

## **Adopción de Herramientas de Minería de Procesos en la Ingeniería de Software**

Silvia Jaqueline Urrea-Contreras<sup>1</sup>, Brenda L. Flores-Rios<sup>2</sup>, María Angélica Astorga-Vargas<sup>3</sup>,  
Jorge Eduardo Ibarra-Esquer<sup>4</sup>, Larysa Burtseva<sup>5</sup>

<sup>1, 2, 5</sup> Instituto de Ingeniería, Universidad Autónoma de Baja California

Blvd. Benito Juárez y Calle de la Normal s/n Col. Insurgentes Este. CP. 21280, Mexicali, Baja California, México

{jaqueline.urrea, brenda.flores, burtseva}@uabc.edu.mx

<sup>3, 4</sup> Facultad de Ingeniería, Universidad Autónoma de Baja California

Calle de la Normal s/n Col. Insurgentes Este. CP. 21280, Mexicali, Baja California, México

{angelicaastorga, jorge.ibarra}@uabc.edu.mx

### **Abstract**

Some software development organizations expose the need for faced selecting and adopting Process Mining tools, algorithms and techniques in order to exploit their event logs as evidence of the execution of their development projects and/or software process improvement. Currently, there is a variety of Process Mining tools with a commercial and open-source approach, but the functionality is limited for experts in the area. A framework that includes a reference model for the process of selection and adoption of Process Mining tools and algorithms and a descriptive guide that provides knowledge to the software roles is presented. The elements of the MoProPEI-MP model and the identification of the main Process Mining tools reported in a literature review are presented as partial results.

### **Resumen**

Se expone la necesidad que enfrentan algunas organizaciones de desarrollo de software al seleccionar y adoptar herramientas, algoritmos y técnicas de Minería de Procesos para explotar sus registros de eventos como una evidencia de la ejecución de sus proyectos de desarrollo y/o de mejora de procesos de software. Actualmente, existe una variedad de herramientas con enfoque de Minería de Procesos de tipo comercial y de uso libre, pero su alcance en funcionalidad queda limitado para expertos del área. Se presenta un marco de trabajo que contemple un modelo de referencia para el proceso de selección y adopción de herramientas y algoritmos de Minería de Procesos y una guía descriptiva que brinde conocimiento a los roles de software. Se presentan como resultados parciales los elementos del modelo MoProPEI-MP y la identificación de las principales herramientas de Minería de Procesos reportadas en una revisión de literatura.

---

*Keywords and phrases:* Process Mining, Process Mining Tools, Software Engineering, Software Process Improvement.

2010 Mathematics Subject Classification: 67N30

---

## 1 Introducción

La Minería de Procesos (MP) es una disciplina que utiliza herramientas de software, algoritmos y técnicas basados en la minería de datos, la cual permite analizar procesos, extraer o explotar la información útil o conocimiento almacenado en los repositorios, intranets o sistemas de información [1, 2]. Además, permite revisar de manera más rigurosa, el cumplimiento de normativas y establecer la validez y confiabilidad de la información acerca de los procesos críticos de una organización [3, 4]. Algunos autores han documentado, por medio de reportes técnicos, casos de estudio y proyectos de MP, cómo las organizaciones se han beneficiado en el diagnóstico, análisis y mejora de procesos, a partir de los datos y registros de eventos relacionados con la ejecución real de los procesos [5, 6]. Un caso particular, es la aplicación de la MP en proyectos de desarrollo de software [7, 8] o mejora de procesos de software (*Software Process Improvement* - SPI) [1, 5, 9].

Las organizaciones dedicadas al desarrollo de software, como parte de sus mejores prácticas, utilizan estándares y modelos de procesos de software para asegurar la alta calidad de sus productos. Sin embargo, sólo en algunas organizaciones la visión de negocios es hacer un uso óptimo de los recursos tecnológicos para elevar su madurez organizacional, tales como intranet, bases de conocimiento, herramientas o sistemas de información. Se han detectado organizaciones de software que enfrentan desafíos en la gestión e implementación de proyectos, de software o SPI, destacando entre ellos la necesidad de contar con herramientas para recopilar, registrar, clasificar y evaluar la evidencia de la ejecución de las actividades cuando el volumen y complejidad de la información son significativos [10, 11].

Ante este escenario, la MP ofrece ventajas con respecto a las técnicas tradicionales de análisis y mejora de procesos, debido a que brinda resultados en menor tiempo y con mayor fiabilidad [4] haciendo uso de herramientas de software, algoritmos y técnicas. Por ejemplo, consideramos que la implementación de herramientas de MP y el análisis de los registros de eventos proveerán de nuevos requerimientos para el desarrollo y mejora de herramientas orientadas a la evaluación de procesos de software. Estas deberán contar con características de una mayor robustez, análisis de información, capacidad de procesamiento de gran volumen de registros de eventos y por ende, impactar en la capacidad de los procesos de software.

En este artículo, se enfatiza como tema crítico y de interés para las organizaciones de desarrollo de software el proceso de selección y adopción de herramientas para el modelado, análisis, seguimiento y explotación de información, así como de herramientas, algoritmos y técnicas de MP. Por tal motivo, existe la necesidad de contar con un marco de trabajo que contemple un modelo de referencia para

el proceso de selección y adopción de herramientas de software, algoritmos y técnicas de MP y una guía descriptiva que presente conocimiento tópico relacionado a los algoritmos y técnicas de MP. Este marco de trabajo se desarrolla como un proyecto de posgrado en la Universidad Autónoma de Baja California; y está orientado a las organizaciones dedicadas al desarrollo de software que requieran apoyo en sus proyectos de ingeniería de explotación de información y/o análisis de sus registros de eventos de procesos de software.

El resto del documento se estructura de la siguiente manera: en la sección 2 se expone la problemática de la insuficiente implementación de la MP en las organizaciones de desarrollo de software, lo que permitió puntualizar objetivos que guíen el desarrollo del proyecto (sección 3); en la sección 4 se definen las etapas de la metodología utilizada, de las cuales en la sección 5 se presentan los resultados obtenidos hasta el momento; por último, las conclusiones y trabajo futuro se muestran en la sección 6.

## **2 Problemática**

En las organizaciones dedicadas al desarrollo de software existe personal o áreas que realizan análisis de tendencias tecnológicas orientadas a la toma de decisiones sobre la adquisición, adopción, perfeccionamiento o comercialización de tecnologías requeridas, procesos de investigación y desarrollo (I+D), innovación y transferencia de conocimiento o tecnología quienes consideran la interacción de estas con su medio ambiente o ecosistema. Por ejemplo, el ciclo de conocimiento es una estrategia de análisis de flujos de conocimiento en la gestión de proyectos o implementación de un proyecto SPI [12], que ayuda a los responsables de la base de conocimiento en la identificación de fuentes externas de información, tanto humanos como tecnológicos, especifica los informes de las auditorías de base tecnológica y necesidades tecnológicas requeridas por la organización de software. El resultado de su implementación es generar nuevo conocimiento, tanto tácito como explícito, que retroalimente los activos de conocimiento existentes en los equipos de trabajo, y a su vez sea entrada para la toma de decisiones según el ecosistema donde se encuentre. Para un líder de proyectos de software o responsable de un proyecto de SPI los datos en sí no son lo más relevante, sino el conocimiento que se encierra en sus relaciones, fluctuaciones y dependencias [13].

Según [14], aun cuando los proyectos de desarrollo de software poseen características distintas a los proyectos de Ingeniería de explotación de información, sobre todo en la parte operativa del proyecto, esta constituye una subdisciplina de la informática que aporta al área de la Inteligencia de Negocios [15]. El objetivo de la Ingeniería de explotación de información es hacer explícita la información que brinde conocimiento para que las organizaciones puedan entender y mejorar sus procesos [5]. Asimismo, especificar metodologías para controlar y gestionar las tareas en la búsqueda de patrones de conocimiento dentro de un gran volumen de información [16]. Otro aspecto del éxito o fracaso de la explotación de la información no sólo se basa en la generación de conocimiento, sino también de la adecuada validación del conocimiento tácito de un experto y adaptarlo para que pueda ser exteriorizado ante todos los miembros del proyecto o grupo de SPI, quienes pueden contar con diferentes niveles de experiencia en el área [17].

Por otro lado, en [18] se indica que el limitado acceso a los datos e información relacionada a la ejecución de las actividades de los proyectos, impide tomar decisiones necesarias en el

descubrimiento de nuevas reglas de negocio, monitoreo y control y mejora continua de los procesos ya establecidos. La tendencia de gestión relacionada a la mejora de procesos es utilizar la MP [3], la cual es una disciplina emergente que está siendo incluida, tanto por investigadores como practicantes, en temas de Ingeniería de Software (IS) [1].

La literatura expone que una organización dedicada al desarrollo de software puede aprovechar sus registros de eventos, para proveer un mejor entendimiento de los procesos y de cómo mejorar su rendimiento ante las actividades requeridas por el proceso de desarrollo de software [7, 8], el modelo de referencia de procesos o un ciclo de SPI [1, 5, 9]. Es decir, los modelos de proceso descubiertos pueden ser utilizados para generar ideas de mejora de procesos, observar el proceso real e identificar problemáticas [1]. Sin embargo, también se expone que la parte más lenta de la MP no es el análisis en sí, sino que la mayor parte del tiempo se invierte en localizar, seleccionar, convertir y filtrar los registros de eventos [19]. Se entiende por registros de eventos a los datos de la ejecución de las actividades de procesos, deben de indicar el identificador de la actividad, una descripción de la actividad, quién realiza la actividad (rol(es)), tiempo de ejecución [19, 20].

Al evaluar las principales metodologías existentes para los proyectos de explotación de información, se identificó la falta de herramientas que soporten de forma completa la fase de gestión de proyectos [17, 21]. En Iberoamérica, la experiencia (conocimiento tácito) que han adquirido las organizaciones de desarrollo de software en proyectos SPI, en un periodo aproximado de una década, las ha desafiado que para alcanzar una alta madurez busquen herramientas, algoritmos y técnicas de MP o de inteligencia de negocios, que les permitan revisar de forma más rigurosa el cumplimiento de normas o modelos de referencia de procesos, establecer la validez y confiabilidad del análisis de evidencias y/o buscar soluciones a problemas en la escalabilidad de información.

Actualmente, existe una diversidad de algoritmos y herramientas de MP [22] disponibles de manera comercial o de uso libre, por mencionar algunas: Disco desarrollada por Fyxicon, Aris Process Performance Manager (PPM) por Software AG, Futura Reflect por Futura Process Intelligence y ProcessAnalyzer por QPR, PROM, entre otras; que al mismo tiempo representan una dificultad de comparación entre ellas debido a que muestran diferencias en funcionalidad y desempeño [1]. Otro problema que se presenta es la falta de estandarización y el hecho de que las herramientas usan sus propios formatos o extensiones requeridas como entrada por ellas, tales como csv, xls, mxml o xes. Esto se vuelve un escenario confuso para el usuario de MP al plantearse lo siguiente: ¿Cuál es la herramienta más adecuada para analizar una actividad [23]?, ¿Necesito utilizar más de un algoritmo o técnica de MP según el tipo de análisis? y/o ¿Requiero de un conocimiento técnico para los diversos tipos de formato que manejan las herramientas de MP?

En virtud de este contexto, se resalta la importancia de contar con un marco de trabajo para la selección de herramientas y algoritmos de MP que oriente a las organizaciones de desarrollo de software en el análisis cualitativo o gestión cuantitativa de procesos y que brinde conocimiento técnico sobre algoritmos y técnicas de MP a roles que participan en gestión de proyectos de desarrollo de software o de SPI [24].

### 3 Objetivos

El objetivo general es: Definir un marco de trabajo para la selección y adopción de herramientas, algoritmos y técnicas de MP, como apoyo a proyectos de desarrollo de software y de SPI. A su vez, se puntualizaron cuatro objetivos específicos para un periodo de dos años:

- O1. Diseñar un modelo de referencia para la adopción y selección de herramientas y algoritmos de MP orientado a equipos de desarrollo de software o de SPI.
- O2. Definir una guía técnica con las características de los principales algoritmos y técnicas de MP.
- O3. Categorizar los criterios, elementos o factores necesarios como aspectos socio-técnico existente en las áreas de análisis de tendencias tecnológicas de las organizaciones de desarrollo de software.
- O4. Validar los resultados con la técnica de *Focus Group* orientada a IS.

Con el propósito de obtener un contexto en la disciplina de IS y la MP, se establecieron dos preguntas de investigación:

**RQ1:** Bajo un proyecto SPI, ¿Cuáles son los modelos de referencia o evaluación de procesos que se han apoyado de la selección y adopción de alguna herramienta o tecnología de MP?

**RQ2:** ¿Cuáles son los algoritmos y herramientas de MP que se han utilizado en casos de estudio, proyectos de desarrollo de software o de SPI?

### 4 Metodología

Para el cumplimiento del objetivo general, se seguirá una metodología bajo un enfoque mixto propuesta por cuatro etapas. La primera etapa trata de una revisión sistemática de literatura siguiendo el protocolo de [25], con el objetivo de contar con el marco teórico y el estado del arte relacionado a la selección y adopción de herramientas con capacidad de análisis de técnicas y/o algoritmos, propias de la MP que les permita a las organizaciones desarrolladoras de software explorar y explotar los registros de eventos contenidos en sus repositorios.

La segunda etapa está enfocada al diseño del modelo de referencia para el proceso de selección y adopción de herramientas de MP tomando en consideración las metodologías para el desarrollo de proyectos y selección de herramientas de explotación de información, bajo un enfoque de MP. La tercera etapa identifica y categoriza los factores o criterios que las organizaciones podrán considerar para hacer uso eficiente del modelo de referencia propuesto y de esa forma guiar, fácil y técnicamente, la selección de la(s) herramientas de MP y adoptarla(s) para el análisis de sus registros de eventos.

Por último, la cuarta etapa cumple con la validación del marco de trabajo propuesto, por medio de la técnica de *Focus Group* con expertos en desarrollo de proyectos de software o de SPI y/o responsables de los repositorios o base de conocimiento en organizaciones de desarrollo de software de México y Colombia.

## 5 Resultados Obtenidos

Una vez realizadas las dos primeras etapas se obtuvieron los siguientes resultados:

### A. Etapa 1: Revisión sistemática de la literatura

Se generó una cadena de búsqueda a partir de los términos clave y relacionados presentados en la Fig. 1. La cadena de búsqueda se ejecutó en las fuentes de información: 1) IEEE Xplore, 2) Elsevier ScienceDirect, 3) Springer Link, y 4) ACM Digital Library; alcanzando un rango de 23,127 estudios encontrados, de los cuales sólo 17 cumplieron con los criterios de inclusión y exclusión, considerándolos como Estudios Primarios (EP). Cada uno de ellos se identificaron bajo el id EP $n$ , año en el que fue publicado (desde 2007-2017), autor(es), título, tipo de documento, validación del escenario de uso y el(los) nombre(s) de herramientas, algoritmos y/o técnicas de MP que se implementaron.

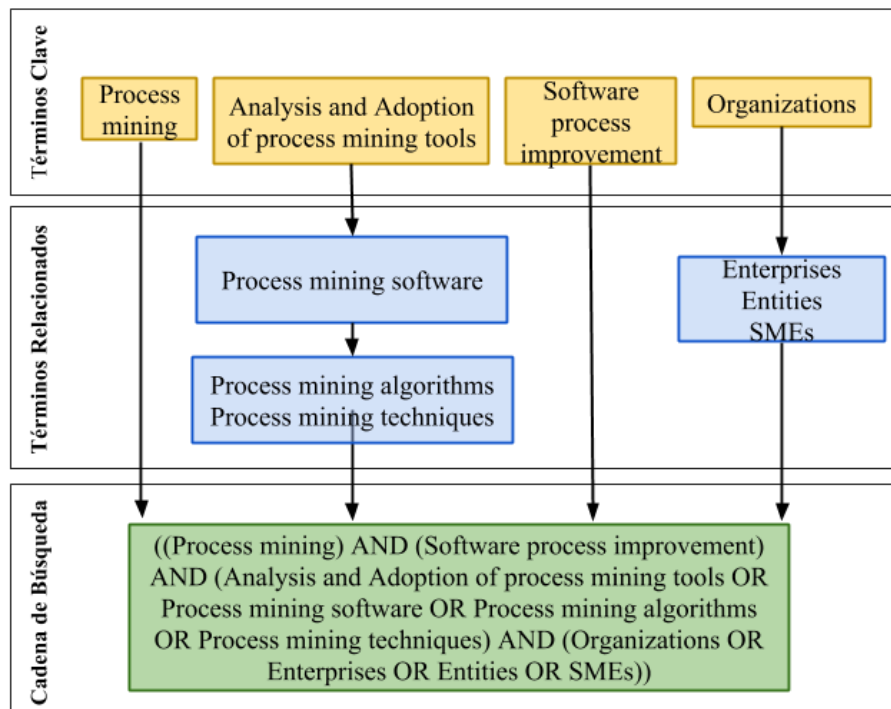


Figura 1. Generación de la cadena de búsqueda. Elaboración propia.

Con el propósito de generar un análisis y evaluación más objetiva y asertiva de los resultados de cada uno de los EP, éstos se distribuyeron en tres categorías: 1) Principios teóricos - expresa los fundamentos teóricos; 2) Propuestas - presenta las propuestas o iniciativas que se han establecido para la utilización de herramientas y técnicas de MP en SPI; y 3) Herramientas y técnicas - representa la variedad de herramientas tanto comerciales como de uso libre existentes. En la Fig. 2 se observa la correspondencia de las 3 categorías de los EP por medio de 3 conjuntos (A, B, C).

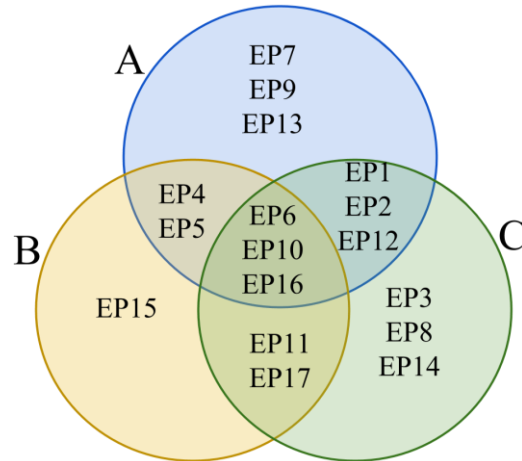


Figura 2. Representación de las 3 categorías relacionadas a los estudios primarios

Donde:

$$A = \{EP1, EP2, EP4, EP5, EP6, EP7, EP9, EP10, EP12, EP13, EP16\}$$

$$B = \{EP4, EP5, EP6, EP10, EP11, EP15, EP16, EP17\}$$

$$C = \{EP1, EP2, EP3, EP6, EP8, EP10, EP11, EP12, EP14, EP16, EP17\}$$

El conjunto A (categoría 1) cuenta con una cardinalidad de 11 EP, siguiendo el conjunto B (categoría 2) con 8 EP, y por último, el conjunto C (categoría 3) con 11 EP. El conjunto B está asociado a dar respuesta a la RQ1 debido a que demuestra la relación entre MP bajo proyectos de SPI. El conjunto C está asociado a la RQ2, la cual se basa en aquellas herramientas y algoritmos identificados y adoptados a un proyecto de MP.

Los EP clasificados en los conjuntos A y B también mencionan herramientas de MP, pero se limitan a sólo indicar los nombres de las herramientas como una referencia de trabajo, no su utilización o implementación a detalle. A diferencia de los elementos del conjunto C que están relacionados a la implementación de herramientas de MP en diversos casos de estudio, en su mayoría de manera exitosa, con una presencia del 64.70% del total de los EP. En la Fig. 2 también se observa que la intersección de los 3 conjuntos  $A \cap B \cap C = \{EP6, EP10, EP16\}$  es la evidencia de cómo convergen las categorías entre sí.

La Fig. 3 muestra el porcentaje del número de veces que fueron identificadas las 21 herramientas orientadas a la MP por los 17 EP. Por ejemplo, ProM es una herramienta de MP de uso libre con una capacidad de registros de eventos ilimitados que genera modelos BPMN, flujos de trabajo, redes Petri, entre otros; y ofrece un filtrado de información, descubrimiento de procesos, verificación de la conformidad y visualización de agrupamientos. Por tal motivo, su implementación se refleja en un 76.4% del total de EP, haciéndola como la herramienta más seleccionada por los autores en la disciplina de MP. A diferencia de las últimas 8 herramientas con un hallazgo de una frecuencia (5.8%) en los EP.

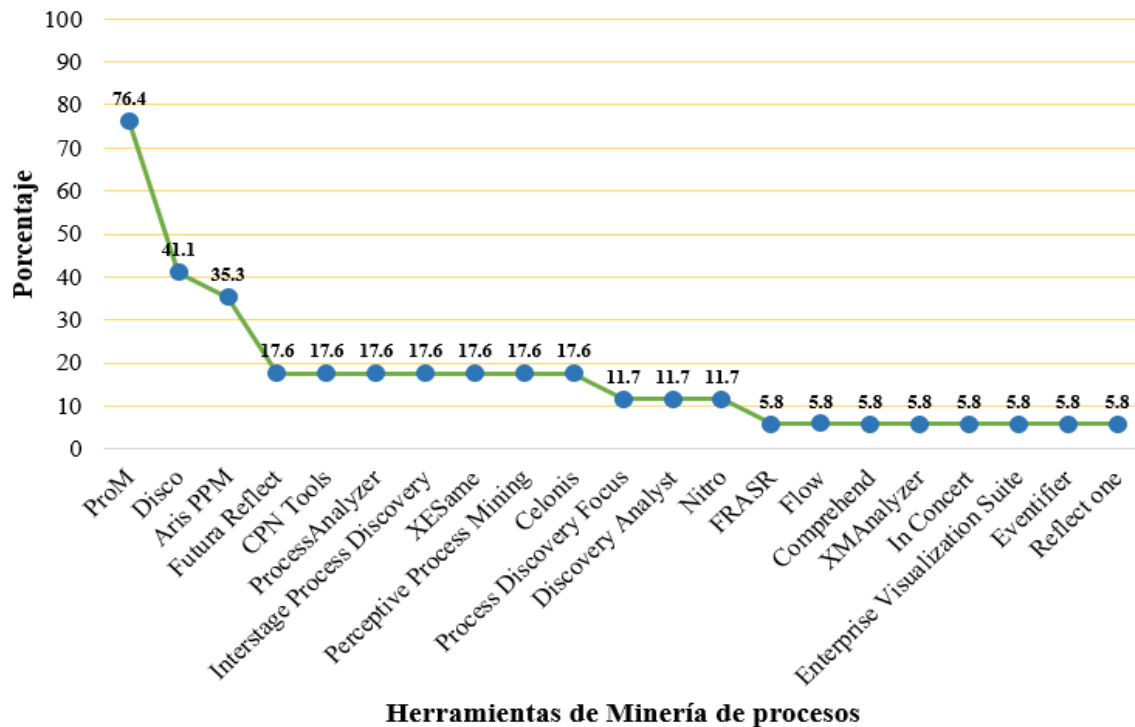


Figura 3. Distribución de porcentajes por tipo de herramienta

*B. Etapa 2: Modelado de referencia de procesos para la selección y adopción de herramientas y algoritmos de Minería de Procesos (MoProPEI-MP)*

Entre las metodologías para el desarrollo de proyectos explotación de información se destacan CRISP, P3TQ, SEMMA y Modelo de procesos para proyectos de explotación de información [14]. Asimismo, [26] propone la metodología para seleccionar herramientas considerando las fases de: 1) análisis de la necesidad – documentar las características del negocio que la herramienta debe atender; 2) Búsqueda en el mercado – Realizar un informe de criterios agrupados por características técnico-funcionales, del proveedor, servicio y costos; y, 3) Evaluación de productos.

Por otro lado, se identificó el Modelo de Procesos para la Ingeniería de Explotación de Información (MoProPEI) [17] que presenta como característica la adaptación a las necesidades de gestión de proyectos que se desarrolla y permite realizar adaptaciones a la ejecución de las actividades en contexto a la MP. De igual manera, el enfoque de MoProPEI hacia la gestión es fortalecer la planificación, administración y documentación de los aspectos necesarios para el desarrollo del proyecto [17]. Por tal motivo, fue seleccionado para adaptarse bajo un análisis de Ingeniería de Procesos. Donde los dos subprocesos de MoProPEI (Gestión y Desarrollo) fueron considerados como Categorías de procesos, las fases como procesos, y las actividades, tareas y Productos de Trabajo (PT) se especificaron con el propósito de identificar y seleccionar herramientas y algoritmos orientados a la MP [24]. El modelo de referencia de procesos resultante fue llamado MoProPEI-MP.



Específicamente, la categoría de Desarrollo se compone de seis procesos: Entendimiento del dominio, Entendimiento de los datos, Modelado, Preparación de los datos, Implementación, Evaluación y presentación. A partir de ellos, el proceso de Modelado contiene dos actividades relacionadas con el A1. Modelado del problema y A2. Configuración del modelo. A1 y A2 se definieron por medio de un patrón de procesos, el cual es un esquema de elementos que sirve para la especificación formal de las tareas, objetivos y los PT de entrada y salida (Tabla 1).

En la Tabla 1 se presentan los objetivos y las tareas de A1 y A2. El objetivo de A1 es definir el problema mediante un vocabulario técnico y representarlo con un registro de eventos de forma que sea posible identificar algoritmos de MP. Esta actividad comprende las siguientes tareas: T1.1 Definir el problema a resolver con base en los problemas del negocio identificados bajo un enfoque de MP y T1.2 Modelar el problema a resolver para la identificación de la solución bajo un enfoque de MP. La actividad A2 tiene tres objetivos específicos para determinar y formalizar la selección de la(s) herramienta(s) y los algoritmos de MP que se aplicarán en función del problema y las técnicas previamente identificadas. Las tareas a realizar son: T2.1 Identificar algoritmos de MP a partir de los problemas a resolver y la identificación de la solución, T2.2 Seleccionar algoritmos de MP con base a los algoritmos de MP identificados, T2.3 Identificar las herramientas de MP a partir de los problemas a resolver con la identificación de la solución y los algoritmos de MP seleccionados y T2.4 Seleccionar la(s) herramienta(s) de MP tomando en cuenta las identificadas [24].

MoProPEI-MP presenta siete PT asociados al proceso de Modelado de la categoría de Desarrollo, para las actividades A1 y A2. Estos son: PT1. Identificación de los problemas de negocio, PT2. Problemas a resolver, PT3. Identificación de la solución, PT4. Algoritmos de MP identificados, PT5. Algoritmos de MP seleccionados, PT6. Herramienta(s) de MP identificada(s), y PT7. Herramienta(s) de MP seleccionada(s). En la Tabla 1 se indican los PT de entrada y salida asociados a las seis tareas de ambas actividades. En el modelo original de MoProPEI, las actividades no estaban sujetas a un flujo de trabajo consecutivo. Sin embargo, en MoProPEI-MP se estableció un flujo de trabajo secuencial para brindar información en la selección más adecuada de la herramienta de MP según el problema identificado. La Fig. 4 muestra el flujo de los PT de todas las tareas del proceso de Modelado para alcanzar el objetivo. Por ejemplo, se visualiza cómo en A2 el PT de salida PT5 de la tarea T2.2 es un PT de entrada de la tarea T2.3 y el PT de salida de la tarea T2.3 (PT6) es un PT de entrada de la tarea T2.4.

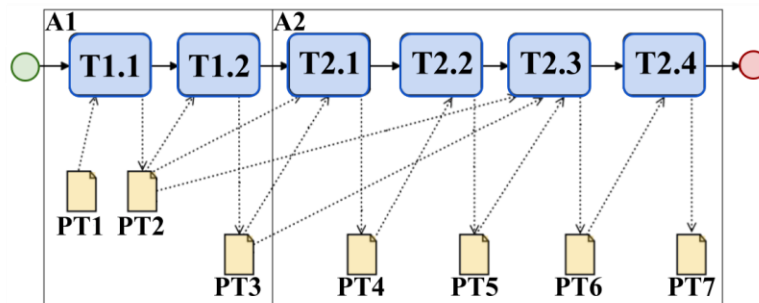


Figura 4. Flujo de los productos de trabajo relacionados a las actividades A1 y A2 para el proceso de Modelado de MoProPEI-MP

Tabla 1: Descripción del proceso de Modelado de MoProPEI-MP [24]

<b>Proceso de Modelado</b>	
Actividades	
A1 Modelado del problema.	A2 Configuración del modelo.
Objetivos	
O1.1 Determinar mediante vocabulario técnico los problemas a resolver de acuerdo a los problemas de negocio identificados. O1.2 Representar el problema mediante un registro de eventos o un modelo de procesos de forma que sea posible identificar los algoritmos de MP a utilizar.	O2.1 Determinar los algoritmos de MP a utilizar con base en las características de la identificación de la solución. O2.2 Determinar el conjunto de herramientas de MP existentes en el mercado que satisfacen las demandas de la identificación de la solución, analizando sus ventajas y desventajas y los algoritmos que soportan. O2.3 Formalizar la selección de la(s) herramienta(s) elegida(s) para la solución del problema.
Tareas	
T1.1 Definir el Problema a resolver con base en los Problemas de negocio identificados bajo un enfoque de MP. T1.2 Modelar el Problema a resolver para la Identificación de la solución bajo enfoque de MP.	T2.1 Identificar los Algoritmos de MP a partir de los Problemas a resolver y la Identificación de la solución. T2.2 Seleccionar los Algoritmos de MP con base a los Algoritmos de MP identificados. T2.3 Identificar las Herramientas de MP a partir de los Problemas a resolver con la Identificación de la solución y los Algoritmos de MP seleccionados. T2.4 Seleccionar la Herramienta de MP tomando en cuenta las Herramientas de MP identificadas.
Productos de trabajo de Entrada	
T1.1: PT1 T1.2: PT2	T2.1: PT2, PT3 T2.2: PT4 T2.3: PT2, PT3, PT5 T2.4: PT6
Productos de trabajo de Salida	
T1.1: PT2 T1.2: PT3	T2.1: PT4 T2.2: PT5 T2.3: PT6 T2.4: PT7

La identificación y selección de algoritmos y herramientas de MP exponen una medida favorable en la gestión del manejo de grandes volúmenes de información generados por las organizaciones de desarrollo de software, originando una pauta para la definición de un modelo de referencia que manifieste un conjunto de actividades orientadas a la aplicación de conocimiento tópico para la mejora de los resultados obtenidos con las herramientas de MP. Por todo lo anterior, se presenta la estructura de un Marco de Trabajo compuesta de una Guía Técnica con definiciones (conocimiento

tópico) relacionada a los algoritmos y técnica de MP, y un modelo de referencia con enfoque de MP factible y de fácil uso para las organizaciones de desarrollo de software (Fig. 5). Se espera que dicho marco de trabajo les permita garantizar la confiabilidad en la identificación y selección de herramientas, algoritmos y técnicas de MP a través de las seis tareas (T1.1, T1.2, T2.1, T2.2, T2.3, T2.4) que define el proceso de Modelado [24].

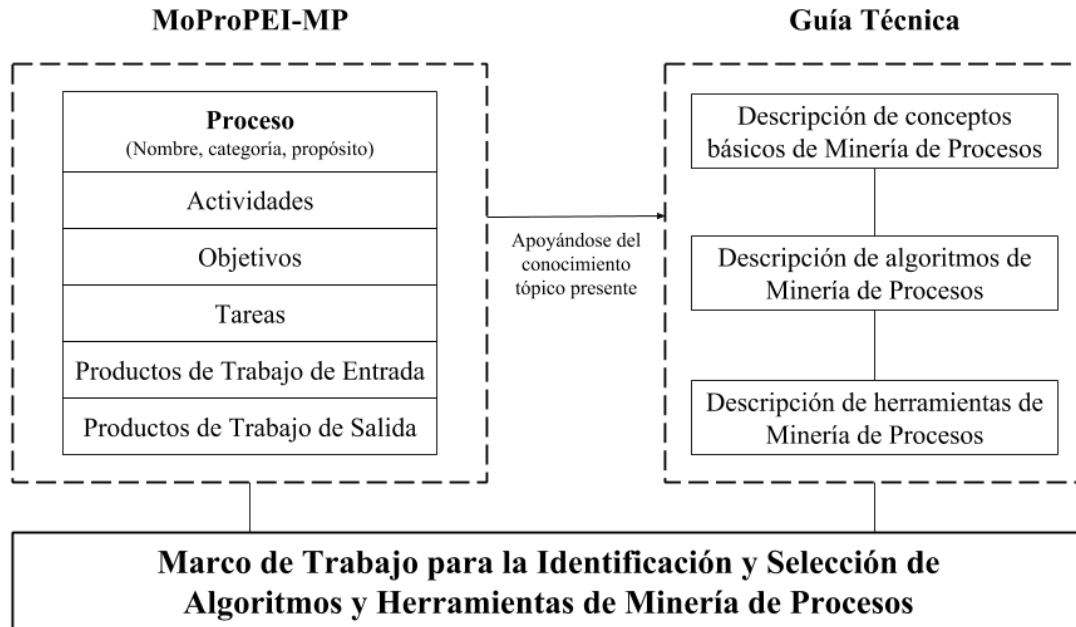


Figura 5. Diagrama General del Marco de trabajo para la selección de algoritmos y herramientas de Minería de Procesos

## 6 Conclusiones

La interdisciplinariedad que la MP ha contribuido a la IS se ha observado en los beneficios de: conocer la ejecución real de los procesos en comparación con el proceso documentado, descubrir nuevos modelos de procesos, generar conocimiento para analizar y mejorar los procesos, entre otros. Como toda disciplina nueva, se presentan desafíos, siendo uno de ellos la decisión de seleccionar herramientas, algoritmos y técnicas adecuadas para el análisis y exploración de los registros de eventos.

Para las etapas 1 y 2, se identifican las 21 herramientas de MP más implementadas por la literatura y se presenta MoProPEI-MP orientado a incitar a los responsables del área tendencias tecnológicas al uso de MP para explotar el conocimiento existente en sus intranets, sistemas de información o bases de conocimiento.

Para dar cumplimiento de las etapas 3 y 4, las cuales se realizarán en el segundo año del proyecto de investigación, se buscará que MoProPEI-MP cuente con una guía tópica de términos específicos

a algoritmos y técnicas de MP con el único fin de hacer accesible la información de MP que se requiera por las organizaciones de desarrollo de software.

### **Agradecimientos**

El primer autor agradece al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo bajo el número de beca 634421. Los autores agradecen al Prof. Michael Schorr Wiener y el Dr. Marcos A. Coronado Ortega por su valioso aporte y sugerencias.

### **Referencias**

- [1] TFPM – IEEE Task Force on Process Mining, “Process Mining Manifesto,” Business Process Management Workshops, BPM 2011, Lecture Notes in Business Information Processing, vol. 99, pp. 169-194, 2012, doi: 10.1007/978-3-642-28108-2\_19.
- [2] H. S. Aguirre and N. Rincón, “Minería de procesos: desarrollo, aplicaciones y factores críticos,” Cuadernos de Administración, vol. 28, no. 50, 2015, doi: 10.11144/Javeriana.cao28-50.mpsda.
- [3] W. M. P. van der Aalst, “Process Mining: Overview and Opportunities,” ACM Transactions on Management Information Systems, vol 3(2), article no. 7, 2012, doi: 10.1145/2229156.2229157.
- [4] D. Pérez-Alfonso, R. Yzquierdo-Herrera, R. Silverio-Castro and M. Lazo-Cortés, “Utilización de técnicas de minería de proceso en el entorno empresarial cubano,” VII Conferencia Internacional de Ciencias Empresariales, Nov. 2012.  
Available: <https://www.researchgate.net/publication/237843306>
- [5] S. Aguirre-Mayorga and C. Parra-Rodriguez, “Metodología para la aplicación de minería de procesos,” Jornadas Chilenas de Computación, 2014.
- [6] A. Valle, E. Santos and E. Loures, “Applying process mining techniques in software process appraisals,” Information and Software Technology, vol. 87, pp. 19-31, July 2017, doi: 10.1016/j.infsof.2017.01.004.
- [7] A. M. Lemos, C. C. Sabino, R. M. F. Lima and C. A. L. Oliveira, “Using Process Mining in Software Development Process Management: A Case Study,” Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on, pp. 1181-1186, Nov. 2011, doi: 10.1109/ICSMC.2011.6083858.
- [8] V. Rubin, I. Lomazova and W. M. P. van der Aalst, “Agile Development with Software Process Mining,” 2014 International Conference on Software and System Process (ICSSP 2014), pp. 70-74, May 2014, doi: 10.1145/2600821.2600842.
- [9] A. M. Valle, E. A. P. Santos and E. R. Loures, “Applying process mining techniques in software process appraisals,” Information and Software Technology, vol. 87, pp. 19-31, July 2017, doi: 10.1016/j.infsof.2017.01.004.

- [10] M. A. Astorga-Vargas, J. Morales-Bustamante, B. L. Flores-Rios and J. E. Ibarra-Esquer, "Determining software process capability in conformity to the process assessment model NMX-I - 15504 applied to the reference model NMX-I-059 supported by the AURAP tool," *Information Systems and Technologies (CISTI)*, 2014, doi: 10.1109/CISTI.2014.6876898.
- [11] P. E. Velazquez-Solis, B. L. Flores-Rios, M. A. Astorga-Vargas, J. E. Ibarra-Esquer, F. F. Gonzalez-Navarro and C. Hernández-Castro, "Process Mining in Software Process Improvement, a Systematic Literature Review," *IEEE 11th Colombian Computing Conference (IEEE 11CCC 2016)*, Sept. 2016. Available: <https://www.researchgate.net/publication/309398450>
- [12] B. L. Flores-Rios, M. A. Astorga-Vargas, O. M. Rodriguez-Elias, J. E. Ibarra-Esquer and M. Andrade, "Interpretación de las normas mexicanas para la implantación de procesos de software y evaluación de la capacidad bajo un enfoque de gestión de conocimiento," *Revista Facultad de Ingeniería Universidad de Antioquia*, no. 71, pp. 85-100, June 2014. Available: <http://www.redalyc.org/articulo.oa?id=43031204007>
- [13] M. F. Pollo-Cattaneo, R. García-Martínez, P. Britos, P. Pesado and R. Bertone, "Elementos para una ingeniería de explotación de información," *Proyecciones*, vol. 10, no. 1, pp. 67-84, 2012. Available: <http://digital.cic.gba.gob.ar/handle/11746/4019>
- [14] J. A. Vanrell and R. A. Bertone, "Modelo de Proceso de Operación para Proyectos de Explotación de Información," *XVI Congreso Argentino de Ciencias de la Computación*, pp. 674-682, 2010. Available: <http://hdl.handle.net/10915/19328>
- [15] S. Negash and P. Gray, "Business Intelligence," *International Handbooks Information System*, pp. 175-193, 2008, doi: 10.1007/978-3-540-48716-6\_9.
- [16] R. García-Martínez, P. Britos, P. Pesado, R. Bertone, F. Pollo-Cattaneo, D. Rodríguez, P. Pytel and J. Vanrell, "Towards an Information Mining Engineering," *Software Engineering, Methods, Modeling and Teaching*, pp. 83-99, 2011.
- [17] S. Martins, P. Pesado and R. García-Martínez, "Propuesta de Modelo de Procesos para una Ingeniería de Explotación de Información: MoProPEI," *Revista Latinoamericana de Ingeniería de Software*, vol. 2, no. 5, pp 313-332, 2014, doi: 10.18294/relais.2014.313-332.
- [18] A. Orellana-García, O. U. Larrea-Armenteros and D. Perez-Alfonso, "Generador de Registros de Eventos para el análisis de procesos en el Sistema de Información Hospitalaria xavia HIS," *Convención Salud 2015*, 2015.
- [19] W.M.P. van der Aalst, "Extracting Event Data from Databases to Unleash Process Mining," *Management for Professionals, BPM - Driving Innovation in a Digital World*, pp 105-128, 2015, doi: 10.1007/978-3-319-14430-6\_8.
- [20] W. Poncin, A. Serebrenik and M. van den Brand, "Process mining software repositories," *15th European Conference on Software Maintenance and Reengineering*, pp. 5-13, 2011, doi: 10.1109/CSMR.2011.5.

- [21] D. Pyle, "Business Modeling and Business intelligence," Morgan Kaufmann Publishers, 2003.
- [22] J. Wang, R. K. Wong, J. Ding, Q. Guo and L. Wen, "Efficient Selection of Process Mining Algorithms," IEEE Transactions on Services Computing, vol. 6, no. 4, pp. 484-496, 2013, doi: 10.1109/TSC.2012.20.
- [23] I. M. Ailenei, "Process Mining Tools: A Comparative Analysis," Partial fulfillment of the requirements for the degree of Master of Science in Business Information Systems, Eindhoven University of Technology, 2011.
- [24] S. J. Urrea-Contreras, B. L. Flores-Rios, M. A. Astorga-Vargas, J. E. Ibarra-Esquer and F. F. González-Navarro, "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, pp 151-162, 2018.
- [25] B. Kitchenham, "Guidelines for performing Systematic Literature Reviews in Software Engineering," Version 2.3, EBSE Technical Report EBSE-2007-01, Keele University and University of Durham, 2007.
- [26] P. Britos, E. Fernández, M. A. Ochoa, H. Merlino, E. Diez and R. García-Martínez, "Metodología de selección de herramientas de explotación de datos," XI Congreso Argentino de Ciencias de la Computación, 2005. Available: <http://hdl.handle.net/10915/23070>