modelado de procesos y desviaciones en su aplicación, lo que conlleva costos significativos. Uno de los aspectos críticos resaltados en estos ejemplos es la disponibilidad de datos adecuados para el análisis, que con frecuencia se encuentran no estructurados y requieren un trabajo de pre-procesamiento.

Es relevante destacar que la Minería de Procesos de Software continúa siendo considerada como un campo de investigación en desarrollo (Rubin, Lomazova, & Van der Aalst, 2014; Erdem, Demirörs, & Rabhi, 2016) por lo que es importante seguir realizando investigaciones desde la representación formal, análisis y aplicación de técnicas, algoritmos y herramientas que apoyen a la innovación del descubrimiento de procesos de software, asociadas a las necesidades de conocimiento y la aplicación de estrategias.

3. METODOLOGÍA

Se implementó una Semi-Revisión Sistemática de Literatura (semi-SLR por sus siglas en inglés) propuesta por Snyder (Snyder, 2019) centrada en las notaciones matemáticas utilizadas en la Minería de Procesos. Una semi-SLR es un enfoque de investigación utilizado que permite contar con una visión general del estado actual del trabajo relacionado y examina cómo ha evolucionado el conocimiento dentro de uno varios campos o disciplinas a lo largo del tiempo (Snyder, 2019).

3.1 Planificación de la Revisión

Esta fase tiene como propósito recopilar, analizar, extraer y sintetizar la investigación sobre las notaciones formales y definiciones utilizadas en la Minería de Procesos para su integración en una guía técnica de MoProPEI-MP (Urrea-Contreras et al., 2018) y aplicabilidad en proyectos de Minería de procesos de software.

Pregunta de investigación

La Pregunta de Investigación (PI) se centra en la importancia de la conceptualización (conocimiento tópico) de los elementos identificados por (Van der Aalst, 2016) para evitar ambigüedades y apoyar en su interpretación y comprensión. Se formuló de la siguiente manera:

PI: ¿Cuáles definiciones o representaciones formales existen sobre elementos de la Minería de procesos? donde el concepto de registros de eventos se basa en los siguientes supuestos (Kneuper, 2018):

- 1) Un registro de eventos contiene datos de varios procesos,
- 2) Un proceso (o su ejecución) consta de casos o instancias individuales y
- 3) Un caso consta de diferentes eventos, donde cada evento pertenece exactamente a un caso.

Selección de fuentes y estrategias de búsqueda

Para abordar la PI, se seleccionaron tres fuentes de datos para la búsqueda de la literatura relevante: *ACM Digital Library*, *Scopus* y *Springer*. Estas fuentes se eligieron estratégicamente debido a su amplio alcance en el campo de estudio y su capacidad para proporcionar un acceso completo a una variedad de estudios académicos, revistas y conferencias. La combinación de estas fuentes permitió realizar una búsqueda de la literatura existente relacionada con la investigación.

Cadena de búsqueda

Después de establecer las fuentes, se procedió a identificar el siguiente conjunto de términos clave para la búsqueda: *Process mining*, *Software process mining*, *Process mining representation*, *Process mining elements*, *Process mining formalisms*, *Process mining analysis*, *Process mining perspectives*, *Discovery*, *Conformance*, *Enhancement*. Los términos claves se ajustaron mediante el uso de operadores lógicos para adaptarlos a las particularidades de cada una de las fuentes.

Criterios de inclusión y exclusión

Con el fin de distinguir entre los estudios relacionados con el tema y aquellos que aportan información pertinente, se han establecido criterios de inclusión y exclusión. Los criterios de inclusión se refieren a las características específicas que deben cumplir los estudios para ser considerados relevantes para la investigación. Por otro lado, los criterios de exclusión consisten en una serie de características que permiten descartar aquellos documentos que no contribuyan de manera significativa al objetivo de la investigación.

Los criterios definidos fueron:

Criterios de inclusión: 1) Estudios en idioma inglés y español; 2) Estudios a partir del año 2013; 3) Estudios que contengan información relacionada a las notaciones matemáticas de la Minería de Procesos; 4) Estudios publicados en revistas y memorias de congresos especializados.

Criterios de exclusión: 1) Estudios repetidos (duplicados); 2) Estudios a los que no se tenga acceso; 3) Estudios que no cumplan los criterios de inclusión.

3.2 Realización de Revisión

Se realizó una búsqueda global de todos los artículos que contenían los términos antes mencionados en sus títulos y resumen. La semi-SLR reveló un conjunto de 1104 artículos, distribuidos de la siguiente manera: *ACM Digital Library* (930 artículos), *Scopus* (89 artículos) y *Springer* (85 artículos). Se procedió a la exclusión de aquellos que no estaban directamente relacionados con el objetivo del estudio actual, centrándose específicamente en la identificación de formalismos y el uso de notaciones matemáticas en la Minería de Procesos.

Posteriormente, a los artículos identificados se les aplicó la técnica de *Snowball Sampling*

(Leemans, van Zelst & Lu, 2023; Parker, Scott, & Geddes, 2019), lo que permitió verificar que las representaciones matemáticas que presentaban algunos artículos eran basados en referencias. Esta técnica se aplicó de manera iterativa, para consultar las referencias de los artículos y se eligieron aquellos que fueron pertinentes para este estudio. En total, se obtuvo una muestra de 13 artículos que exploran el tema de notaciones matemáticas en el contexto de la Minería de Procesos.

4. INTERPRETACIÓN DE RESULTADOS

Para conocer el conocimiento semántico del significado y representación de las relaciones entre los distintos elementos de la Minería de Procesos, se compararon sus definiciones y representaciones formales. De esta manera, el enfoque consistió en analizar cómo se conciben y se expresan matemáticamente estas relaciones dentro del ámbito de la Minería de Procesos.

4.1 Universos

Se establecen universos conceptuales con el propósito de proporcionar un contexto para las definiciones formales, permitiendo una comprensión estructurada de los elementos de la Minería de Procesos con el propósito de contextualizar las definiciones formales.

Definición 1 (Universos). Se utilizan las siguientes notaciones para los universos:
 ϵ denota el universo de eventos únicos; es decir, el conjunto de todos los posibles identificadores que se refieren a eventos concretos,
A denota el conjunto de todos los posibles nombres de atributos,
V denota el conjunto de todos los posibles valores de atributos,
C denota el universo de casos.

4.2 Evento

Un proceso describe la ejecución del conjunto de actividades. Cada actividad ejecutada da lugar a uno o más eventos que registran información que describe la instancia del proceso, la actividad ejecutada, el tiempo de ejecución y datos adicionales como los atributos y valores involucrados (Van Eck, 2022).

Tabla 1. Comparativa de definiciones y representaciones formales para evento

Cita	Definición y representación formal
Van Eck, 2022	<i>Un evento $e \in \varepsilon$ es una tupla de n pares nombre-pares de valores. Es decir, $e = ((a_1, v_1), (a_2, v_2), \dots, (a_n, v_n))$ con atributos $a_1, a_2, \dots, a_n \in Attr$, $\forall 1 \leq i < j \leq n : a_i \neq a_j$, y valores $v_1, v_2, \dots, v_n \in Val$.</i>
De Leoni, Van der Aalst, & Dees, 2016	<i>Un evento e es una asignación de valores a características, es decir, $e \in C \rightarrow U$. En el resto $\varepsilon = C \rightarrow U$ es el universo de eventos.</i>
Diamantini, Genga, & Potena, 2016	<i>Un evento σ es una instancia de una especificación de entidad, y se identifica por un identificador único. Denotaremos por \odot el universo de eventos, es decir, el conjunto de todos los posibles identificadores de eventos. Cada evento se describe mediante un conjunto de atributos. Sea A el conjunto de nombres de atributos; para cualquier evento $\sigma \in \odot$ y atributo $\alpha \in A$, $\# \alpha(\sigma)$ es el valor del atributo α para el evento σ. Cada evento se describe al menos con un nombre de atributo de etiquetado, tal que $\# \text{nombre}(\sigma) = (\# \text{recurso}(\sigma), \# \text{actividad}(\sigma)) \in ES$, y por un atributo timestamp que identifica el momento en que se produjo el suceso.</i>
Senderovich, 2017	<i>Sea ε el conjunto de todos los eventos posibles, es decir, identificadores únicos de eventos. Sea ε^* el conjunto de todas las secuencias finitas sobre ε siendo $\epsilon \in \varepsilon^*$ la traza vacía. Los eventos se asocian con atributos, por ejemplo, marcas de tiempo, actividades, ubicaciones y recursos. Denotamos por A_Σ el conjunto de todos los espacios de atributos de eventos: $A_\Sigma = \{A_i \mid i \in I\}$, siendo A_i el espacio de atributos i_{th} y I es el conjunto de índices de atributos de eventos.</i>

4.3 Caso

Los casos, como los eventos, pueden tener atributos y por cada caso se puede registrar una traza de eventos (Marin-Castro & Tello-Leal, 2021). Los casos al ser representaciones de instancias de la actividad están ligadas a poseer atributos y características únicas que pueden pasar por diferentes actividades durante la ejecución (Tabla 2).

Tabla 2. Comparativa de definiciones y representaciones formales para caso

Cita	Definición y representación formal
Senderovich, 2017	<i>El conjunto de casos es $C \subseteq \varepsilon^*$, es decir, un conjunto de secuencias finitas de eventos. Exigimos que cada evento aparezca como máximo una vez en algún caso. Un caso $c = \langle c_1, \dots, c_n \rangle \in C$ es, por tanto, una secuencia finita de sucesos tal que $c_i \in \varepsilon$, $i = 1, \dots, n$, siendo n el número de sucesos por caso.</i>
Hompes et al., 2015	<i>Sea C el universo de casos, es decir, el conjunto de todos los posibles identificadores de casos. $n(c)$ es el valor del atributo $n \in N$ para el caso $c \in C$ ($n(c) = \perp$ si c no tiene ningún atributo llamado n).</i>
Seeliger, Nolle, & Mühlhäuser, 2018	<i>Sea C el conjunto de todos los posibles identificadores de casos. Los casos también pueden tener atributos, así que para cada caso $c \in C$ y un atributo $a \in A$, definimos $\#_a(c)$ como el valor de un atributo a para el caso c.</i>
Jokonowo et al., 2019	<i>Sea C el caso. Un caso son las actividades del ciclo de vida del proceso para cualquier evento $c \in C$.</i>

4.4 Traza

Los eventos de un caso se representan en forma de traza, es decir, una secuencia de únicos eventos (Van der Aalst, 2016). Estas trazas constituyen la entrada tradicional de muchos enfoques de descubrimiento de procesos y otras técnicas de minería de procesos (Van Eck, 2022). Las trazas capturan el seguimiento de las actividades realizadas y las decisiones tomadas en el proceso (Tabla 3).

Tabla 3. Comparativa de definiciones y representaciones formales para traza

Cita	Definición y representación formal
Van Eck, 2022	<i>Una traza $\sigma = \langle e_1, e_2, \dots, e_n \rangle \in \varepsilon^*$ es una secuencia finita de n eventos ordenados en el tiempo que han sido registrados para una instancia de proceso específica. Para todos los $e_i, e_j \in \sigma$ con $1 \leq i < j \leq n$, se cumple que (1) $\#_{time}(e_i) \leq \#_{time}(e_j)$, (2) $\#_{pi}(e_i) = \#_{pi}(e_j)$ y que (3) $e_i \neq e_j$.</i>
De Leoni, Van der Aalst, & Dees, 2016	<i>Una traza $t \in \varepsilon^*$ es una secuencia de eventos. Sea $T \in \varepsilon^*$ el universo de las trazas.</i>
Diamantini, Genga, & Potena, 2016	<i>Una traza l es una secuencia finita de eventos, $l = \langle \sigma_1, \sigma_2, \dots, \sigma_n \rangle$, $\sigma_i \in \Sigma$, $\#_{timestamp}(\sigma_i) < \#_{timestamp}(\sigma_j)$, $i < j$.</i>
Hompes et al., 2015	<i>Una traza es una secuencia finita de eventos $\sigma \in \varepsilon^*$, tal que cada evento aparece sólo una vez, es decir, para $1 \leq i \leq j \leq \sigma$: $\sigma(i) \neq \sigma(j)$.</i>

4.5 Registro de eventos

Un registro de eventos almacena datos sobre la ocurrencia de actividades que fueron registradas por los sistemas de información mientras soportan la ejecución de un proceso. Cada ejecución de una instancia de proceso da lugar a una secuencia de eventos (Van Eck, 2022; De Leoni, Van der Aalst & Dees, 2016). La mayoría de los registros de eventos almacenan información adicional (Van der Aalst, 2016) que dependen de las técnicas de Minería de Procesos y las preguntas en cuestión que se necesiten responder con su análisis, relacionadas con la conformidad y el rendimiento. La Tabla 4 presenta una comparativa de definiciones de este término.

Tabla 4. Comparativa de definiciones y representaciones formales de registro de eventos

Cita	Definición y representación formal
Van Eck, 2022	<i>Un registro de eventos $L = \{\sigma_1, \dots, \sigma_n\} \subset \varepsilon^*$ es un conjunto finito de n trazas. Para todos los $\sigma, \sigma' \in L$ con dos eventos cualesquiera $e \in \sigma, e' \in \sigma'$ se cumple que $\#_{pi}(e) \neq \#_{pi}(e')$.</i>
De Leoni, Van der Aalst, & Dees, 2016	<i>Un registro de eventos L es un conjunto múltiple de trazas, es decir $L \in (T)$.</i>
Diamantini, Genga, & Potena, 2016	<i>Un registro de eventos L es el registro de una colección de trazas. Cada evento en una traza (registro de eventos) es único, es decir, $\forall \sigma_i, \sigma_j \in L, i \neq j \Rightarrow \sigma_i \neq \sigma_j$.</i>
Hompes et al., 2015	<i>Un registro de eventos es un conjunto de casos $L \subseteq C$ tal que cada evento aparece como máximo una vez en todo el registro, es decir, para cualquier $c, c' \in L$ tal que $c \neq c'$: $set(c) \cap set(c') = \emptyset$.</i>

4.6 Red Petri

Los modelos de procesos, como técnicas de descubrimiento de procesos, permiten representar el comportamiento de un proceso o sistema, por medio de varias notaciones de modelado o representaciones gráficas (Van der Aalst, 2016, Van Eck, 2022). De esta forma, los modelos de procesos apoyan la descripción del orden de ejecución de las actividades y las opciones que se pueden realizar, también conocido como flujo de control.

Tabla 5. Comparativa de definiciones y representaciones formales de registro de red Petri

Cita	Definición y representación formal
Li, 2019	<p>Una red Petri es una tupla $N = (P, T, F)$ donde P es un conjunto finito de lugares, T es un conjunto finito de transiciones tal que $P \cap T = \emptyset$, y $F \subseteq (P \times T) \cup (T \times P)$ es un conjunto de arcos dirigidos, llamados relaciones de flujo. Una red Petri marcada es un par (N, M), donde $N = (P, T, F)$ es una red de Petri y $M \in \mathbb{B}(P)$ es un subconjunto sobre P que denota el marcado de la red.</p>
Leemans, 2018	<p>Sea A el alfabeto de actividades y sea τ una etiqueta especial (silenciosa) tal que $\tau \notin A$. Una red Petri es una tupla $PN = (P, T, F, \ell)$, donde:</p> <ul style="list-style-type: none"> - P es un conjunto finito de lugares - T es un conjunto finito de transiciones, tal que $P \cap T = \emptyset$ - F es un subconjunto finito de arcos dirigidos: $F \in \mathcal{B}((P \times T) \cup (T \times P))$ <p>ℓ es una función de etiquetado de transiciones: $\ell: T \rightarrow A \cup \{\tau\}$</p>
Kalenkova et al., 2016	<p>Una red Petri es una tupla $PN = (P, T, F)$ con P el conjunto de lugares, T el conjunto de transiciones, $P \cap T = \emptyset$, y $F \subseteq (P \times T) \cup (T \times P)$ la relación de flujo</p>
Van Zelst, 2019	<p>Sea P un conjunto de lugares y sea T un conjunto de transiciones s.t. $P \cap T = \emptyset$. Sea $F = (P \times T) \cup (T \times P)$ la relación de flujo. Además, dejemos que Σ denote el universo de etiquetas, que $\tau \notin \Sigma$ y que $\lambda: T \rightarrow \Sigma \cup \{\tau\}$ denota la función de etiquetado de las transiciones. Una red de Petri N, es una tupla $N = (P, T, F, \lambda)$.</p>
Syamsiyah, 2020	<p>Una red de Petri es una tupla $\mathcal{N} = (P, T, F, \ell)$ donde:</p> <ul style="list-style-type: none"> - P es un conjunto de lugares - T es un conjunto de transiciones, $P \cap T = \emptyset$ - $F \subseteq (T \times P) \cup (P \times T)$ es una relación de flujo <p>$\ell: T \rightarrow U_A \cup \{\tau\}$ es una función de etiquetado.</p>

4.7 Aplicabilidad de las definiciones formales

Representación en lenguaje formal: Los formalismos proporcionan un lenguaje formal y una notación matemática estandarizada que los autores muestran para una comunicación clara y precisa entre los expertos en Minería de procesos. Al tener un lenguaje común, los profesionales pueden compartir y discutir sus ideas, hallazgos y resultados de manera más efectiva, lo que facilita la colaboración y el intercambio de conocimientos en el campo. La notación matemática utilizada en los formalismos identificados muestra varias ventajas. En primer lugar, brinda un nivel de dominio y exactitud que no siempre se puede lograr con descripciones en lenguaje natural. La notación matemática permite una representación más rigurosa de los conceptos (por lo regular los artículos seleccionados contemplaban una sección de conceptos preliminares) y relaciones dentro de los procesos, lo que evita ambigüedades en la interpretación de los resultados. Otra ventaja de la notación matemática es su capacidad para simplificar y generalizar conceptos complejos. Al representar los procesos de manera abstracta y simbólica, se pueden identificar patrones y regularidades que pueden pasar desapercibidos en las descripciones verbales o visuales.

Esto permite una comprensión más profunda de los procesos y la capacidad de aplicar resultados y conocimientos a diferentes contextos y dominios.

Algoritmos y técnicas: Además de permitir la representación de los procesos, los formalismos en la Minería de procesos desempeñan un papel crucial en la aplicación de algoritmos y técnicas para analizar los datos de los procesos y extraer conocimiento que permite la mejora y optimización de los procesos (Urrea-Contreras et al., 2021). Los formalismos proporcionan una estructura y un marco de referencia para aplicar estas técnicas de análisis de datos. Por ejemplo, mediante el uso de redes de Petri, se pueden aplicar algoritmos de descubrimiento de procesos que permiten identificar patrones recurrentes, desviaciones, cuellos de botella y oportunidades de mejora en los procesos. Además, los formalismos permiten la definición de métricas y medidas de rendimiento que se utilizan en la evaluación y comparación de los procesos. Estas métricas pueden ser utilizadas para identificar áreas problemáticas, evaluar la eficiencia y efectividad de los procesos, y tomar decisiones basadas en datos para la optimización y reingeniería de los mismos.

Identificación de atributos: Las representaciones matemáticas muestran los diferentes atributos relevantes en el análisis de los procesos. Al utilizar fórmulas matemáticas, es posible expresar de manera precisa las relaciones entre los atributos, lo que facilita la identificación de patrones y tendencias entre ellas. Las representaciones matemáticas proporcionan un marco para la identificación y el análisis de los atributos en la Minería de Procesos, lo que contribuye a una comprensión más profunda y precisa de los procesos analizados. Tal es el caso de la definición de perspectiva por (Hompe et al., 2015) (Tabla 6), la cual contempla los atributos y la proyección de múltiples perspectivas. Por otro lado, más en lo particular se encuentra la descripción de la perspectiva de flujo de control por (De Leoni, Van der Aalst, & Dees, 2016), donde la descripción de esta perspectiva se realiza basada en las notaciones de manipulación de traza, exponiendo definiciones relacionadas al número de ejecuciones de actividad, primera ocurrencia de actividad, actividad previa en la traza y actividad actual.

Tabla 6. Definición y representación formal de perspectiva

Cita	Definición y representación formal
Hompe et al., 2015	<p>Sea P una perspectiva. $\uparrow_P : C \rightarrow R^m$ denota la función que asigna un caso a un vector real de longitud m según la perspectiva P. m es el número de atributos en P, por ejemplo, el número de recursos diferentes en el registro. c/P denota la proyección del caso $c \in L$ a una perspectiva P. Además que $c/\{P_1, P_2, \dots, P_K\} = c[\uparrow_{P_1}]_1 \parallel (c[\uparrow_{P_2}]) \parallel \dots \parallel \uparrow_{P_K}$, es decir, se concatenan los vectores de perfil resultantes de la proyección a múltiples perspectivas.</p>

Por otro lado, el modelar un proceso de software por medio de una red Petri permite conocer el flujo de trabajo, las actividades, transiciones y condiciones que pueden presentar. En la Fig. 1, se observa como en la red Petri las posiciones (P) se representan como círculos y se numeran de P_1 al P_{14} para indicar su identificación única. Las transiciones (T) se representan como barras verticales y también se numeran del T_1 al T_{14} . Mediante flechas dirigidas, se indican las relaciones de incidencia entre las posiciones y las transiciones. Una flecha que va de una posición a una transición indica que la transición consume recursos de esa posición (por ejemplo de P_2 a T_2 o T_3), mientras que una flecha que va de una transición a una posición indica que la transición genera o produce recursos en esa posición (por ejemplo de T_{10} a P_{11}). Esta representación permite visualizar las interacciones entre las posiciones y las transiciones, así como sus conjuntos, proporcionando una comprensión clara de cómo se produce el flujo de recursos en el sistema.

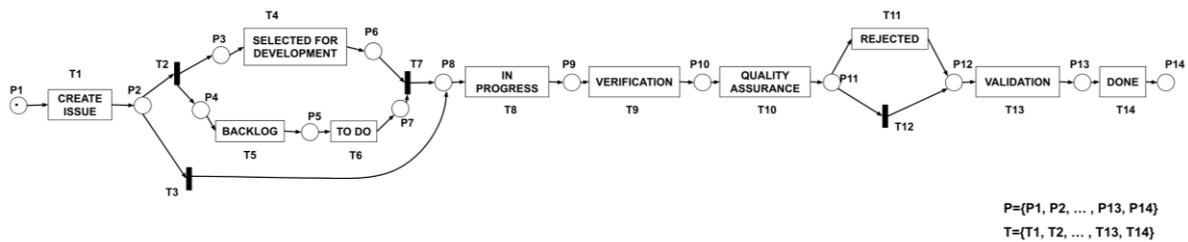


Figura 1. Red Petri de los estados de transición de un proceso de software

Se observa como existen flujos paralelos entre las posiciones P3 y P4 que dependen de los eventos que describen el inicio de la actividad y su terminación entre P6 y P7 respectivamente. El analizar en conjunto y aplicar las definiciones de caso y traza se pueden institucionalizar mejores prácticas y/o enfoques para la gestión de proyectos de software, la asignación de recursos pero sobre todo, la toma de decisiones basadas en diagnósticos, la obtención de nuevos modelos o los datos existentes y generados por las empresas de desarrollo de software.

5. DISCUSIÓN DE RESULTADOS

Tomando las descripciones anteriores y las definiciones basadas en conjuntos de los elementos de la Minería de Procesos (de la Tabla 1 a la Tabla 5), se establecieron las siguientes definiciones:

Definición 2 (Evento). Un evento e es una asignación de valores, el cual se describe mediante un conjunto de atributos. Sea $e \in \varepsilon$ para cualquier evento, $a \in A$ para los atributos y $v \in V$ para los valores, denotamos que, $e = ((a_1, v_1), (a_2, v_2), \dots, (a_n, v_n))$, en donde $a \subseteq v$.

Definición 3 (Caso). El conjunto de casos es $C \subseteq \varepsilon^*$, es decir, un conjunto de secuencias finitas de eventos. Cada evento aparece como máximo una vez en algún caso. Un caso $c = \langle c_1, c_2, \dots, c_n \rangle \in C$. Los casos también pueden tener atributos, así que para cada caso c , un atributo $a \in A$, definimos $c = \langle (c_1, a_1), (c_2, a_2), \dots, (c_n, a_n) \rangle$.

Definición 4 (Traza). Una traza puede ser vista como un caso, secuencia finita de eventos. Sea t una traza y ε^* el conjunto de todas las secuencias finitas sobre ε , $t \in \varepsilon^*$ por lo que $t = \langle e_1, e_2, \dots, e_n \rangle$, donde n es el número de eventos por traza.

Definición 5 (Registro de eventos). Un registro de eventos L es una secuencia finita de trazas donde $L = (t_1, t_2, \dots, t_n)$.

Definición 6 (Petri net). Una red de Petri es una tupla $N = (P, T, F, \ell)$ donde:
 P es un conjunto de lugares (posiciones).
 T es un conjunto de transiciones, $P \cap T = \emptyset$ (transiciones, conjuntos finitos).
 $F \subseteq (T \times P) \cup (P \times T)$ es una relación de flujo (conjunto de arcos dirigidos).
 $\ell: T \mapsto \mathbb{A} \cup \{\tau\}$ es una función de etiquetado.

De acuerdo a la definición de los elementos de Minería de Procesos y sus relaciones al extraer los datos de los repositorios se sugiere el siguiente orden: identificar los eventos (Definición 2) agrupándolos en sus correspondientes casos (Definición 3) para crear un registro de eventos (Definición 5). Una vez definido el registro de eventos, se genera un modelo basado en los datos extraídos que puede ser representado con una red Petri (Definición 6) para posteriormente verificar la conformidad de los eventos que es representado por medio de las trazas (Definición 4). El modelo podría extenderse con las perspectivas de Minería de Procesos con la identificación de los atributos correspondientes (Tabla 7). De esta manera, la Tabla 7 presenta la relación de atributos que se utilizan en el proceso de minado de las perspectivas de Minería de Procesos. Estos atributos son variables que se extraen y analizan para obtener información relevante sobre el flujo, el tiempo y la relación organizacional de los procesos. Asimismo, la Tabla 7 muestra una lista de los atributos relacionados a las perspectivas de flujo de control, organizacional, del caso y del tiempo, junto con ejemplificaciones de atributos como el orden de actividades, *timestamp* y caso. Estos atributos pueden incluir características como la duración de una actividad, el estado de un evento, el rol de un participante, la secuencia de actividades, entre otros.

La relación de atributos en la Tabla 7 proporcionaría una guía para los investigadores y profesionales de Ingeniería de Software que realizan minado de procesos, ayudándoles a identificar los atributos relevantes que deben considerar en su análisis. Asimismo, esta relación de atributos sirve como un recurso útil para comprender y aplicar las perspectivas de Minería de Procesos en diferentes contextos y proyectos.

Es importante destacar que los atributos específicos y su relevancia pueden variar dependiendo del dominio de aplicación y los objetivos del minado. Por lo tanto, esta tabla brindaría una visión general de los atributos comunes utilizados en el minado de perspectivas, pero se recomienda adaptarla y ajustarla según las necesidades y particularidades de cada proyecto.

Tabla 7. Atributos para el minado de perspectivas de Minería de Procesos de software

Flujo de Control	Organizacional	Caso	Tiempo
Actividad	ID Actividad	Caso***	<i>Timestamp*</i>
<i>Orden Actividades*</i>	Actividades (Tipo)		Horas invertidas
	<i>Timestamp*</i>		Horas efectivas
	Horas invertidas		
	<i>Recursos**</i>		
<i>Orden Actividades*</i> Timestamp (Inicio/Fin) <ul style="list-style-type: none"> - Fecha DD/MM/AA - Hora HH:MM 			
	<i>Recursos**</i> Roles Estructuras organizacionales <ul style="list-style-type: none"> - Departamentos - Equipos/proyectos 		
		Caso*** El atributo de caso dependerá del análisis	

CONCLUSIONES

Actualmente, la comunidad de Ingeniería de Software desea desarrollar estrategias basadas en conocimiento y orientadas al análisis de datos, Inteligencia Artificial, la Minería de Procesos de Software, entre otras, aplicables a las diversas problemáticas del desarrollo de proyectos con el propósito de fortalecer o alcanzar altos niveles de capacidad de procesos. En este documento, se presentan los formalismos de los elementos de la Minería de Procesos junto a su representación matemática, proporcionando una representación semántica de los conceptos y relaciones dentro de los procesos, evitando posibles confusiones derivadas del lenguaje natural. Las representaciones matemáticas del marco referencial también permiten identificar patrones que pueden pasar desapercibidos en las descripciones verbales o visuales, para su comprensión e implementación en la mejora de procesos de software. Como trabajo futuro, se realizará una sesión de *Focus Group* para la socialización y validación del marco referencial obtenido.

AGRADECIMIENTOS

El primer autor agradece al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT) por el financiamiento brindado bajo la referencia CVU-841596.

REFERENCIAS

Caldeira, J., & e Abreu, F. B. (2016). Software development process mining: Discovery, conformance checking and enhancement. In 2016 10th International Conference on the Quality of Information and Communications Technology (pp. 254-259). IEEE.

Dakic, D., Stefanovic, D., Lolic, T., Narandzic, D., & Simeunovic, N. (2020). Event log extraction for the purpose of process mining: a systematic literature review. In Innovation in Sustainable Management and Entrepreneurship: 2019 International Symposium in Management (SIM 2019) (pp. 299-312). Springer International Publishing.

De Leoni, M., Van der Aalst, W. M., & Dees, M. (2016). A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. *Information Systems*, 56, 235-257.

Diamantini, C., Genga, L., & Potena, D. (2016). Behavioral process mining for unstructured processes. *Journal of Intelligent Information Systems*. 47 (1), 5-32.

Dong, L., Liu, B., Li, Z., Wu, O., Babar M. A., & Xue, B. (2017). A mapping study on mining software process. *Software Engineering Conference (APSEC)* (pp. 51-60). IEEE.

Erdem, S., Demirörs, O., & Rabhi, F. (2018). Systematic mapping study on process mining in agile software development. In *Software Process Improvement and Capability Determination: 18th International Conference (SPICE 2018)* (pp. 289-299). Springer International Publishing. https://doi.org/10.1007/978-3-030-00623-5_20

Hompes, B., Buijs, J. C. A. M., Van der Aalst, W. M., P. Dixit, & Buurman, J. (2015). Discovering deviating cases and process variants using trace clustering. In *Proceedings of the 27th Benelux Conference on Artificial Intelligence (BNAIC)*, 5 (6).

Jokonowo, B., Sarno, R. Rochimah, S., & Priambodo, B. (2019). Process mining: Measuring key performance indicator container dwell time. *Indones. J. Electr. Eng. Comput. Sci*, 16 (1), 401-411.

- Kalenkova, A. A., Van der Aalst, W. M., Lomazova, I. A., & Rubin, V. A. (2016). Process mining using BPMN: relating event logs and process models. In Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems (pp. 123-123). IEEE.
- Keith B., & Vega, V. (2016). Process mining applications in software engineering, In Trends and Applications in Software Engineering: Proceedings of CIMPS 2016 (pp. 47-56). [https://doi: 10.1007/978-3-319-48523-2_5](https://doi.org/10.1007/978-3-319-48523-2_5)
- Kneuper, R. (2018). Selected Current Trends in Software Processes. In Software Processes and Life Cycle Models, (pp. 303-326). Springer.
- Leemans, M. (2018). Hierarchical Process Mining for Scalable Software Analysis. Ph.D. dissertation, Eindhoven University of Technology.
- Leemans, S. J., van Zelst, S. J., & Lu, X. (2023). Partial-order-based process mining: a survey and Outlook. Knowledge and Information Systems, 65 (1), 1-29.
- Li, G. (2019). Process mining based on object-centric behavioral constraint (OCBC) models. Ph.D. Thesis, Eindhoven University of Technology, Mathematics and Computer Science, Eindhoven, The Netherlands.
- Liu, C., Van Dongen, B., Assay N., & Van der Aalst, W. M. (2016). Component behavior discovery from software execution data. In 2016 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1-8). IEEE. [https://doi: 10.1109/SSCI.2016.7849947](https://doi.org/10.1109/SSCI.2016.7849947)
- Marin-Castro, H. M., & Tello-Leal, E. (2021). Event log preprocessing for process mining: a review. Applied Sciences, 11 (22), 10556. <https://doi.org/10.3390/app112210556>
- Marques, R., da Silva, M. M., & Ferreira, D. R. (2018). Assessing agile software development processes with process mining: A case study. In 2018 IEEE 20th Conference on Business Informatics (CBI) (pp. 109-118). IEEE Press.
- Parker, C., Scott, S., & Geddes, A. (2019). Snowball sampling. SAGE research methods foundations.
- Rivas, M. H., & Bayona-Oré, S. (2019). Algoritmos de minería de proceso para el descubrimiento automático de procesos. Revista Ibérica de Sistemas e Tecnologías de Informação (RISTI), 31, 33-49.
- Rubin, V., Lomazova, I., & Van der Aalst, W. M. (2014). Agile development with software process mining, In Proceedings of the 2014 International conference on software and system process (pp. 70-74). <https://doi.org/10.1145/2600821.2600842>
- Seeliger, A., Nolle, T., & Mühlhäuser, M. (2018). Process explorer: an interactive visual recommendation system for process mining. In KDD Workshop on Interactive Data Exploration and Analytics.
- Senderovich, A. (2017). Queue Mining: Service Perspectives in Process Mining. In BPM (Demos).
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of business research, 104, 333-339.
- Štolfa, J. (2018). Application of Process Mining in Intelligent Process Support. Ph.D. Thesis.

- Syamsiyah, A. (2020). In-database preprocessing for process mining. Technische Universiteit Eindhoven.
- Urrea-Contreras, S. J., Flores-Rios, B. L., Astorga-Vargas, M. A., & Ibarra-Esquer, J. E. (2021). Process Mining Perspectives in Software Engineering: A Systematic Literature Review. In 2021 Mexican International Conference on Computer Science (ENC) (pp. 1-8). IEEE Press. [https://doi: 10.1109/ENC53357.2021.9534824](https://doi.org/10.1109/ENC53357.2021.9534824)
- Urrea-Contreras, S. J., Flores-Rios, B. L., Astorga-Vargas, M. A., Ibarra-Esquer, J. E., & González-Navarro, F. F. (2018). MoProPEI-MP an Adaptation in the Selection of Process Mining Tools for Software Processes. Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento 2018 (JIISIC) (pp. 151-162).
- Urrea-Contreras, S. J., Flores-Rios, B. L., Astorga-Vargas, M. A., Ibarra-Esquer, J. E., González-Navarro, F. F., García Pacheco, I. A., & Pacheco Agüero, C. L. (2022). Process Mining Model Integrated with Control Flow, Case, Organizational and Time Perspectives in a Software Development Project. International Conference in Software Engineering Research and Innovation (CONISOFT) (pp. 92-101). IEEE. [https://doi: 10.1109/CONISOFT55708.2022.00022](https://doi.org/10.1109/CONISOFT55708.2022.00022)
- Van der Aalst, W. M. (2015). Extracting event data from databases to unleash process mining, In BPM-Driving innovation in a digital world. (pp. 105-128). Springer.
- Van der Aalst, W. M. (2016). Data science in action. In Process mining, Springer, Berlin, Heidelberg.
- Van Eck, M. L. (2022). Process Mining for Smart Product Design. Eindhoven University of Technology.
- Van Zelst, S. J. (2019). Process mining with streaming data. Technische Universiteit Eindhoven.

NOTAS BIOGRÁFICAS

Silvia Jaqueline Urrea Contreras Doctora en Ciencias en el área de la Computación por la Universidad Autónoma de Baja California. Sus áreas de investigación son Minería de procesos aplicada a la Ingeniería de software y Mejora de procesos de software. Cuenta con experiencia profesional como miembro de equipos de proyectos de software.

Brenda Leticia Flores Rios Doctora en Ciencias en Computación por UABC. Actualmente, es Coordinadora de Investigación y Posgrado del Instituto de Ingeniería UABC. Miembro del Cuerpo Académico de Cómputo Científico nivel consolidado, de la Academia Mexicana de Computación y de la Red Mexicana de Ingeniería de software. Cuenta con el reconocimiento PRODEP-SEP y SNI nivel I. Sus áreas de investigación son Ingeniería de software, Mejora de procesos de software e Ingeniería del conocimiento.

María Angélica Astorga Vargas Doctora en Ciencias en Computación por la Universidad Autónoma de Baja California. Profesora-investigadora del Programa Educativo de Licenciado en Sistemas Computacionales en la Facultad de Ingeniería UABC. Miembro del Cuerpo Académico de Cómputo Científico y de la Academia Mexicana de Computación. Cuenta con el reconocimiento al Perfil Deseable PRODEP-SEP y SNI nivel I. Sus áreas de investigación son la Ingeniería de software y la mejora de procesos de software con enfoque en la conformación efectiva de equipos de proyectos de software.

Jorge Eduardo Ibarra Esquer es profesor investigador adscrito al programa educativo Ingeniero en Computación y al programa de Maestría y Doctorado en Ciencias e Ingeniería en la Facultad de Ingeniería de la UABC. Cuenta con el reconocimiento como Investigador Nivel I y con el Reconocimiento a Perfil Deseable que otorga el Programa PRODEP-SEP. Entre sus áreas de interés se encuentran el Internet de las Cosas, la Minería de Datos, y la aplicación de esta última al descubrimiento de conocimiento.

Félix Fernando González-Navarro Doctor por la Universidad Politécnica de Cataluña, España. Cuenta con el reconocimiento como Investigador Nivel I y con el Reconocimiento a Perfil Deseable PRODEP-SEP. Es líder del Cuerpo Académico de Cómputo Científico nivel consolidado, miembro de la Academia Mexicana de Computación. Las líneas de investigación son Aprendizaje Máquina y Minería de Datos.

Iván A. Gacía Pacheco cursó los estudios de posgrado en la Universidad Politécnica de Madrid, España. Actualmente se desempeña como profesor-investigador en la División de Estudios de Posgrado de la Universidad Tecnológica de la Mixteca. Es miembro del SNI nivel 1. Los intereses son la Ingeniería de Software, específicamente los métodos de evaluación y mejora del proceso software aplicados a las pequeñas empresas, el desarrollo de herramientas educativas para mejorar la formación de los futuros ingenieros de software, y el desarrollo de herramientas computacionales para reducir el esfuerzo y maximizar el beneficio económico de los pequeños entornos de desarrollo.

