

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

Instituto de Ingeniería

Maestría y Doctorado en Ciencias e Ingeniería



**Marco de Trabajo para la Identificación y Selección de
Algoritmos y Herramientas de Minería de Procesos**

Tesis que para obtener el grado de
MAESTRO EN INGENIERÍA

Presenta

Silvia Jaqueline Urrea Contreras

Director de tesis

Dra. Brenda Leticia Flores Ríos

Mexicali, Baja California, mayo de 2019.

RESUMEN de la Tesis de **SILVIA JAQUELINE URREA CONTRERAS**, presentada como requisito parcial, para la obtención del **GRADO DE MAESTRO EN INGENIERÍA**. Mexicali, Baja California, México. Mayo del 2019.

MARCO DE TRABAJO PARA LA IDENTIFICACIÓN Y SELECCIÓN DE ALGORITMOS Y HERRAMIENTAS DE MINERÍA DE PROCESOS

Resumen aprobado por:

Dra. Brenda Leticia Flores Ríos
Director de Tesis

Las organizaciones de software producen evidencia de la ejecución de sus actividades, que pueden procesarse con algoritmos, con el apoyo de herramientas de Minería de Procesos y usarse para descubrir, monitorear y mejorar los procesos. 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, lo que ha resultado en desafíos para su selección e identificación de acuerdo con los algoritmos o técnicas que maneja cada uno de ellos.

Ante este contexto, se presenta el **MARCO DE TRABAJO PARA LA IDENTIFICACIÓN Y SELECCIÓN DE ALGORITMOS Y HERRAMIENTAS DE MINERÍA DE PROCESOS** que contempla un modelo de referencia para el proceso de identificación y selección de algoritmos y herramientas de Minería de Procesos denominado MoProPEI-MP y una Guía Técnica que brinde conocimiento a los roles de software. MoProPEI-MP se presenta como una adaptación de un modelo de proceso para proyectos de explotación de información, orientado a guiar en la identificación y selección de algoritmos y herramientas de Minería de Procesos.

TÉRMINOS CLAVE: Minería de Procesos, Herramientas y Algoritmos de Minería de Procesos, Ingeniería de Software.

ABSTRACT of the Thesis, presented by **SILVIA JAQUELINE URREA CONTRERAS**, as a partial requirement to obtain the **MASTER IN ENGINEERING DEGREE**. Mexicali, Baja California, México. May 2019.

FRAMEWORK FOR THE IDENTIFICATION AND SELECTION OF PROCESS MINING ALGORITHMS AND TOOLS

Abstract approved by:

Dra. Brenda Leticia Flores Ríos
Thesis Advisor

Software organizations produce evidence of the execution of their activities, which can be processed with algorithms, supported by process mining tools, and used to discover, monitor and improve processes. Some software development organizations expose their problems in 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, which has resulted in challenges for their selection and identification according to the algorithms or techniques that each of them handles.

In this context, the **FRAMEWORK FOR THE IDENTIFICATION AND SELECTION OF PROCESS MINING ALGORITHMS AND TOOLS** is presented, that includes a reference model for the process of identification and selection of Process Mining algorithms and tools called MoProPEI-MP and a Technical Guide that provides knowledge to the software roles. MoProPEI-MP is presented as an adaptation of a process model for information exploitation projects, oriented to guide in the identification and selection of algorithms and tools of process mining.

KEYWORDS: Process Mining, Process Mining Tools and Algorithms, Software Engineering.

Dedicatoria

A mi madre Silvia Contreras, Sofía Isabella y Nahir Sánchez por haberme impulsado a creer en mí y enseñarme la motivación de seguir adelante.

Reconocimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado para realizar mis estudios de posgrado a nivel maestría y de esa forma poder desarrollar la tesis presentada.

Al Instituto de Ingeniería por brindarme la oportunidad de ser alumna del programa Maestría y Doctorado en Ciencias e Ingeniería (MyDCI) de la Universidad Autónoma de Baja California (UABC).

A los doctores Dr. Félix Fernando González Navarro, Dr. Jorge Eduardo Ibarra Esquer y la Dra. Larysa Burtseva por su colaboración en los productos académicos logrados.

Al Dr. Gabriel Alejandro López Morteo, la Dra. Araceli Celina Justo López y la M.C. Linda Eugenia Arredondo Acosta por su aceptación de participar en la conclusión de la tesis presentada.

A la M.C. María Angélica Astorga Vargas por el tiempo brindado en la revisión de cada producto académico logrado, así como por sus aportaciones que representaron una continuidad para seguir avanzando.

A mi tutora y directora de tesis, la Doctora Brenda Leticia Flores Ríos por todo lo que me ha hecho aprender y crecer tanto en lo profesional como en lo personal y por la valiosa enseñanza de que toda mujer tiene el derecho de poderse superar en todo ámbito, haciendo frente a los obstáculos y adversidades que se puedan presentar.

Índice

Capítulo 1. Introducción.....	1
1.1 Planteamiento del Problema.....	2
1.2 Justificación.....	5
1.3 Objetivos	6
1.4 Metodología	6
1.5 Estructura del Documento.....	7
Capítulo 2. Revisión de la Literatura	8
2.1 Planificación.....	9
2.1.1 Identificación de la necesidad de la revisión.....	9
2.1.2 Formulación de Preguntas de Investigación.....	10
2.1.3 Cadena de Búsqueda	10
2.1.4 Identificación y Selección de las Fuentes.....	11
2.2 Ejecución.....	11
2.2.1 Criterios de Inclusión y Exclusión	12
2.2.2 Selección de Estudios Primarios	13
2.2.3 Extracción de Información	15
2.3 Análisis de Resultados	17
2.3.1 Categorización de los 17 EP por medio de la Teoría de Conjuntos	18
2.3.2 Principios Teóricos.....	20
2.3.3 Propuestas	21
2.3.4 Herramientas y Técnicas de Minería de Procesos.....	23
2.3.5 Minería de Procesos bajo proyectos de desarrollo de software y SPI.....	29
Capítulo 3. Marco de Trabajo para la identificación y selección de algoritmos y herramientas de Minería de Procesos	32
3.1 Modelo de Referencia para la Minería de Procesos: MoProPEI-MP.....	33
3.1.1 MoProPEI.....	33
3.1.2 Descripción general del Proceso de Modelado en MoProPEI-MP	34

3.1.3	Descripción de Productos de Trabajo.....	35
3.1.4	Descripción de Actividades.....	38
3.2	Implementación y Evaluación del Proceso de Modelado de MoProPEI-MP.....	40
Capítulo 4. Algoritmos y Herramientas para la Minería de Procesos.....		43
4.1	Algoritmos y Técnicas de Minería de Procesos	43
4.1.1	Algoritmo Alpha (α).....	44
4.1.2	Algoritmo Genético.....	45
4.1.3	Algoritmo Heurístico	46
4.1.4	Algoritmo Difuso	46
4.1.5	Análisis comparativo de algoritmos de Minería de Procesos.....	47
4.2	Herramientas de Minería de Procesos.....	49
4.2.1	ProM.....	49
4.2.2	Disco	51
4.2.3	Aris PPM.....	53
4.2.4	Perceptive Process Mining (antes Futura Reflect)	54
4.2.5	CPN Tools.....	56
4.2.6	ProcessAnalyzer.....	57
4.2.7	Interstage Process Discovery	59
4.2.8	Análisis comparativo de herramientas de Minería de Procesos	60
Capítulo 5. Evaluación del Marco de Trabajo		63
5.1	Proceso de Evaluación (validación empírica).....	63
5.1.1	Estructura General de Focus Group	64
5.1.2	Ejecución de la técnica Focus Group para la evaluación del Marco de Trabajo.....	69
5.1.2.1	F1. Planteamiento de la investigación.....	69
5.1.2.2	F2. Definición de grupos de discusión	71
5.1.2.3	F3. Conducción de la sesión de debate.....	73
5.1.2.4	F4. Análisis de información y reporte de resultados	74
5.2	Análisis y Reporte de Resultados Cualitativos, Cuantitativos y Mixtos	74
5.2.1	Análisis Cualitativo	75
5.2.2	Análisis Cuantitativo	78

5.2.3	Análisis Mixto.....	83
5.3	Modificaciones realizadas a partir de la evaluación.....	86
Capítulo 6. Conclusiones y Trabajo Futuro.....		88
6.1	Conclusiones	88
6.2	Logros Académicos.....	90
6.2.1	Productos Académicos	90
6.2.2	Conferencias en Congresos Internacionales.....	90
6.2.3	Ponencia en Seminario de Investigación.....	91
6.2.4	Registro de Modelo de Utilidad	91
6.3	Trabajo Futuro.....	91
Referencias.....		93
Anexo A. Guía Técnica: Conceptos Básicos de la Minería de Procesos.....		97
	Descripción de Términos Básicos	97
	Tipos de Minería de Procesos	98
	Perspectivas de la Minería de Procesos.....	98

Lista de Tablas

Tabla 1. Preguntas de investigación definidas.	10
Tabla 2. Generación de la cadena de búsqueda.....	11
Tabla 3. Documentos extraídos del proceso de obtención de estudios primarios.....	16
Tabla 4. Distribución de estudios primarios por país y universidades.....	17
Tabla 5. Desglose de herramientas y técnicas utilizadas en los estudios primarios.....	24
Tabla 6. Distribución de herramientas y algoritmos por estudio primario.....	26
Tabla 7. Productos de trabajo relacionados al proceso de Modelado en MoProPEI-MP.....	35
Tabla 8. Preguntas relacionadas al cumplimiento de los PT.....	37
Tabla 9. Descripción del proceso de Modelado de MoProPEI-MP en un patrón de procesos....	39
Tabla 10. Trazabilidad de los EP con las tareas del proceso de Modelado de MoProPEI-MP...	40
Tabla 11. Identificación e implementación de herramientas de Minería de Procesos por EP. ...	41
Tabla 12. Comparativo de Algoritmos de MP y la herramienta ProM. Extendido de (Gupta, 2014).	48
Tabla 13. Características de la herramienta ProM v. 6.8.	50
Tabla 14. Características de la herramienta Disco.	52
Tabla 15. Características de la herramienta Aris PPM.....	54
Tabla 16. Características de la herramienta Perceptive Process Mining.....	55
Tabla 17. Características de la herramienta CPN Tools.....	56
Tabla 18. Características de la herramienta ProcessAnalyzer.....	58
Tabla 19. Características de la herramienta Interstage Process Discovery.	59
Tabla 20. Ventajas y Desventajas de herramientas de Minería de Procesos.....	61
Tabla 21. Descripción de los roles de Focus Group.....	64
Tabla 22. Descripción de la Fase Planteamiento de la investigación.....	64
Tabla 23. Descripción de la Fase Definición de grupos de discusión.....	65
Tabla 24. Descripción de la Fase Conducción de la sesión de debate.	65
Tabla 25. Descripción de la Fase Análisis de información y reporte de resultados. Adaptado de (Mendoza-Moreno et al., 2013).	66
Tabla 26. Relación de los elementos para el proceso de aplicación de Focus Group.	67

Tabla 27. Producto Reporte de contexto.	69
Tabla 28. Estructura del Protocolo de Focus Group	70
Tabla 29. Metodología Mixta: Diseño exploratorio secuencial (DEXPLOS).....	71
Tabla 30. Reporte de selección de participantes.	72
Tabla 31. Caracterización de los participantes.....	72
Tabla 32. Reporte de participantes segmentados en grupos.....	73
Tabla 33. Lineamientos de Conformidad del debate cumplido.....	73
Tabla 34. Preguntas de la Entrevista Semiestructurada (Análisis Cualitativo).	76
Tabla 35. Cuestionario de Evaluación de Aspectos Generales (Análisis Cuantitativo).....	79
Tabla 36. Clasificación de las preguntas del cuestionario por categoría.....	80
Tabla 37. Valoración del Marco de Trabajo.	80
Tabla 38. Valoración de MoProPEI-MP.	81
Tabla 39. Valoración de la Guía Técnica.....	81
Tabla 40. Valoración de las herramientas de MP.....	81
Tabla 41. Valoración de los algoritmos de MP.....	81
Tabla 42. Trazabilidad de los instrumentos cualitativos y cuantitativos para el análisis mixto..	83
Tabla 43. Modificaciones y Trabajo Futuro con base a los resultados de la evaluación.	86
Tabla 44. Descripción de Términos Básicos de la Minería de Procesos.....	97
Tabla 45. Tipos de Minería de Procesos.	98
Tabla 46. Perspectivas de la Minería de Procesos.	99

Lista de Figuras

Figura 1. Proceso de Revisión Sistemática. Traducida de (Biolchini et al., 2005)	8
Figura 2. Proceso para la obtención de estudios primarios.	13
Figura 3. Ejecución del proceso para la obtención de estudios primarios.....	15
Figura 4. Representación de las categorías por conjuntos de los estudios primarios.....	19
Tomando en cuenta las operaciones entre conjuntos se observa la relación entre los conjuntos de la Figura 4. Se detecta que los conjuntos A y B muestran una intersección con cardinalidad de 5 EP, los conjuntos A y C tienen una intersección de 6 estudios y los conjuntos B y C poseen una intersección de 5 estudios. Las intersecciones de los conjuntos se representan en la Figura 5.	
Figura 5. Intersecciones entre los conjuntos.	19
Figura 6. Los tres tipos básicos de minería de procesos explicados en términos de entradas y salidas: (a) descubrimiento, (b) verificación de conformidad, y (c) mejoramiento (Van der Aalst et al., 2012).....	21
Figura 7. Distribución de frecuencias por tipo de herramientas.....	27
Figura 8. Distribución de porcentajes por tipo de herramientas.....	28
Figura 9. Esquema de los componentes del Marco de Trabajo para la Minería de Procesos.....	32
Figura 10. Estructura General MoProPEI (Martins et al., 2014).	33
Figura 11. Flujo de trabajo de las tareas relacionadas a las actividades A1 y A2 para el proceso de Modelado de MoProPEI-MP.....	38
Figura 12. Interfaz de la herramienta ProM 6.	51
Figura 13. Interfaz de la herramienta Disco.....	53
Figura 14. Interfaz de la herramienta Aris PPM.	54
Figura 15. Interfaz de la herramienta Perceptive Process Mining (Van der Aalst, 2015).....	56
Figura 16. Interfaz de la herramienta CPN Tools.	57
Figura 17. Interfaz de la herramienta ProcessAnalyzer.	59
Figura 18. Interfaz de la herramienta Interstage Process Discovery.....	60
Figura 19. Diagrama de actividad para el proceso general de la aplicación de Focus Group. (Mendoza-Moreno et al., 2013).	66
Figura 20. Diagrama de actividad para el proceso de la aplicación de Focus Group. Adaptado de (Mendoza-Moreno et al., 2013).	68

Figura 21. Resultados de las categorías en porcentaje.	82
Figura 22. Esquema de los componentes del Marco de Trabajo para la Minería de Procesos (Modificado).	87
Figura 23. Flujo de trabajo de las tareas relacionadas a las actividades A1 y A2 para el proceso de Modelado de MoProPEI-MP (Modificado).....	87

Lista de Abreviaturas

MP	Minería de Procesos
SPI	Software Process Improvement (Mejora de Procesos de Software)
IS	Ingeniería de Software
EP	Estudio Primario
PT	Producto de Trabajo

Capítulo 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 (TFPM – IEEE, 2012; Aguirre y Rincón, 2015). 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 (Van der Aalst, 2012; Pérez-Alfonso et al., 2012). 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 (Aguirre-Mayorga y Parra-Rodríguez, 2014; Valle et al., 2017). Un caso particular, es la aplicación de la MP en proyectos de desarrollo de software (Lemos et al., 2011; Rubín et al., 2014) o mejora de procesos de software (Software Process Improvement - SPI) (TFPM – IEEE, 2012; Aguirre-Mayorga y Parra-Rodríguez, 2014; Valle et al., 2017).

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 (Astorga-Vargas et al., 2014; Velazquez-Solis et al., 2016).

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 (Pérez-Alfonso et al., 2012) 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 documento, 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 técnica que presente conocimiento tópicamente relacionado a los algoritmos y técnicas de MP. Este marco de trabajo 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.

1.1 Planteamiento del Problema

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 innovación y desarrollo (I+D), investigación 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 (Flores-Rios et al., 2014) 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 (Pollo-Cattaneo et al., 2012).

Según (Vanrell y Bertone, 2010), 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 (Negash y Gray, 2008). 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 (Aguirre-Mayorga y Parra-Rodríguez, 2014). Asimismo, especificar metodologías para controlar y gestionar las tareas en la búsqueda de patrones de conocimiento dentro de un gran volumen de información (García-Martínez et al., 2011). 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 (Martins et al., 2014).

Las empresas de software dedican mucho esfuerzo a modelar y aplicar procesos de software para mejorar la calidad de sus productos. Sin embargo, se podrían afrontar desafíos tales como fallas durante el modelado del proceso y desviaciones al aplicar los procesos definidos que pueden incurrir en un costo significativo (Van der Aalst, 2015), por tal motivo, actualmente se están adoptando nuevas herramientas y metodologías para hacer frente a la creciente complejidad de los datos que se generan de los proyectos y las demandas de calidad de los usuarios finales del producto de software (Rubin et al., 2014).

Por otro lado, en (Orellana-García, 2015) 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 (Van der Aalst, 2012), la cual es una disciplina emergente que está siendo incluida, tanto por investigadores como practicantes, en temas de Ingeniería de Software (IS) (TFPM – IEEE, 2012).

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 (Lemos et al., 2011; Rubín et al., 2014), el modelo de referencia de procesos o un ciclo de SPI (TFPM – IEEE, 2012; Aguirre-Mayorga y Parra-Rodríguez, 2014; Valle et al., 2017). Es decir, los modelos de proceso descubiertos pueden ser utilizados para generar ideas de mejora de procesos, observar el proceso real e identificar problemáticas (TFPM – IEEE, 2012). 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 (Van der Aalst, 2015). 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 (timestamp) (Van der Aalst, 2015; Poncin et al., 2011).

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 (Martins et al., 2014; Pyle, 2003). 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, se han desafiado 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 (Wang et al., 2013) disponibles de manera comercial o de uso libre. Algunos ejemplos son Disco desarrollada por Fyxicon, Aris Process Performance Manager (PPM) por Software AG, Futura Reflect por Futura Process Intelligence y ProcessAnalyzer por QPR y PROM; que al mismo tiempo representan una dificultad de comparación entre ellas debido a que muestran diferencias en funcionalidad y desempeño (TFPM – IEEE, 2012). 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 un usuario de MP al decidir cuál es la herramienta más adecuada para analizar una actividad (Ailenei, 2011), si requiere utilizar más de un algoritmo o técnica de MP según el tipo de análisis o contar con el conocimiento técnico requerido dependiendo el tipo de formato que se utilice.

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 (Urrea-Contreras et al., 2018a).

1.2 Justificación

La minería de datos está relacionada con la tecnología (algoritmos) y la exploración de información, con los procesos y las metodologías propias de la ingeniería estando más cerca de los procesos de la IS (Pollo-Cattaneo et al., 2012). Esta es utilizada para encontrar patrones en grandes volúmenes de datos, en una relación de patrones y datos; mientras que la MP encuentra las relaciones de los procesos de la organización dentro de los datos, explorando la información de los registros de eventos (Velazquez-Solis et al., 2016). Antes del 2011, esta disciplina se refería a un grupo de técnicas con el objetivo de descubrir, mejorar y monitorear procesos reales a través de la extracción del conocimiento (Nakatumba, 2010).

Existen herramientas o sistemas de información que cuentan con la capacidad de procesamiento y manejo de información, automatización de actividades, asignación de tareas, definir flujos y rastrear su progreso (Valenzuela Ruiz y Flores Rios, 2008). A partir de las iniciativas de estandarización de lenguajes de procesos como XPDL, BPML, BPEL4WS y el modelo de referencia de flujo de trabajo, definido por WfMC (Workflow Management Coalition) se expone la necesidad de utilizar herramientas que hagan uso del modelado de procesos de negocio o diagramas tipo Business Process Diagram (BPD) bajo el estándar BPMN.

De igual manera, el contar con herramientas o sistemas de flujos de trabajo para el seguimiento de las actividades definidas en un modelo de referencia de procesos brindaría una ventaja adicional al análisis de hallazgos en las evaluaciones. Cuando una organización de desarrollo de software, gestiona un proyecto de software o uno de mejora de procesos genera volumen de información, entonces se recomienda la implementación de técnicas o algoritmos de Inteligencia de Negocio, Inteligencia Artificial o MP que permitan explorar registros de eventos (Yzquierdo et al., 2012).

Bajo la perspectiva de MP, los flujos de trabajo están incluidos implícitamente en cada registro de eventos asociado a cada una de las actividades de los procesos, permitiendo realizar el descubrimiento, mejora y monitoreo de proceso a través de la aplicación de herramientas de software para el análisis de datos conociendo el estado actual de las actividades, por ende de los procesos, y/o diagnosticar el nivel de capacidad de los mismos (Velazquez-Solis et al., 2016). El desafío es aprovechar los datos de eventos en una forma significativa, por ejemplo, para proveer un mejor entendimiento, identificar cuellos de botella, anticipar problemas, registrar violaciones de políticas, recomendar contramedidas, y simplificar procesos.

1.3 Objetivos

Para este proyecto de investigación, se ha definido el siguiente objetivo general: Definir un Marco de Trabajo para la Identificación y Selección de Algoritmos y Herramientas de MP, como apoyo a proyectos de desarrollo de software y de SPI. A su vez, se cuentan con cuatro objetivos específicos:

- O1. Diseñar un modelo de referencia para la identificación y selección de algoritmos y herramientas 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 herramientas de MP.
- O3. Identificar los factores/criterios técnicos necesarios para hacer uso del modelo de referencia a través de guía técnica.
- O4. Validar los resultados con la técnica Focus Group enfocada a la IS.

1.4 Metodología

Para el cumplimiento del objetivo general presentado en la sección anterior, se seguirá una metodología con enfoque mixto propuesta por cuatro etapas.

La primera etapa trata de una revisión sistemática de literatura siguiendo el protocolo de (Kitchenham, 2007), 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 los factores/criterios técnicos que las organizaciones podrán considerar para hacer uso eficiente del modelo de referencia propuesto a través de una guía técnica y de esa forma guiar, fácil y técnicamente, la identificación de la(s) herramientas de MP y seleccionada(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 las bases de conocimiento en organizaciones de México y Colombia.

1.5 Estructura del Documento

El Capítulo 2 muestra la revisión sistemática de literatura enfocada a los temas de identificación y selección de herramientas de MP.

En el Capítulo 3 se presenta el Marco de Trabajo para la Identificación y Selección de Algoritmos y Herramientas de Minería de Procesos describiendo sus componentes y enfocándose en su modelo de referencia MoProPEI-MP, explicando su descripción general dentro de un patrón de procesos, los productos de trabajo que se generan de la ejecución de las tareas y las actividades que posee MoProPEI-MP.

El Capítulo 4 aborda parte del contenido de la Guía Técnica: la descripción de los algoritmos y herramientas de MP. Dentro de los algoritmos se presenta el algoritmo alpha (α), genético, heurístico y difuso, por el lado de las herramientas se exhibe ProM, Disco, Aris PPM, Pereptive Process Mining, CPN Tools, ProcessAnalyzer e Interstage Process. Al final de la descripción de cada sección (Algoritmos y Herramientas) se expone un análisis comparativo de información.

En el Capítulo 5 se desarrolla la evaluación del Marco de Trabajo a través de la técnica de Focus Group dentro de la cual, se implementó una metodología mixta para poder derivar datos cualitativos, generados por una entrevista semiestructurada; datos cuantitativos, obtenidos de la aplicación de un cuestionario y datos mixtos resultantes de la correlación entre los tipos de datos mencionados anteriormente.

El Capítulo 6 demuestra las conclusiones y trabajo futuro así como los productos académicos logrados. También se mencionan las conferencias en congresos internaciones, ponencia en Seminario de Investigación, el registro de un modelo de utilidad y las aportaciones que se alcanzaron a cumplir en el tiempo de la investigación.

El Anexo A presenta el contenido faltante de la Guía Técnica: los conceptos básicos de la MP donde se puntualizan términos básicos de la disciplina, los tipos de MP (descubrimiento de procesos, verificación de la conformidad y la mejora de procesos), así mismo las perspectivas de la MP (flujo de control, organizacional, del caso y del tiempo).

Capítulo 2. Revisión de la Literatura

La revisión de la literatura se considera como un paso inicial en cualquier organización de investigación y desarrollo. Es el medio por el cual el investigador puede realizar un mapeo de los conocimientos existentes y previamente desarrollados e iniciativas en el campo. Debido a su importante papel en la área científica, se han desarrollado reglas generales para la realización de reseñas bibliográficas, a fin de garantizar al investigador una buena calidad de información del material cubierto (Biolchini et al., 2005).

En este capítulo, se implementa el método de Revisión Sistemática de la Literatura (Systematic Literature Review - SLR) con el fin de presentar el estado del arte relacionado a la selección y adopción de herramientas de MP que cuenten con la capacidad de análisis de técnicas y/o algoritmos que permitan dentro de las organizaciones en su mayoría PyMEs o entidades de desarrollo de software la explotación y exploración de los registros de eventos en sus repositorios así como su papel en proyectos de desarrollo de software y de SPI. Una SLR es un método para identificar, evaluar e interpretar toda la investigación disponible pertinente a una pregunta de investigación en particular, área temática o fenómeno de interés. Los estudios que contribuyen a una revisión sistemática se denominan Estudios Primarios (EP) (Kitchenham, 2007).

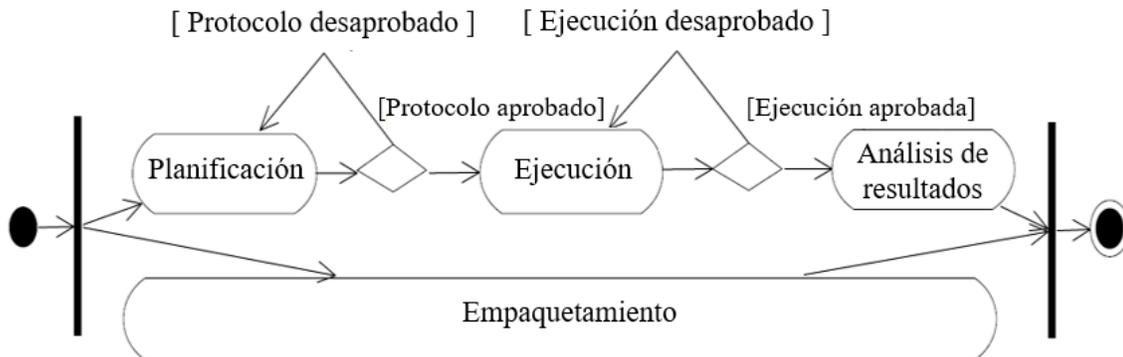


Figura 1. Proceso de Revisión Sistemática. Traducida de (Biolchini et al., 2005)

Como se puede observar en la Figura 1, el proceso de SLR consta de tres fases principales (Biolchini et al., 2005): 1) Planificación, 2) Ejecución y 3) Análisis de resultados. Cada fase cuenta con actividades propias que se detallan en las siguientes secciones. Al final, se obtiene como resultado una conclusión general sobre los estudios primarios del tema de interés. A

continuación, se detallan las actividades de cada una de las tres fases con su procedimiento de realización.

2.1 Planificación

En esta fase se confirma la identificación de la necesidad de la revisión, la especificación de la pregunta central de la investigación y la selección de fuentes primarias que se usarán para ejecutar la revisión; así como las cadenas de búsqueda a implementar dentro de las fuentes seleccionadas. Se analizarán los trabajos con mayor similitud al objetivo de la presente investigación. Se tomará como base los temas en torno a la selección y adopción de herramientas, algoritmos y técnicas de MP, su participación y apoyo a proyectos de desarrollo de software y de SPI así como la localización de las herramientas de MP con mayor identificación y selección por parte de los autores y el proceso de selección de dichas herramientas, algoritmos y técnicas.

2.1.1 Identificación de la necesidad de la revisión

En la actualidad, la creciente acumulación de información en repositorios y/o bases de datos de las organizaciones desarrolladoras de software ha conllevado una exploración y explotación de los registros de eventos que se encuentran a disposición. Esta situación resulta beneficiosa para las organizaciones y representa una oportunidad para poder realizar una mejora de sus procesos y/o gestión de sus proyectos con información que se encuentra disponible y se genera día tras día. Sin embargo, la facilidad de acceso a dicha información no simboliza una interpretación directa de resultados, la información debe ser procesada y analizada por un experto del área.

Obtener resultados de los registros de eventos por parte de las herramientas, algoritmos y técnicas de MP no es la parte difícil del proceso de minado, la interpretación de los mismos es el desafío que se enfrenta para discernir entre la información irrelevante y la información esencial; por consiguiente es necesario conocer el proceso de selección y adopción de herramientas para el modelado, análisis, seguimiento de los registros de eventos así como las herramientas, algoritmos y técnicas propias de la MP para poder fijar un uso asertivo de la MP y obtener resultados relevantes.

La estructuración de lo antes mencionado muestra un desafío en la revisión de literatura por la descentralización de la información y la profundidad de alcance que representan los cuestionamientos y objetivos de la investigación. Por lo tanto, la SLR se realiza con el objetivo de conocer el estado actual de la selección y adopción de herramientas, con capacidad de analizar

los registros de eventos almacenados en los repositorios de las empresas de software y que ayuden a la toma de decisiones y acciones en función de elevar su nivel de madurez.

2.1.2 Formulación de Preguntas de Investigación

La pregunta de investigación (Research Question - RQ) determina el punto de interés de la SLR con el fin de obtener un contexto de investigación claro y contar con una identificación de resultados posibles al contestarla. Para determinar el punto de partida de la investigación y con el propósito de obtener un contexto en la disciplina de IS y la MP, se establecieron dos RQ con su respectiva motivación las cuales se presentan en la Tabla 1 (Urrea-Contreras et al., 2018b).

Tabla 1. Preguntas de investigación definidas.

Pregunta de investigación	Motivación
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?	Conocer cuales modelos de referencia o evaluación de procesos han contemplado la selección y adopción de herramientas de la MP dentro de un proyecto SPI y la forma en que se manejó las tecnologías de la 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?	Determinar cuáles son los algoritmos y herramientas de MP (nombres y características) que se han utilizado en casos de estudio, proyecto de desarrollo de software o de SPI para conocer la frecuencia de identificación y selección por parte de los autores.

2.1.3 Cadena de Búsqueda

A partir de las RQ, se especificaron las palabras clave para poder realizar búsquedas más acertadas y concretas. La cadena de búsqueda la conforman aquellas palabras principales o términos relacionados para poder contestar las RQ y de ese modo conocer los resultados más relevantes de la búsqueda. La Tabla 2 muestra los términos clave (relacionadas a las áreas de MP, IS y SPI) junto con la cadena de búsqueda utilizada, de igual forma, también se consideró la derivación de términos relacionados para los términos clave para la formulación de la cadena de búsqueda.

Tabla 2. Generación de la cadena de búsqueda.

Términos Clave	Términos relacionados	Cadena de búsqueda
<ul style="list-style-type: none"> • Process mining 		(Process Mining) AND (Software process improvement) AND (Process Mining Software OR Analysis and Adoption of Process Mining tools) AND (Organizations OR Enterprises OR Entities OR SMEs)
<ul style="list-style-type: none"> • Analysis and Adoption of process mining tools 	<ul style="list-style-type: none"> • Process mining software 	
<ul style="list-style-type: none"> • Software process improvement 		
<ul style="list-style-type: none"> • Organizations 	<ul style="list-style-type: none"> • Enterprises • Entities • SMEs 	

2.1.4 Identificación y Selección de las Fuentes

Las búsquedas iniciales de los EP se pueden realizar utilizando bibliotecas digitales (Kitchenham, 2007), por lo que la identificación y selección de las fuentes se centra en la ejecución de obtención de los EP por medio de la cadena de búsqueda y que son consideradas relevantes en el área de IS. Estas fueron: IEEE Xplore, Elsevier Science (ScienceDirect), Springer y ACM Digital Library.

La aplicación de la cadena de búsqueda en las diversas bibliotecas digitales proporcionara un número considerable de estudios, el cual se puede elevar por la cantidad de documentos repetidos entre estas bibliotecas, por ello, es sustancial agregar un criterio que limite la aparición de dichos estudios.

2.2 Ejecución

Dentro de la etapa de ejecución, se obtienen los EP de acuerdo a los criterios de inclusión y exclusión los cuales delimitan el contexto de la búsqueda con base a las características que se deben presentar en la SLR.

Es importante que en esta etapa se cuente con un proceso de selección claro de los estudios potenciales a considerar para su incorporación. Por ello, la definición de criterios de inclusión y exclusión deben ser aplicables para todos los estudios encontrados.

2.2.1 Criterios de Inclusión y Exclusión

Los criterios de inclusión y exclusión deben basarse en las RQ. Tienen que ser probados para garantizar que puedan ser interpretados de manera confiable y que clasifiquen los EP correctamente (Kitchenham, 2007).

Los criterios de inclusión definen las características que debe de cumplir la ejecución de la SLR para poderse considerar un elemento relevante de la búsqueda; mientras los criterios de exclusión especifican las características que no deben de estar presentes en las búsquedas realizadas, de esta manera, dichos criterios ayudan a discernir la información con mayor enfoque hacia el beneficio de la investigación.

Definición de criterios de inclusión:

1. Estudios en los idiomas inglés o español.
2. Estudios a partir del año 2007, un periodo de diez años.
3. Estudios que contengan al menos dos palabras clave en el título y abstract.
4. Estudios que contengan análisis y adopción de herramientas dentro del ámbito de MP.
5. Estudios que contengan implementación de SPI.
6. Estudios que muestren resultados sobre el registro de eventos manipulados por herramientas o software específico para la MP.

Definición de criterios de exclusión:

- 1 Estudios que no contengan información sobre registros de eventos relacionados a procesos de software o herramientas de MP.
- 2 Estudios repetidos (duplicados).
- 3 Estudios a los que no se tiene acceso.
- 4 Estudios que no cumplan con al menos 4 criterios de inclusión.

2.2.2 Selección de Estudios Primarios

Una vez que los criterios de inclusión y exclusión son apropiados para la necesidad de la revisión, la cadena de búsqueda se especifica de acuerdo a los términos claves, las fuentes donde será implementada la cadena de búsqueda quedan determinadas y las RQ quedan específicas para delimitar el espectro de la búsqueda, se procede a continuar con la etapa de selección de EP.

La selección de EP conlleva un proceso a seguir para extraer aquellos documentos que son más relevantes para la investigación a partir de estudios filtrados por medio de los criterios de inclusión y exclusión y otorgando un número finito de EP obtenidos. Se deben escoger solo los documentos que realmente supongan una aportación determinante y resalten aquellos consultados que reflejen una verdadera contestación hacia las RQ.

Esta etapa representa reducir la información disponible, incrementar la validez de los estudios seleccionados e identificar la relevancia de publicación que se tiene por parte de los términos claves establecidos en la sección 2.1.3.

Para realizar la búsqueda exhaustiva de los documentos potencialmente relevantes se llevó a cabo un proceso de selección, el cual consiste en 7 pasos divididos en 5 etapas que se encuentran claramente descritos en la Figura 2 por medio de un diagrama de flujo. La aplicación de dicho proceso de obtención de EP en las diversas fuentes proporcionará un número finito de documentos notables. Como estrategia complementaria se tomó en cuenta la lectura de la bibliografía de los documentos con mayor número de criterios de inclusión cumplidos.

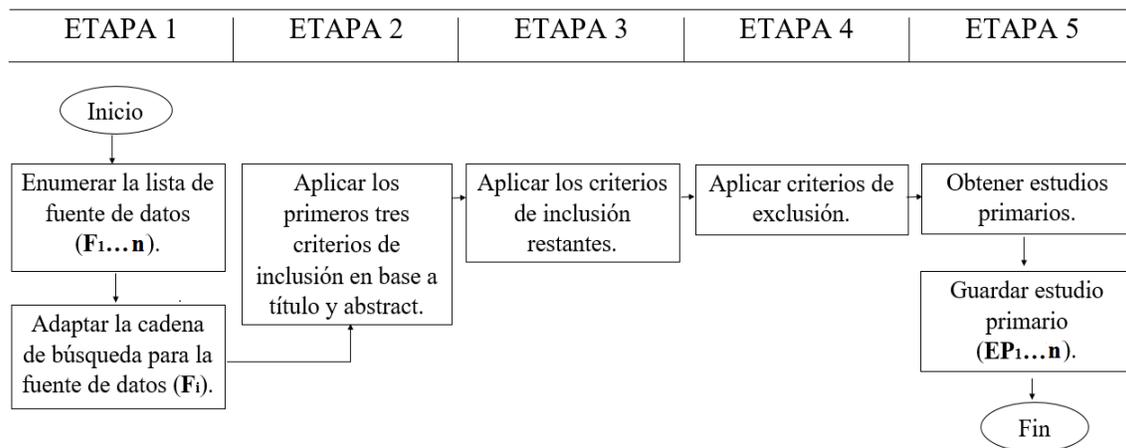


Figura 2. Proceso para la obtención de estudios primarios.

A continuación se describen a mayor detalle las etapas del proceso de obtención de EP con sus respectivos pasos a seguir:

Etapas 1

1. Enumerar las fuentes seleccionadas para supervisar y comparar los resultados obtenidos de las diversas fuentes de datos durante el proceso de búsqueda.
2. Adaptar la cadena de búsqueda dependiendo de la fuente de datos con el fin de lograr una búsqueda acertada.

Etapas 2

3. Aplicar los 3 primeros criterios de inclusión con base a los estudios en los idiomas inglés o español realizados a partir del año 2007 y que contengan al menos dos palabras de la cadena de búsqueda en el título y abstract.

Etapas 3

4. Aplicar los criterios de inclusión restantes para filtrar un documento con relevancia importante en el tema de investigación.

Etapas 4

5. Aplicar criterios de exclusión a fin de eliminar aquellos documentos no relevantes en información, repetidos en dos o más fuentes de datos o que no se tenga acceso a él.

Etapas 5

6. Obtener EP depurados conforme a lo establecido en los criterios de inclusión y exclusión.
7. Guardar cada estudio primario con una numeración de identificación para un manejo posterior en la extracción de datos.

En la Figura 3, se muestra la ejecución del proceso para la obtención de EP y la cantidad de estudios obtenidos en cada etapa durante su ejecución. Se observa que de los 23,127 estudios encontrados al ejecutar la cadena de búsqueda, sólo 17 cumplieron con los criterios de inclusión y exclusión, los cuales fueron analizados y utilizados para la presente investigación.

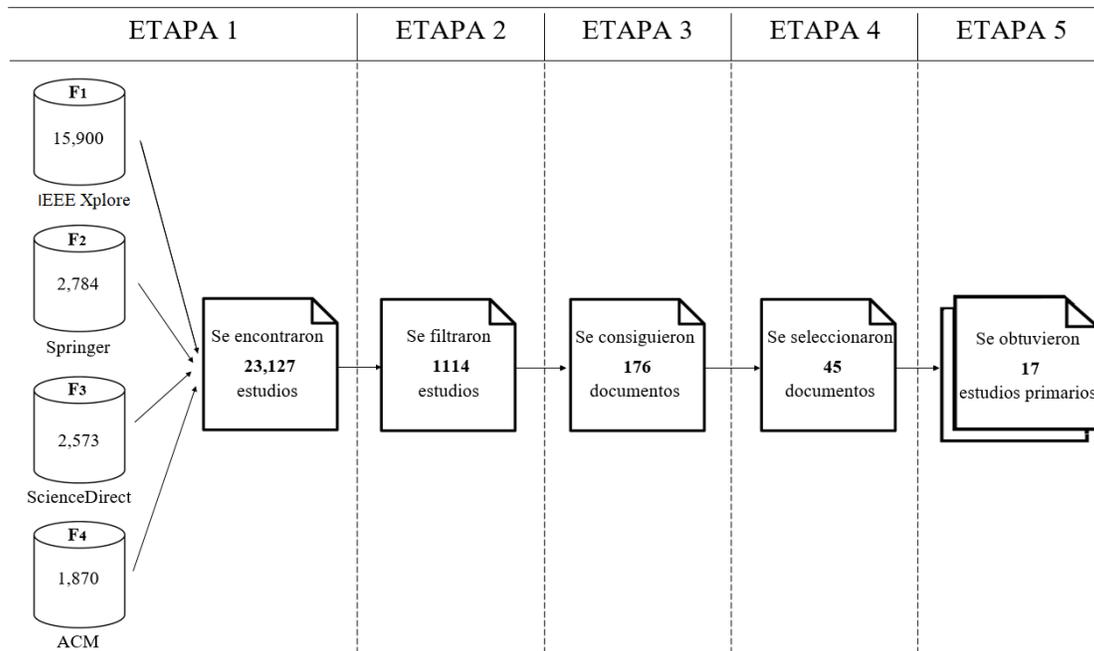


Figura 3. Ejecución del proceso para la obtención de estudios primarios.

La biblioteca con mayor número de artículos respecto al tema de MP fue IEEE Xplore, al momento de la ejecución de la etapa 2 del proceso se obtuvo el 4.81% del total de estudios encontrados en la etapa 1, para finalizar, la etapa 5 concluyó con una muestra representativa del 37.77% a comparación con los 45 estudios depurados de la etapa 4.

2.2.3 Extracción de Información

La extracción de información requiere una máxima fiabilidad de los datos que se extrae de cada EP, dicha información debe permitir la contestación de las RQ. De forma resumida, los datos obtenidos deben dar reiteración a la selección y adopción de herramientas de MP en torno a proyectos de desarrollo de software o de SPI.

Para la extracción de información es importante contar con un formato donde se correlacione la información básica de los EP encontrados y enmarcar una fiabilidad del proceso descrito en la subsección anterior. En la presente investigación se utilizó como formato la síntesis de datos cuantitativa de los EP por medio de tablas, los cuales se registraron y analizaron bajo los siguientes indicadores (Tabla 3): ID (numeración de identificación del EP), año (en que fue publicado el documento), autor(es), fuente, título y tipo de documento (paper, tesis). El orden cronológico de EP se aprecia en la misma tabla cubriendo el periodo desde 2011 a 2017.

Tabla 3. Documentos extraídos del proceso de obtención de estudios primarios.

ID	Año	Fuente	Autor(es)	Título	Tipo de documento
EP1	2011	IEEE	Wouter Poncin, Alexander Serebrenik, Mark van den Brand	Process mining software repositories.	Artículo
EP2	2011	tue.nl	Irina-Maria Ailenei	Process Mining Tools: A Comparative Analysis.	Tesis
EP3	2011	IEEE	Artini M. Lemos, et al.	Using Process Mining in Software Development.	Artículo
EP4	2012	Springer	Wil van der Aalst, et al.	Process Mining Manifesto.	Artículo
EP5	2012	ACM	Wil van der Aalst	Process Mining: Overview and Opportunities.	Artículo
EP6	2012	Researchgate	Damián Pérez Alfonso, Raykenler Yzquierdo Herrera, Rogelio Silverio Castro, Manuel Lazo Cortés	Utilización de técnicas de minería de proceso en el entorno empresarial cubano.	Artículo
EP7	2013	Researchgate	Damián Pérez Alfonso, Raykenler Yzquierdo Herrera, Manuel Lazo Cortés	Recommendation of Process Discovery Algorithms: a Classification Problem.	Artículo
EP8	2014	ACM	Vladimir Rubin, Irina Lomazova, Wil van der Aalst	Agile Development with Software Process Mining.	Artículo
EP9	2014	Researchgate	Esmita .P. Gupta	Process Mining A Comparative Study.	Artículo
EP10	2014	jcc2014	Santiago Aguirre Mayorga, Carlos Parra	Metodología para la aplicación de minería de procesos.	Artículo
EP11	2015	Springer	Wil van der Aalst	Extracting Event Data from Databases to Unleash Process Mining.	Artículo
EP12	2015	Redalyc	Hugo Santiago Aguirre Mayorga, Nicolás Rincón García	Minería de procesos: desarrollo, aplicaciones y factores críticos.	Artículo
EP13	2015	Proquest	Swapnali B. Sonawane, Ravi P. Patki	Process Mining by using Event Logs.	Artículo
EP14	2015	Springer	Anna A. Kalenkova, Wil van der Aalst, Irina A. Lomazova, Vladimir A. Rubin	Process mining using BPMN: relating event logs and process models.	Artículo
EP15	2016	ReCIBE	Paola E. Velazquez-Solis, et al.	Empiric Evidence of Process Mining in CMMI-DEV Implementation.	Artículo
EP16	2016	SSRN	Arturo Orellana García, Damián Pérez Alfonso, Vivian Estrada Sentí	Revisión de los principales modelos para aplicar técnicas de Minería de procesos.	Artículo
EP17	2017	ScienceDirect	Arthur M. Valle, Eduardo A.P. Santos, Eduardo R. Loures	Applying process mining techniques in software process appraisals.	Artículo

2.3 Análisis de Resultados

En la sección de análisis de resultados, se presentan de forma resumida los resultados obtenidos a partir de los 17 EP seleccionados en la sección 2.2.3, se discuten posibles limitaciones e implementaciones encontradas para exponer la evidencia y respaldar las acciones a tomar en relación a la necesidad de la revisión. El análisis considera organizar y categorizar la información extraída para responder determinadamente las RQ.

En esta sección, se hace un análisis y representación de los 17 EP con el objetivo de mostrar cómo ha evolucionado la aplicación de herramientas y técnicas de MP en la última década (2007-2017) y su relación con proyectos SPI. En Tabla 4 se muestra la distribución de EP por país y universidades. Los Países Bajos son los que tienen una mayor frecuencia de investigaciones en el área con 5 estudios provenientes de la Universidad Tecnológica de Eindhoven, seguido por Cuba con 3 estudios derivados de la Universidad de las Ciencias Informáticas. En tercer lugar, con 2 estudios realizados en Rusia con investigaciones de la Escuela Superior de Economía de la Universidad Nacional de Investigación, Colombia en la Pontificia Universidad Javeriana, Brasil en Universidad Federal de Pernambuco y la Pontificia Universidad Católica de Paraná e India en la Universidad de Mumbai y Universidad Savitribai Phule Pune. Finalmente, México con sólo 1 estudio por parte de la Universidad Autónoma de Baja California.

Tabla 4. Distribución de estudios primarios por país y universidades.

País	Universidad	ID del estudio primario
Países Bajos	Universidad Tecnológica de Eindhoven	EP1, EP2, EP4, EP5, EP11
Cuba	Universidad de las Ciencias Informáticas	EP6, EP7, EP16
Rusia	Escuela Superior de Economía de la Universidad Nacional de Investigación	EP8, EP14
Colombia	Pontificia Universidad Javeriana	EP10, EP12
Brasil	Universidad Federal de Pernambuco	EP3
	Pontificia Universidad Católica de Paraná	EP17
India	Universidad de Mumbai	EP9
	Universidad Savitribai Phule Pune	EP13
México	Universidad Autónoma de Baja California	EP15

Para generar un análisis y evaluación más objetiva e identificar los datos sustanciales de los resultados con la finalidad de ensamblar la información consultada de cada EP, éstos se categorizaron bajo la teoría de conjuntos ya que permite diferenciar la relación de elementos a partir de un conjunto único y conocer sus diferencias y similitudes. De esta forma, se podrán contestar las RQ con base a la organización y categorización de la información extraída de los conjuntos que serán originados a partir de los 17 EP.

2.3.1 Categorización de los 17 EP por medio de la Teoría de Conjuntos

Se revisaron los 17 EP de donde se observó el potencial de categorizar la información en 3 temas centrales, debido a las siguientes razones: la categoría 1 exhibe los principios teóricos de la MP por parte de diversos autores y los elementos que intervienen en ella para poder brindar una entendimiento fijo de la MP, la categoría 2 produce la contestación de la RQ1 ya que se encuentra enfocada a las propuestas en relación de la MP con la SPI y por último, la categoría 3 origina la respuesta de la RQ2 debido al encuadre de la identificación e implementación entre herramientas y técnicas de MP.

Las tres categorías que surgieron fueron: 1) Principios teóricos, 2) Propuestas, y 3) Herramientas y técnicas. De lo anterior, se deduce que la categoría 1 expresa los fundamentos necesarios para poder llevar a cabo esta investigación; la categoría 2 muestra las propuestas o iniciativas que se han establecido para la utilización de herramientas y técnicas de MP en la SPI; y la categoría 3 representa la variedad de herramientas tanto comerciales como de código abierto que existen.

Las categorías se analizaron por medio de la Teoría de Conjuntos para delimitar cuales elementos (EP) pertenecen a cada una de ellas, los conjuntos están representados en la Figura 7. El conjunto A (categoría 1) consta con una cardinalidad de 11 estudios, siguiendo el conjunto B (categoría 2) con 8 estudios y finalmente, el conjunto C (categoría 3) que posee 11 EP. Los elementos relacionados a los EP que pertenecen a cada conjunto se describen a continuación:

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

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

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

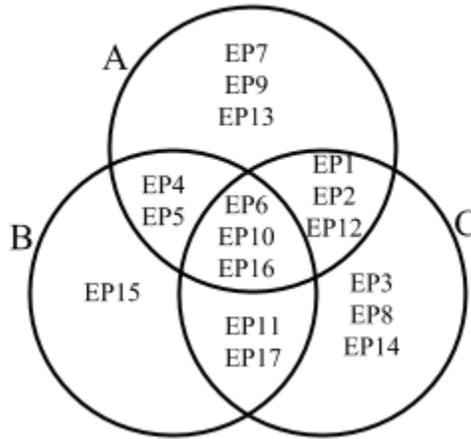


Figura 4. Representación de las categorías por conjuntos de los estudios primarios.

Tomando en cuenta las operaciones entre conjuntos se observa la relación entre los conjuntos de la Figura 4. Se detecta que los conjuntos A y B muestran una intersección con cardinalidad de 5 EP, los conjuntos A y C tienen una intersección de 6 estudios y los conjuntos B y C poseen una intersección de 5 estudios. Las intersecciones de los conjuntos se representan en la Figura 5.

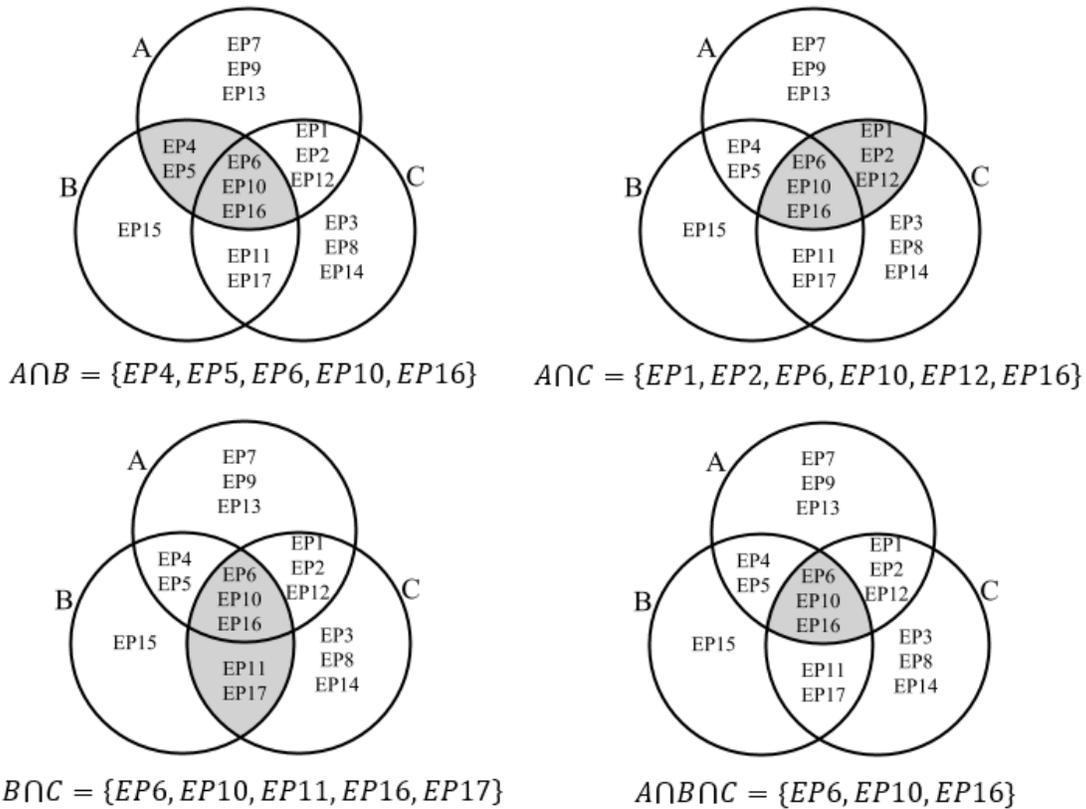


Figura 5. Intersecciones entre los conjuntos.

Los EP de la intersección $A \cap B$ 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 directa 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. 8 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.

Con base a lo anterior, a continuación se desglosa la información obtenida de cada EP en consideración al conjunto que le corresponde. Se ha contemplado relevante que la información se estructure a partir de las categorías identificadas que pasaran a ser subsecciones y de esa forma, establecer un marco de identificación de evidencia y análisis de los resultados logrados.

2.3.2 Principios Teóricos

La MP es una disciplina que permite descubrir, monitorear y mejorar los procesos identificados que poseen las organizaciones a partir de sus registros de eventos (Van der Aalst et al., 2012), está compuesta por una serie de herramientas y técnicas basadas en la minería de datos para analizar procesos cuyo registro de eventos de ejecución real se encuentran disponibles en sistemas de información (Aguirre y Rincón, 2015).

Se evidencia el reciente interés sobre las técnicas y herramientas de MP las cuales ofrecen una medida para 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 (Van der Aalst et al., 2012). Además, dichas herramientas permiten extraer conocimiento de los repositorios, intranets o sistemas de información de las organizaciones para poder llevar una gestión relacionada con la SPI.

La MP enfoca sus esfuerzos en la extracción de información útil y relevante sobre los procesos de la organización, tomando como punto de partida los datos contenidos en los registros de eventos. Cubre tres áreas fundamentales: descubrimiento de procesos, verificación de conformidad y mejora de modelos de procesos (Pérez-Alfonso et al., 2012), como se muestra en la figura 4, dichas áreas se encuentran explicadas en el anexo A.

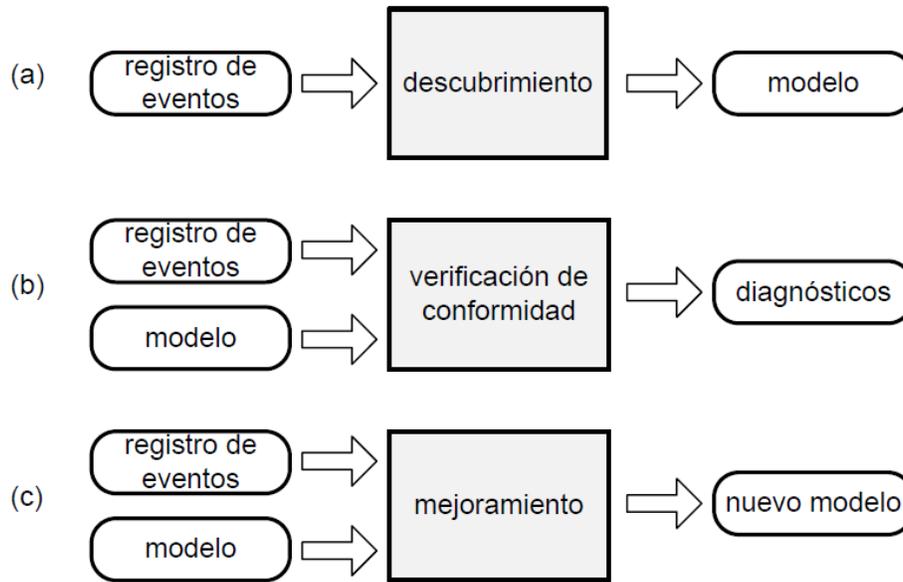


Figura 6. Los tres tipos básicos de minería de procesos explicados en términos de entradas y salidas: (a) descubrimiento, (b) verificación de conformidad, y (c) mejoramiento (Van der Aalst et al., 2012).

Un registro de eventos contiene datos de una serie de procesos (generalmente uno). Cada proceso tiene varias instancias de proceso que pueden ser únicas identificado. Además, cada proceso tiene una cantidad de actividades, y cada instancia de proceso, también conocido como caso ejemplo: una serie de eventos, que consisten en una actividad siendo ejecutado en un momento determinado en el tiempo y asociado con ciertos datos (Poncin et al., 2011). Hoy en día, los registros de eventos ya están automatizados en algunos sistemas de información (van der Aalst, 2011; Orellana García et al., 2015).

2.3.3 Propuestas

Las organizaciones actuales reconocen la ventaja que supone incorporar los nuevos avances tecnológicos a sus procesos. Con el fin de obtener resultados eficientes, constituye una prioridad, mejorar la administración y desempeño organizacional a través de la incorporación de la gestión o enfoque basado en procesos, por tanto se hace necesario romper viejos paradigmas e incorporar nuevos para optimizar los procesos en las organizaciones (Orellana-García et al., 2016). La MP puede considerarse un medio competente para ayudar a las organizaciones a comprender su forma real de trabajar y puede servir como base para la mejora de procesos (Valle et al., 2017).

Las técnicas, algoritmos y aplicaciones de software de MP paulatinamente se han transferido a las empresas a través de casos de estudio y proyectos de MP que han permitido a las organizaciones diagnosticar, analizar y mejorar los procesos basado en el análisis de los datos de ejecución real de los procesos (Aguirre et al., 2014). Es importante descubrir los procesos reales utilizando registro de eventos. Los modelos de proceso descubiertos pueden ser utilizados para generar ideas de mejora de procesos, ver el proceso real y sus problemas (Van der Aalst et al., 2012).

La MP puede ser utilizada para verificar la conformidad y sugerir mejoras en los procesos. Sin embargo, la parte más lenta de la MP no es el análisis real. La mayor parte del tiempo se usa para localizar, seleccionar, convertir y filtrar el registro de eventos (Van der Aalst, 2015).

Desde la perspectiva de la MP, los flujos de trabajo están incluidos implícitamente en cada registro de evento (actividad) asociado a cada uno de los procesos de la organización, permitiendo de esta manera realizar el descubrimiento, la mejora y el monitoreo de los procesos a través de la aplicación de herramientas de software para el análisis de datos (Velazquez-Solis et al., 2016).

Los datos sobre los eventos han comenzado a estar disponibles y las técnicas de MP han madurado. Además, las tendencias de gestión relacionadas al mejoramiento de procesos se pueden beneficiar de la MP (Van der Aalst et al., 2012), con el objetivo de hacer explícita información que puede generar conocimiento para que las organizaciones puedan entender y mejorar sus procesos (Aguirre et al., 2014).

Ante el desafío de aprovechar los registros de eventos para proveer un mejor entendimiento de los procesos y de cómo mejorar su rendimiento la MP al realizar su análisis sobre la evidencia de la ejecución de los procesos brinda numerosas ventajas con respecto a las técnicas tradicionales de análisis y mejora de procesos, ya que ofrece resultados en menor tiempo y con mayor fiabilidad (Pérez-Alfonso et al., 2012).

Analizar los procesos de software desde una perspectiva de sus registros de eventos generados por la ejecución de las actividades por medio de herramientas de trabajo provee nuevas ideas para proponer herramientas de SPI más robustas que permitan un análisis y monitoreo de los registros como tal.

2.3.4 Herramientas y Técnicas de Minería de Procesos

La MP es un tema emergente que ha atraído la atención tanto de los investigadores como de proveedores, esto genera como resultado, varias soluciones de MP disponibles en el mercado. Un problema que se presenta con estas soluciones es la falta de estandarización y el hecho de que las herramientas usan sus propios estándares y nombres. Esto puede ser bastante confuso para un usuario de MP cuando decidir cuál es el sistema o herramienta más adecuado para una tarea (Ailenei, 2011).

Durante la búsqueda de utilización de herramientas y técnicas de MP se visualizó en los estudios categorizados en la categoría 3 una aplicación a diversos casos de estudio/uso de manera exitosa en su mayoría, los estudios clasificados en las categorías 1 y 2 también muestran mención de herramientas pero se limitan a solo presentar los nombres de las herramientas como una referencia de trabajo.

En la Tabla 5 se muestra un desglose de las herramientas y técnicas utilizadas en cada EP ya sea por la aplicación de la herramienta y una breve descripción del motivo de la utilización o por hacer mención de ellas para llevar un registro de las herramientas y técnicas que presenta cada EP encontrando un total de 21 herramientas distintas. Como derivación de esta tabla, se obtiene la Tabla 6 en la cual se representa una distribución de herramientas y algoritmos por EP permitiendo conocer concretamente sin repetir cuales herramientas fueron mencionadas en los EP. La Figura 7 muestra la distribución de frecuencias por tipo de herramientas obtenidas de la Tabla 6 para contestar la RQ2.

La Figura 8 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.

Tabla 5. Desglose de herramientas y técnicas utilizadas en los estudios primarios.

ID	Herramientas y técnicas	Descripción
EP1	FRASR (FRamework for Analyzing Software) + ProM y Futura Reflect.	Se presenta FRASR en función con ProM aplicado en los casos de estudio 1 y 2. Se menciona que la herramienta Futura Reflect podría ser utilizada para este fin también.
EP2	CPN Tools, ProM, ARIS PPM, Flow, Futura Reflect, ProcessAnalyzer, Process Discovery Focus, Comprehend, Intersage Process Discovery y Discovery Analyst.	Se generaron y convirtió registros de eventos artificiales con CPN Tools y ProMimport5. Las herramientas ARIS PPM, Flow, Futura Reflect y ProcessAnalyzer fueron analizadas conforme al marco de evaluación que se propone evaluando las fortalezas y debilidades de las herramientas. La lista inicial de sistemas considerados incluía también a Iontas - Process Discovery Focus, Open Connect – Comprend, Fujitsu - Interstage Automated Process Discovery y StereoLOGIC - Discovery Analyst.
EP3	ProM.	Se utilizó ProM para extraer los registros de eventos de una empresa real y descubrir el modelo de proceso que realmente es empleado por ellos.
EP4	ProM, XESame, y Nitro.	En el 2010 se estandarizó XES (www.xes-standard.org) un formato de registro estándar que es extensible y compatible con la biblioteca OpenXES (www.openxes.org) y con herramientas como ProM, XESame, Nitro, entre otros, orientados a la MP.
EP5	ProM, XESame, y Nitro.	Se actualizó EP4 con énfasis agregando la página http://www.win.tue.nl/ieeetfpm/
EP6	ProM.	Se identificó los casos y eventos en los que se produjo una anomalía con la herramienta ProM.
EP7	ProM, model-log y model-model	Mediante el diseño de un marco de evaluación que permite validar los resultados de la aplicación de MP, se combinan los métodos model-log y model-model para la evaluación de algoritmos.
EP8	ProM, Disco, ARIS, ProcessAnalyzer, Perceptive Process Mining y Celonis.	Se mencionó herramientas exitosas en el área de MP tales como ProM, Disco, ARIS, Process Analyzer, Perceptive Process Mining y Celonis. Se comprobó la funcionalidad de Disco con una breve entrada de registro que contenía una instancia de un proceso, actividad, timestamp y ID del usuario para derivar un modelo visual del comportamiento del usuario analizado y filtrado de modo que se centró en las actividades particulares.

EP9	Algoritmos: Heuristic miner, Genetic miner y Fuzzy miner.	Se evaluó y comparo los algoritmos más utilizados en la MP de manera eficiente.
EP10	Disco.	Se utilizó la herramienta Disco en el caso de estudio para verificar la coherencia de los datos a través del análisis del flujo del proceso con el software de MP.
EP11	ProM, Disco, Perceptive Process Mining, ARIS PPM, Celonis, ProcessAnalyzer, Interstage Process Discovery, Discovery Analyst y XMAalyzer.	Se muestran 4 screenshots de las herramientas ProM, disco, Perceptive Process Mining y Celonis analizando el mismo evento. No se elaboran más técnicas de MP ni se discuten herramientas específicas.
EP12	Disco.	Se planteó alternativas de mejora para los casos de estudios 1, 2 y 3 con el apoyo de la herramienta Disco y los resultados que se obtuvieron.
EP13	ARIS PPM e In Concert.	Se hizo referencia que ARIS e In Concert generan modelos de instancia.
EP14	ProM, Disco, ARIS PPM, Perceptive Process Mining, ProcessAnalyzer y Celonis.	Se implementó ProM considerando la arquitectura de paquetes BPMN y su funcionalidad en ProM.
EP15	ProM, ARIS PPM, Disco, Enterprise Visualization Suite, Interstage BPME, Process Discovery Focus, Reflect one y Futura Reflect.	Se mencionó herramientas comerciales y de código abierto (ProM) de MP.
EP16	XESame, Eventifier, ProM y Disco.	Se explicó el funcionamiento de las herramientas XESame y Eventifier para la obtención de un registro de eventos que puede ser analizado en ProM o Disco para generar un modelo.
EP17	ProM.	Se ejecutó la herramienta ProM en los casos de estudio A y B donde se obtuvo la aplicación de un algoritmo que comparo el modelo de Jure con el modelo de Facto y la verificación de conformidad respecto a algunas reglas comerciales y el registro de eventos.

Tabla 6. Distribución de herramientas y algoritmos por estudio primario.

Herramientas	EP1	EP2	EP3	EP4	EP5	EP6	EP7	EP8	EP9	EP10	EP11	EP12	EP13	EP14	EP15	EP16	EP17
FRASR	x																
ProM	x	x	x	x	x	x	x	x			x			x	x	x	x
Futura Reflect	x	x													x		
CPN Tools		x															
Aris PPM		x						x			x		x	x	x		
Flow		x															
ProcessAnalyzer		x						x						x			
Process Discovery Focus		x													x		
Comprehend		x															
Interstage Process Discovery		x									x				x		
Discovery Analyst		x									x						
XESame				x	x											x	
Nitro				x	x												
Disco								x		x	x	x		x	x	x	
Perceptive Process Mining								x			x			x			
Celonis								x			x			x			
XMAalyzer											x						
In Concert													x				
Enterprise Visualization Suite															x		
Eventifier																x	
Reflect one															x		
Algoritmos: Heuristic miner, Genetic miner y Fuzzy miner.									x								

Distribución de frecuencias por tipo de herramientas

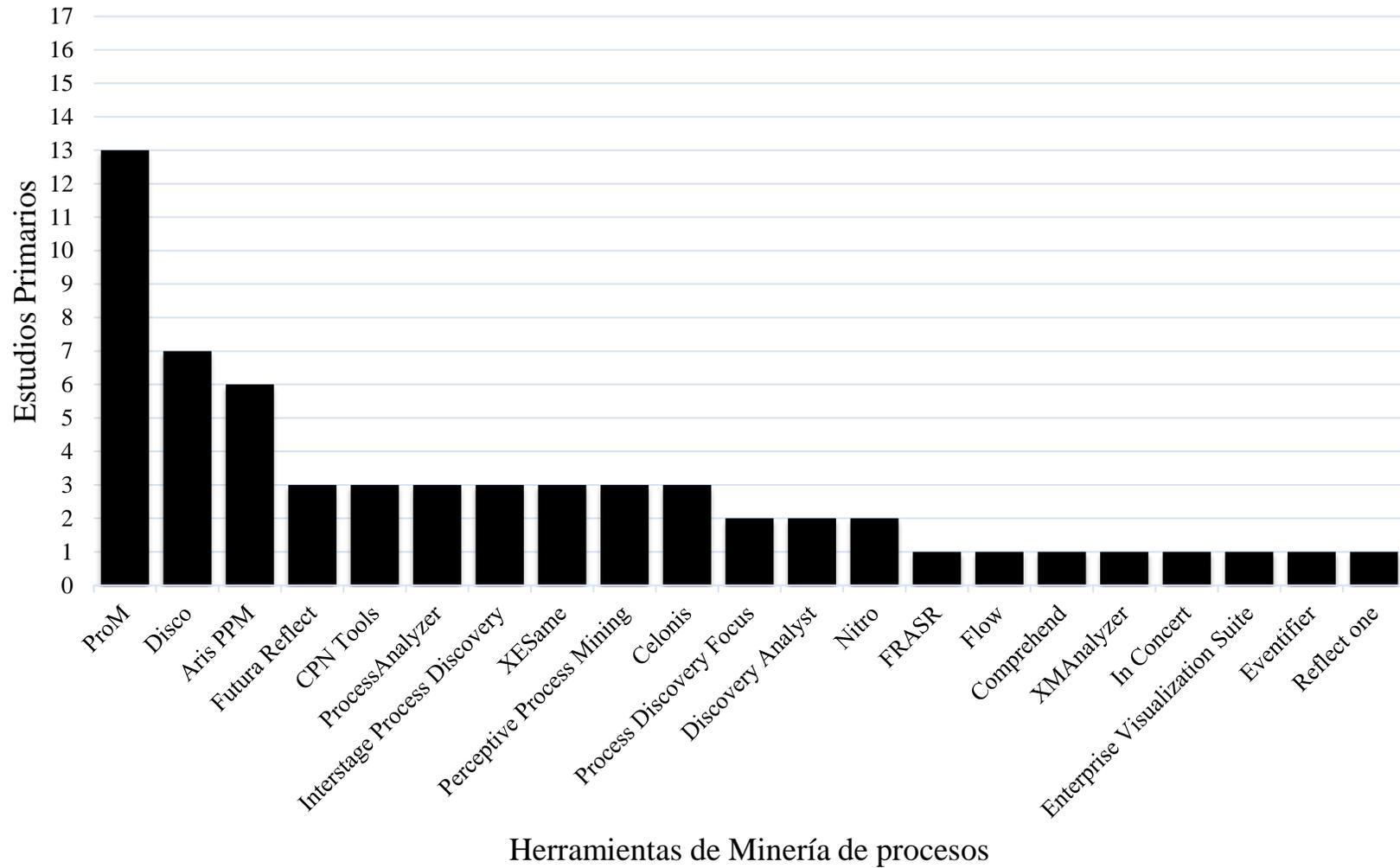


Figura 7. Distribución de frecuencias por tipo de herramientas.

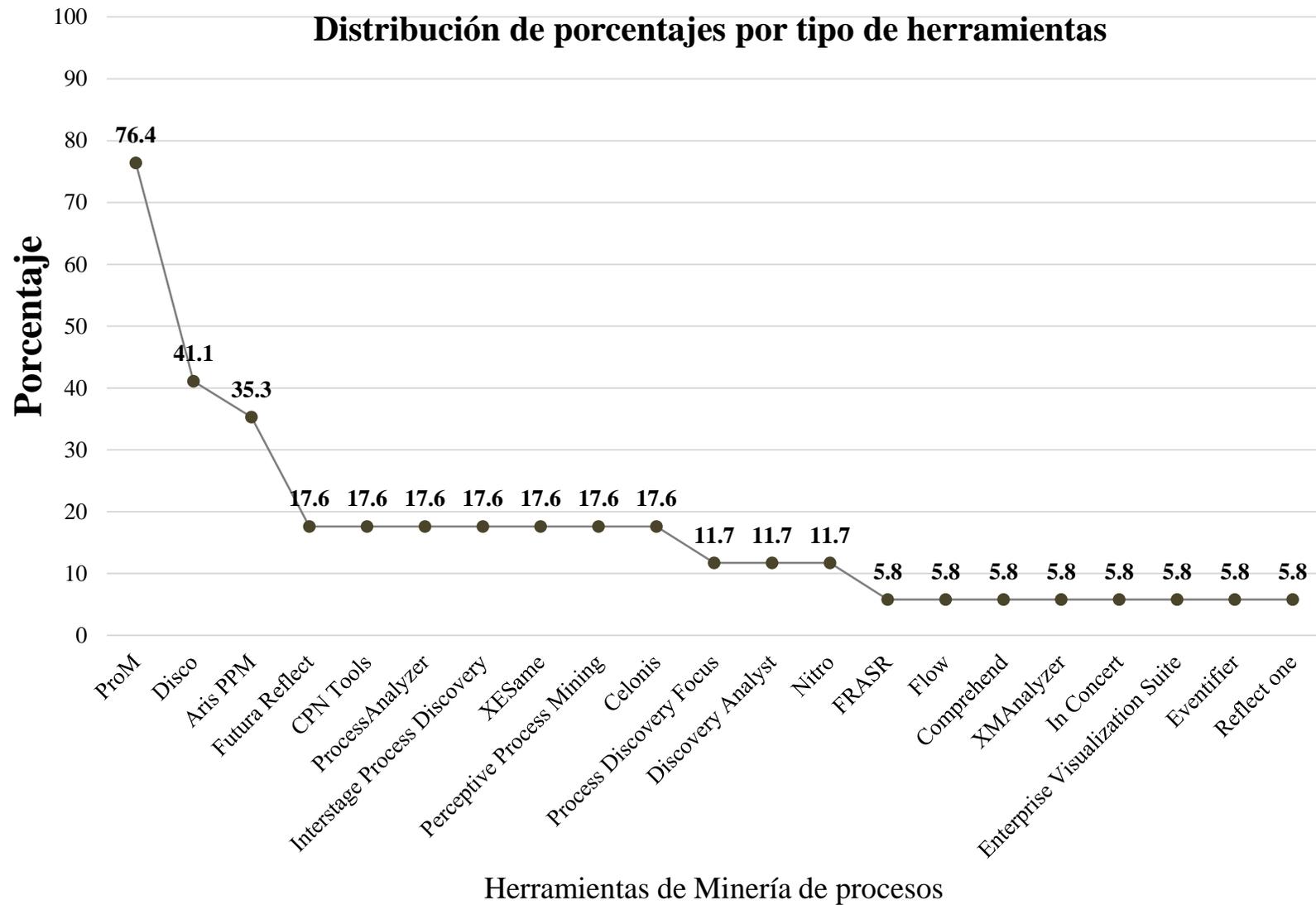


Figura 8. Distribución de porcentajes por tipo de herramientas.

Los EP que evidencian una aplicación de la MP con proyectos de desarrollo de software y SPI fueron aquellos donde se aplicó uno o varios casos de estudio apoyándose de las herramientas de MP., los cuales se describen con mayor detalle en la sección 2.2.4.

2.3.5 Minería de Procesos bajo proyectos de desarrollo de software y SPI

EP1 (Poncin et al, 2011) se plantea el desafío de analizar simultáneamente diferentes repositorios y combinar la información resultante, como solución a lo antes mencionado se propuso la implementación de FRASR (FRamework for Analyzing Software Repositories) junto con técnicas y herramientas de MP. De esta forma, FRASR es el encargado de la combinación y coincidencia de eventos mientras que son analizados en la herramienta ProM. Para evaluar la funcionalidad FRASR+ProM se realizaron dos casos de estudio: 1) Roles de desarrollador y 2) Ciclo de vida de errores, teniendo como resultado 3 roles en el primer caso la asignación de cada uno de los roles disponibles por medio de un registro exportado en combinación con la visualización del Gráfico de puntos ProM; en el caso 2 se utilizó FRASR en combinación con el plugin ProM Fuzzy Miner para tener de resultado un grafo difuso correspondiente al ciclo de vida de los informes de fallas en el repositorio.

Se investigaron 19 casos de uso en EP2, en donde por medio de un marco de evaluación propuesto por Ailenei (2011), se analizaron las ventajas y desventajas de herramientas como ARIS PPM, Flow, Futura Reflect y ProcessAnalyzer, cada caso de uso posee una descripción del caso, el registro de evento usado, los criterios de aceptación y los resultados del procesamiento del registro de eventos en las herramientas mencionadas.

Dentro de EP3 se realizó un caso de estudio en una casa de software brasileña en donde hace cinco años, la compañía definió un proceso de desarrollo de software e implementó un sistema de monitoreo para registrar información relevante sobre su ejecución. Es aquí donde se generó los registros de eventos en los cuales se aplicaron técnicas de MP para extraer información importante sobre el proceso real de desarrollo de software. El objetivo principal fue descubrir el proceso real que ejecutan los equipos de desarrollo y compararlo con el proceso documentado. La base de datos incluída más de 2,000 instancias de proceso y se utilizó la herramienta ProM por su perspectiva del flujo de control al proceso que ofrece para el análisis. Se evaluó el modelo del proceso real y comparó con el proceso que se suponía que debía ejecutarse. Dicho análisis de conformidad demostró como resultados que, aunque algunos proyectos siguen el proceso formal,

la mayoría de ellos no ocurren como se esperaba, también se produjo un conjunto de informes que indican inconsistencias detectadas entre modelo de proceso y su registro de ejecución correspondiente.

En EP6 se enfoca en el Sistema Único de Identificación Nacional (SUIN), el cual antes de poder ejecutar cualquier proceso de negocio sobre él, es necesario transitar por varios procesos de configuración como es el caso del proceso de Gestionar los roles de la solución que posee un registro de eventos (31 casos, 804 eventos, 52 clases de eventos y 3 tipos de eventos) que permiten determinar anomalías en el proceso analizado. A partir del registro de eventos se obtuvo un modelo representativo del proceso aplicando el algoritmo Alpha++ el cual no reveló anomalías por lo que se aplicó una alineación de trazas y de la matriz obtenida se le aplicó el algoritmo antes mencionado para conseguir un árbol de bloques de construcción. Una vez realizado el trabajo anterior y conociendo los casos y eventos que produjeron anomalías se pudo identificar por medio de la herramienta ProM el usuario que ejecutó cada acción en el proceso.

EP8 presenta una breve entrada de registro que contenía una instancia de un proceso, actividad, timestamp y ID del usuario, por lo que se importó el registro: Caso 1, Mostrar reserva, 08:30, Usuario 1, en la herramienta Disco (Fluxicon) 2 y se derivó un modelo visual del comportamiento del usuario. El modelo se analizó de modo que se enfocó en actividades particulares, casos o fallas exitosas, comportamiento frecuente o sospechoso, etc.

EP11 muestra un registro de eventos conformado por los campos: case id, timestamp, activity, resource y cost. Se utilizan las herramientas ProM, Disco, Perceptive Process Mining y Celonis para analizar el mismo registro de eventos y enfocarse en los datos de eventos usados para la MP.

Se presentan en EP12 tres casos de estudio: 1) Ventas y Distribución, 2) Compras, 3) Servicios legales, donde fueron analizados por medio de la herramienta Disco y se mostró el modelo real de ejecución del proceso del caso 1 en comparación con la modelación del proceso teórico en BPMN para poder verificar las diferencias entre el proceso teórico y el proceso real. Del caso 2 no se mostraron modelos resultantes por la alta complejidad debido al volumen de información y del caso 3 se tuvo un alto nivel de confiabilidad con la información por lo que no se mostraron diagramas ni datos relacionados al caso.

En EP17 se mostraron los casos: A) Número mínimo recomendado de instancias y B) Todas las instancias. La intención era utilizar diferentes tamaños de muestreo, por lo que, en el caso A, la evaluación se realizó utilizando el número de instancias mínimo recomendado (10, según la

fórmula SCAMPI) y en el caso B, el número máximo de instancias de procesos disponibles (1911), aunque el número mínimo requerido por la fórmula del método SCAMPI (Valle et al., 2017) resultó en solo 4 instancias (Valle et al., 2017). Lo obtenido de los casos de estudio fue muy similar, desplegando los resultantes del caso A en donde se mostró los resultados de la aplicación de un algoritmo que comparó el modelo de Jure con el modelo de Facto, la aplicación de un algoritmo que realizó la verificación de conformidad entre un registro de eventos y su modelo relacionado de Jure y por último los resultados de la verificación de conformidad con respecto a algunas reglas comerciales y el registro de eventos.

Basarse en trabajos anteriores para obtener el contexto de la investigación y antecedentes relevantes permitió una extracción de información que se analizó y criticó los resultados pertinentes, esto permitió poder afirmar que utilizar herramientas y técnicas de MP puede mejorar el monitoreo y control de los registros de eventos asociados a las actividades del método de evaluación SCAMPI y auxiliar a los proyectos de desarrollo de software y/o de SPI.

Capítulo 3. Marco de Trabajo para la identificación y selección de algoritmos y herramientas de Minería de Procesos

En este capítulo se enfatiza, como tema crítico y de interés para las organizaciones de software el proceso de identificación y selección de herramientas para el modelado, análisis, seguimiento y explotación de información, así como de herramientas, algoritmos y técnicas de MP (Urrea-Contreras et al., 2018b). Por tal motivo, existe la necesidad de contar con un Marco de Trabajo que contemple un Modelo de referencia (MoProPEI-MP- Modelo de referencia de procesos para la selección y adopción de algoritmos y herramientas de Minería de Procesos) para el proceso de identificación y selección de herramientas de software, algoritmos y técnicas de MP y una Guía Técnica que presente conocimiento tópico relacionado a los algoritmos y técnicas de MP. La Figura 9 muestra el esquema de los dos componentes del Marco de Trabajo propuesto en esta investigación.

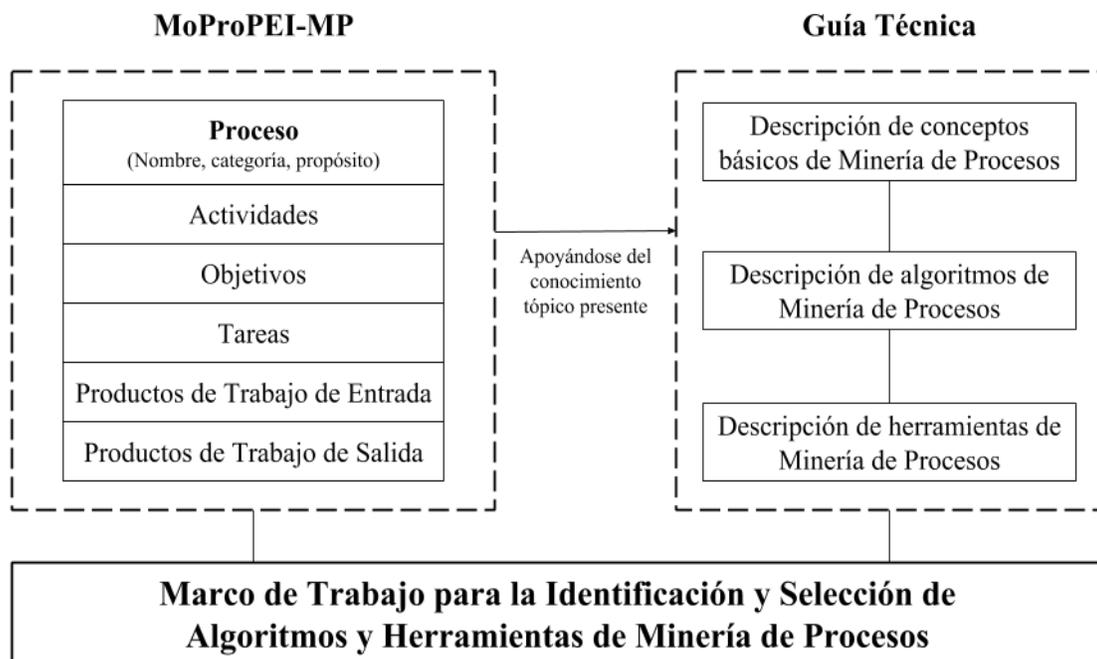


Figura 9. Esquema de los componentes del Marco de Trabajo para la Minería de Procesos.

3.1 Modelo de Referencia para la Minería de Procesos: MoProPEI-MP

Un modelo de referencia ayuda a brindar una perspectiva integradora de los elementos de una disciplina, asistido de la Ingeniería de Procesos permite diseñar, planificar y alcanzar resultados favorables en la ejecución de los procesos para que éstos evolucionen hacia la mejora. Por lo anterior, se tomó como base el Modelo de Procesos para la Ingeniería de Explotación de Información MoProPEI, propuesto por Martins et al. (2014), el cual fue adaptado en esta investigación a un modelo de referencia en la disciplina de MP bajo un enfoque de Ingeniería de Procesos.

3.1.1 MoProPEI

En la Figura 10 se observa la estructura general de MoProPEI con los componentes de cada uno de los dos subprocesos de Gestión y Desarrollo que conforman MoProPEI. Específicamente, el subproceso de Desarrollo se compone de seis fases: Entendimiento del dominio, Entendimiento de los datos, Modelado, Preparación de los datos, Implementación, Evaluación y Presentación (Martins et al., 2014).

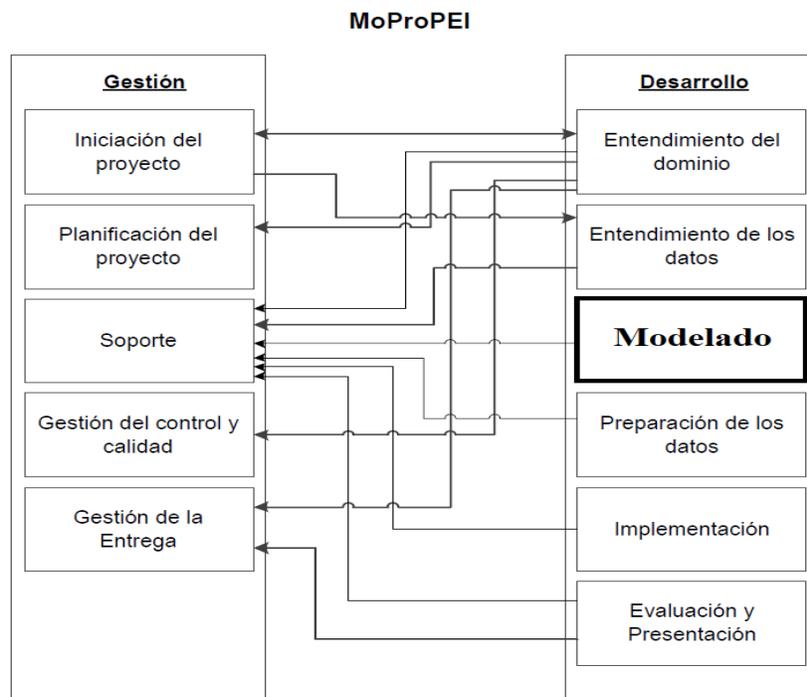


Figura 10. Estructura General MoProPEI (Martins et al., 2014).

Los autores Martins et al. (2014) presentan como una de las características de MoProPEI la adaptación a las necesidades del proyecto de minería de datos que se desarrolla permitiendo realizar un ajuste a la ejecución de las actividades entorno a la MP. De igual manera, su enfoque hacia la gestión fortalece la planificación, administración y documentación de los aspectos necesarios para el desarrollo del proyecto. Partiendo de esta flexibilidad y adaptabilidad, los subprocesos de Gestión y Desarrollo de MoProPEI fueron considerados como Categorías de Procesos. Las fases se analizan como procesos y las actividades, tareas y Productos de Trabajo (PT) se orientan hacia la identificación y selección de algoritmos y herramientas para la MP. Esta adaptación se llamó MoProPEI-MP (Urrea-Contreras et al., 2018a).

La fase de Modelado de MoProPEI se tomó como base para el modelo de referencia MoProPEI-MP debido a que contiene dos actividades relacionadas con la identificación y selección de algoritmos y herramientas pero para la Minería de Datos. El objetivo de la primera actividad, **A1. Modelado del problema**, es definir el problema mediante un vocabulario técnico y determinar técnicas para explotar la información que se implementan a fin de resolver los problemas identificados. Esta actividad comprende las siguientes tareas: T1.1 *Definir el problema de la explotación de la información* y T1.2 *Modelar el problema de la explotación de la información*. La segunda actividad, **A2. Configuración del modelo**, tiene el objetivo de establecer las herramientas y los algoritmos que se aplicarán en función del problema y las técnicas previamente identificadas. Incluye las tareas: T2.1 *Identificar herramientas alternativas*, T2.2 *Seleccionar herramientas*, T2.3 *Identificar algoritmos de minería de datos* y T2.4 *Seleccionar algoritmos de minería de datos* (Martins et al., 2014). Por lo anterior, las actividades y tareas de la fase de Modelado se consideran con gran potencial para ser adaptadas a la disciplina de MP.

3.1.2 Descripción general del Proceso de Modelado en MoProPEI-MP

De la categoría de Desarrollo, el proceso de Modelado de MoProPEI-MP se centra en las actividades A1 y A2, que comprenden las tareas T1.1, T1.2, T2.1, T2.2, T2.3 y T2.4 y los PT. En este proceso, las tareas se modificaron para su mejor adaptabilidad a la MP de forma que las tareas de **A1. Modelado del problema** son: T1.1. *Definir el problema* que se va a resolver en función de los problemas de negocio identificados bajo un enfoque de MP, y T1.2. *Modelar el problema a resolver* para la identificación de la solución bajo enfoque de MP. Las tareas de **A2. Configuración del modelo** son: 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, y T2.4. *Seleccionar la herramienta de MP* tomando en cuenta las herramientas de MP identificadas.

Con la modificación de las tareas anteriores, se originaron nuevos PT de entrada y salida con base en las necesidades que mostraba MoProPEI-MP para cumplir con una integración de todos sus elementos y lograr la realización de las tareas de forma aprobatoria. En la siguiente sección se describen los PT de MoProPEI-MP así como las técnicas empleadas para su aceptación y desempeño.

3.1.3 Descripción de Productos de Trabajo

Los PT son el resultado de la ejecución de las tareas; éstos se manejan como PT de entrada y salida pero con el comportamiento de que pueden ser tanto de salida para una tarea así como de entrada para otra. La Tabla 7 enumera los siete PT definidos y utilizados por las tareas de las actividades A1 y A2 establecidas en la sección 3.1.2, así como las técnicas que se deberán de emplear para determinar su cumplimiento.

Tabla 7. Productos de trabajo relacionados al proceso de Modelado en MoProPEI-MP.

Producto de Trabajo	Técnicas empleada para su cumplimiento
PT1 Problemas de negocio identificados.	Técnica grupo nominal
PT2 Problemas a resolver.	Lluvia de ideas
PT3 Identificación de la solución.	Diagrama Ishikawa
PT4 Algoritmos de MP identificados.	Lista de verificación - Guía Técnica
PT5 Algoritmos de MP seleccionados.	Comparación
PT6 Herramientas de MP identificadas.	Lista de verificación - Guía Técnica
PT7 Herramientas de MP seleccionadas.	Comparación

Para producir el PT1 se necesita identificar aquellos problemas que obstaculizan la ejecución correcta de los procesos definidos o que tengan potencial para ser mejorados con base en la respuesta de las preguntas *¿Qué problemas están surgiendo?* y/o *¿Qué se puede hacer para que mejore lo que ya existe?*. La respuesta a dichas preguntas se puede establecer con la Técnica de Grupo Nominal ya que brinda ayuda técnica en las reuniones que solicitan la opinión de expertos y reduce errores en la toma de decisiones en grupo.

Con respecto al PT2, se refiere a qué tipo de problemas se afronta (identificados en el PT1), cómo actuar ante ellos y la prioridad de cada uno para optimizar tiempo. En general, se tiene que establecer la prioridad si se trata de un problema oculto (aquel que no se es consciente que se posee), inactivo (se tienen noción de él pero su resolución puede esperar), activo (se ha buscado una solución pero sin éxito) y urgente (necesaria próxima solución al problema). Para lo anterior, la técnica de lluvia de ideas ayuda a imponer una prioridad a cada problema identificado con el fin de realizar una selección unánime y de esa forma el problema con mayor ponderación será el primero en darle solución.

En relación con el PT3, se debe identificar y definir las soluciones existentes con el fin de poder validar la mejor solución para el problema con mayor prioridad establecido en el PT2. Utilizar un diagrama Ishikawa ayudará a tener una visión mejorada al enfocarse en los factores que tiene mayor fundamentación para la solución, tomar decisiones eficientes y acciones pertinentes.

Para dar hincapié a la MP, es necesario obtener el PT4 con base en la realización de una precisa identificación de todos los algoritmos de MP que pueden adaptarse para la solución del PT3. La exploración de las características de los algoritmos puede ser ejecutado con la técnica de lista de verificación con ayuda de la Guía Técnica, en donde se valore las similitudes y diferencias que puede ofrecer los algoritmos al beneficio de la solución con la comparación de criterios aceptables.

En el PT5 será la selección de los algoritmos de MP identificados en el PT4, utilizar la técnica de comparación permitirá que el PT5 conteste la pregunta: ¿Cuál(es) algoritmos de MP debo seleccionar?, ya que se podrá conocer las ventajas y desventajas de la aplicación de los algoritmos de MP.

Una vez que el PT5 se encuentre terminado y definido, se deberá obtener el PT6, el cual podrá ser realizado igual que el PT4 por medio de la técnica lista de verificación que conforme al seguimiento de los algoritmos de MP del PT5, evaluará aquellas herramientas que puedan procesar el(los) algoritmo(s) que se necesiten.

Si un responsable de la base de conocimiento no cuenta con información de algoritmos y herramientas no podrá hacer la identificación de éstas, por lo que la Guía Técnica le brindará conocimiento tópico sobre los PT4 y PT6.

Para finalizar la adquisición de todos los PT, se concluirá con el PT7, de tal forma que con la técnica de comparación se analizarán aquellas ventajas y desventajas de seleccionar la

herramienta de MP adecuada, verificar si los resultados que se buscan alcanzar los pueden brindar las herramientas de libre uso o se tendrá que hacer uso de las comerciales, así como constatar la capacidad de registros con que se cuenta y puedan manejar las herramientas de MP.

Con el objetivo de seguir las técnicas vinculadas por cada PT en la ejecución de las tareas, se presentan las preguntas de la Tabla 8. Estas preguntas están relacionadas al cumplimiento de los siete PT reflejando soluciones a las incógnitas que pueda identificar el equipo o responsables de las bases de conocimiento al inicio del proceso.

Tabla 8. Preguntas relacionadas al cumplimiento de los PT.

PT	Preguntas
PT1	¿Qué problemas están surgiendo? ¿Por qué? ¿Cuál es el origen de los problemas identificados? ¿Qué se puede hacer para que mejore lo que ya existe?
PT2	¿A qué tipo de minería se está afrontando el problema? ¹ ¿Cuál es la prioridad de dichos problemas?
PT3	¿Cómo actuar frente a los problemas identificados? ¿Qué soluciones existentes se le puede asignar al problema? ¿Cómo se puede definir la solución apropiada?
PT4	¿Cuáles algoritmos de MP pueden implementarse para la solución identificada?
PT5	¿Cuál(es) algoritmos de MP debo seleccionar? ¿Qué ventajas y desventajas tiene cada algoritmo de MP?
PT6	¿Cuáles herramientas de MP pueden procesar los algoritmos de MP seleccionados?
PT7	¿Cuál(es) herramientas de MP debo seleccionar? ¿Qué ventajas y desventajas tiene cada herramienta de MP?

Proveer un contexto para los PT con base en las técnicas de la Tabla 7, es uno de los fundamentos hacia la mejora. Se desea reducir la variación de la ejecución al establecer preguntas vinculadas a la obtención del PT (Tabla 8) y el manejo adecuado de los resultados del cumplimiento de las tareas, que originan tanto PT de salida como PT de entrada, como se abordará en el modelado de flujo de trabajo.

¹ Consultar Anexo A.

3.1.4 Descripción de Actividades

En esta sección, se describe, de manera general, el proceso de Modelado de MoProPEI-MP por medio de un patrón de procesos². Este patrón comprende las actividades A1 y A2 (Tabla 9), la cual se compone por el Nombre del proceso, actividad, objetivos, tareas, PT de entrada y salida, bajo el objetivo de identificar y seleccionar algoritmos y herramientas de MP (Urrea-Contreras et al., 2018a).

Un flujo de trabajo involucra un número de pasos lógicos asociados a una tarea y éstas a su vez a 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 Figura 11 muestra el flujo de las tareas así como de los PT del proceso de Modelado para alcanzar los objetivos correspondientes (Urrea-Contreras et al., 2018a). 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. Por lo anterior, se observa que el flujo de trabajo debe ser respetado en la forma de realización y estructuración de las tareas que fueron redefinidas ya que el orden en que se encuentran es cómo fluye la información y el seguimiento del PT de salida a la tarea anterior es el PT de entrada de la tarea actual.

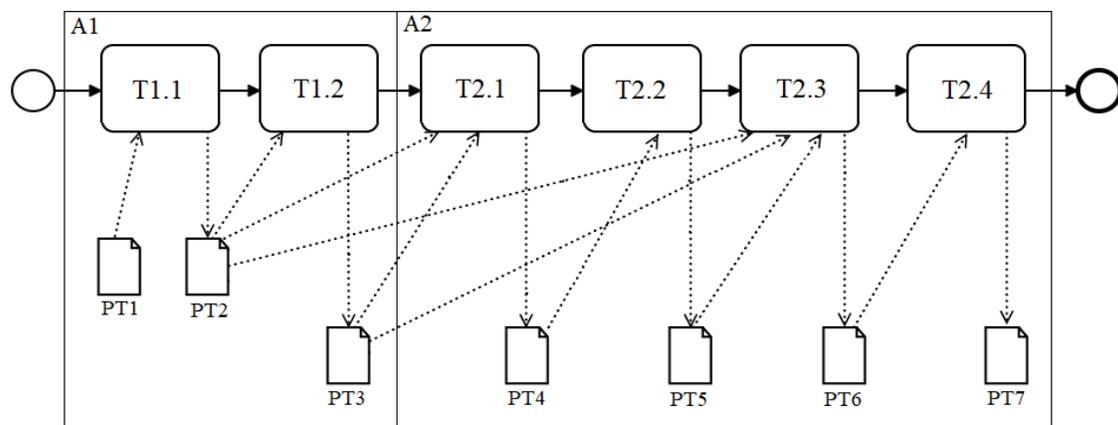


Figura 11. Flujo de trabajo de las tareas relacionadas a las actividades A1 y A2 para el proceso de Modelado de MoProPEI-MP.

² El patrón de procesos es un esquema de elementos que sirve para la documentación de los procesos (Oktaba et al., 2005).

Tabla 9. Descripción del proceso de Modelado de MoProPEI-MP en un patrón de procesos.

Proceso de Modelado	
Actividades	
A1 Modelado del Problema.	A2 Configuración del Modelo.
Objetivos	
<p>O1.1 Determinar mediante vocabulario técnico los Problemas a Resolver de acuerdo a los Problemas de Negocio Identificados.</p> <p>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.</p>	<p>O2.1 Determinar los algoritmos de MP a utilizar con base en las características de la identificación de la solución.</p> <p>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.</p> <p>O2.3 Formalizar la selección de la(s) herramienta(s) elegida(s) para la solución del problema.</p>
Tareas	
<p>T1.1 Definir el Problema a resolver con base en los Problemas de negocio identificados bajo un enfoque de MP.</p> <p>T1.2 Modelar el Problema a resolver para la Identificación de la Solución bajo enfoque de MP.</p>	<p>T2.1 Identificar los Algoritmos de MP a partir de los Problemas a resolver y la Identificación de la solución.</p> <p>T2.2 Seleccionar los Algoritmos de MP con base a los Algoritmos de MP identificados.</p> <p>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.</p> <p>T2.4 Seleccionar la Herramienta de MP tomando en cuenta las Herramientas de MP identificadas.</p>
Productos de trabajo de Entrada	
<p>T1.1: PT1</p> <p>T1.2: PT2</p>	<p>T2.1: PT2, PT3</p> <p>T2.2: PT4</p> <p>T2.3: PT2, PT3, PT5</p> <p>T2.4: PT6</p>
Productos de trabajo de Salida	
<p>T1.1: PT2</p> <p>T1.2: PT3</p>	<p>T2.1: PT4</p> <p>T2.2: PT5</p> <p>T2.3: PT6</p> <p>T2.4: PT7</p>

3.2 Implementación y Evaluación del Proceso de Modelado de MoProPEI-MP

Para evaluar una trazabilidad de las tareas de MoProPEI-MP con el trabajo relacionado en la disciplina de MP durante el periodo de 2007 hasta 2017, presentado en el capítulo anterior, se asociarán los 17 EP con las tareas T1.1, T1.2, T2.1, T2.2, T2.3, T2.4 de las actividades A1 y A2 del proceso de Modelado (Tabla 10). De esta forma, se validará cuáles especifican a un mayor detalle la información requerida por las dos actividades. El resultado se visualiza en la Tabla 10 resaltando con color gris el renglón del EP que cumple con lo requerido por las tareas. Cabe señalar que el 100% de los EP cumplieron con las tareas T2.1 y T2.3 orientadas a la identificación de algoritmos y herramientas. Por otro lado, el 76.47% de los EP indicaron una selección de algoritmos y herramientas de MP por lo que se asocian con las tareas T2.2 y T2.4.

Tabla 10. Trazabilidad de los EP con las tareas del proceso de Modelado de MoProPEI-MP.

ID	A1 Modelado de problema		A2 Configuración del modelo	
	T1.1	T1.2	T2.1, T2.3	T2.2, T2.4
EP1	X	X	X	X
EP2	X	X	X	X
EP3	X	X	X	X
EP4			X	
EP5			X	X
EP6	X	X	X	X
EP7	X		X	X
EP8	X	X	X	X
EP9			X	X
EP10			X	X
EP11	X	X	X	X
EP12	X	X	X	X
EP13	X	X	X	
EP14		X	X	X
EP15			X	
EP16		X	X	
EP17	X	X	X	X

Para la actividad A1, se observa que solo el 58.82% de los EP especificaron el problema (T1.1) y el 64.7% modelaron el problema (T1.2). Se identificaron las herramientas de MP que fueron seleccionadas e implementadas por los 17 EP (Tabla 11), algunas de ellas fueron: ProM, Futura Reflect, Aris PPM, Flow, ProcessAnalyzer, Disco, Perceptive Process Mining y Celonis. Sin embargo, algunos EP no implementaban el total de herramientas identificadas. En la Tabla 11 se resalta de color gris las herramientas que fueron implementadas por los EP (Urrea-Contreras et al., 2018a).

Tabla 11. Identificación e implementación de herramientas de Minería de Procesos por EP.

Herramientas	EP1	EP2	EP3	EP6	EP8	EP11	EP12	EP17
ProM	X	X	X	X	X	X		X
Futura Reflect	X	X						
CPN Tools		X						
Aris PPM		X			X	X		
Flow		X						
ProcessAnalyzer		X			X			
Process Discovery Focus		X						
Comprehend		X						
Interstage Process Discovery		X				X		
Discovery Analyst		X				X		
Disco					X	X	X	
Perceptive Process Mining					X	X		
Celonis					X	X		
XMAnalyzer						X		

La gestión de flujo de trabajo para procesos requiere de herramientas de software que incluyan elementos para definirlos, ejecutarlos, interfaces para utilizarlos, compartirlos, monitorearlos y controlarlos (Pressman, 2014). Las organizaciones de software han ido madurando y siendo eficaces en la recopilación, organización y almacenamiento de una gran cantidad de datos relacionada a la ejecución de sus actividades diarias (registro de eventos). Sin embargo, algunas de ellas no cuentan con algoritmos ni herramientas de MP que les permita reorganizar o mejorar

sus procesos o generar conocimiento útil para conocer la forma real de cómo realizan sus actividades de desarrollo de software o SPI (Urrea-Contreras et al., 2018b).

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 base para la definición de un modelo de referencia que manifieste un conjunto de actividades orientadas a la generación y aplicación de conocimiento para la mejora de los resultados obtenidos por la utilización de herramientas de MP.

En este capítulo se presentó el componente MoProPEI-MP. A continuación, se presenta el componente faltante del Marco de Trabajo: la Guía Técnica, la cual describe los conceptos e información relevante relacionados a las herramientas, algoritmos y técnicas de MP. Dicha guía tendrá como propósito el ser una base para el cumplimiento del flujo de trabajo de las tareas de MoProPEI-MP.

Capítulo 4. Algoritmos y Herramientas para la Minería de Procesos

Este capítulo tiene como propósito presentar una descripción tónica apoyada de una Guía Técnica (Anexo A) que enfatice los fundamentos relacionados a la disciplina de MP, la cual ayude a las organizaciones dedicadas al desarrollo de software a unificar y simplificar: 1) la identificación de los algoritmos y técnicas haciendo uso del modelo MoProPEI-MP, y 2) la selección de las principales características de las herramientas de MP, presentado en el capítulo anterior. De esta manera, dichas organizaciones podrán crear y hacer uso de tanto conocimiento explícito y tónico (Guía Técnica) como de nuevo conocimiento tácito al capacitar técnicamente a responsables de áreas de análisis de tendencias tecnológicas orientadas a la toma de decisiones sobre la adquisición, adopción, perfeccionamiento o comercialización de tecnologías requeridas, procesos de innovación y desarrollo (I+D), investigación y transferencia de conocimiento o tecnología, quienes consideran la interacción de herramientas, algoritmos y técnicas orientadas a la MP con su alta madurez organizacional, medio ambiente o ecosistema (Urrea-Contreras et al., 2018b). Por ejemplo, para un líder de proyectos de software o responsable de un proyecto SPI los datos en sí no son lo más relevante, sino el conocimiento que se encierra en sus relaciones, fluctuaciones y dependencias (Pollo-Cattaneo et al., 2012).

4.1 Algoritmos y Técnicas de Minería de Procesos

Las técnicas de extracción de procesos pueden extraer conocimiento de los registros de eventos. Para encontrar un modelo de proceso sobre la base de un registro de eventos, se debe analizar el registro para determinar las dependencias causales, por ejemplo, si a una actividad siempre le sigue otra actividad, es probable que exista una relación de dependencia entre ambas actividades (Gupta, 2014).

Para aplicar con éxito la MP en la práctica, se tiene que lidiar con el ruido y la falta de los registros de eventos completos. Esta sección se centra en técnicas de descubrimiento de procesos más avanzadas, con el objetivo de proporcionar una visión general de los enfoques más relevantes hacia las técnicas de MP; esto brindará una manera asertiva de seleccionar la técnica de descubrimiento de proceso apropiada. Además, los conocimientos sobre las fortalezas y

debilidades de los diversos enfoques apoyan la interpretación correcta y el uso efectivo de los modelos descubiertos (Van der Aalst, 2016).

4.1.1 Algoritmo Alpha (α)

El algoritmo α toma un registro de eventos y produce una red de Petri que explica el comportamiento en el registro, es capaz de construir automáticamente la red de Petri sin utilizar ningún conocimiento adicional. Si el registro de eventos contiene información sobre recursos, también se pueden descubrir modelos relacionados con los recursos.

El pseudocódigo del algoritmo α se describe L como un registro de eventos sobre un conjunto T de actividades. Sea L un registro de eventos sobre $T \subseteq A$. $\alpha(L)$ se define como (Van der Aalst, 2004):

1. $T_L = \{t \in T \mid \exists \sigma \in L \ t \in \sigma\}$,
2. $T_I = \{t \in T \mid \exists \sigma \in L \ t = first(\sigma)\}$,
3. $T_O = \{t \in T \mid \exists \sigma \in L \ t = last(\sigma)\}$,
4. $X_L = \{(A, B) \mid A \subseteq T_L \wedge A \neq 0 \wedge B \subseteq T_L \wedge B \neq 0 \wedge \forall a \in A \forall b \in B \ a \rightarrow_L b \wedge \forall a_1, a_2 \in A \ a_1 \#_L a_2 \wedge \forall b_1, b_2 \in B \ b_1 \#_L b_2\}$,
5. $Y_L = \{(A, B) \in X_L \mid \forall (A', B') \in X_L \ A \subseteq A' \wedge B \subseteq B' \Rightarrow (A, B) = (A', B')\}$,
6. $P_L = \{p_{(A,B)} \mid (A, B) \in Y_L\} \cup \{i_L, o_L\}$,
7. $F_L = \{(a, p_{(A,B)}) \mid (A, B) \in Y_L \wedge a \in A\} \cup \{(p_{(A,B)}, b) \mid (A, B) \in Y_L \wedge b \in B\} \cup \{(i_L, t) \mid t \in T_I\} \cup \{(t, o_L) \mid t \in T_O\}$, and
8. $\alpha(L) = (P_L, T_L, F_L)$.

En el Paso 1, se verifica qué actividades aparecen en el registro (T_L). Estos corresponderán a las transiciones de la red de flujo de trabajo generada. T_I es el conjunto de actividades de inicio, es decir, todas las actividades que aparecen primero en alguna traza (Paso 2). T_O es el conjunto de actividades finales, refiriéndose a todas las actividades que aparecen en último lugar en alguna traza (Paso 3). Los pasos 4 y 5 forman el núcleo del algoritmo α . El desafío es determinar los lugares de la red de flujo de trabajo y sus conexiones. El objetivo es construir los lugares nombrados $P_{(A, B)}$ de manera que A es el conjunto de transiciones de entrada ($\bullet P_{(A, B)} = A$) y B es el conjunto de transiciones de salida ($P_{(A, B)} \bullet = B$) de $P_{(A, B)}$ (Van der Aalst, 2016).

El algoritmo α básico de 8 líneas descrito anteriormente posee limitaciones cuando se trata de patrones de procesos particulares como bucles cortos o dependencias no locales (Van der Aalst, 2016, Alves de Medeiros, s.f.). Otras limitaciones que presenta se encuentran la sensibilidad al ruido y el estado incompleto del registro de eventos, incluso si el registro de eventos se encuentra

completo, algunas redes de flujo de trabajo diferentes tienen el mismo comportamiento, es decir, dos modelos pueden ser estructuralmente diferentes pero equivalentes a la traza.

El algoritmo α fue uno de los primeros algoritmos de descubrimiento de procesos con capacidad de descubrir una gran clase de modelos. Hoy en día hay algoritmos más eficaces que superan las debilidades del algoritmo α . Estas son variantes del algoritmo α o algoritmos que utilizan un enfoque completamente diferente, por ejemplo la síntesis basada en regiones o el algoritmo genético que se presentará en la siguiente sección (4.2.2).

4.1.2 Algoritmo Genético

El algoritmo genético se utiliza para descubrir una red de Petri dado un conjunto de trazas de eventos. Los algoritmos genéticos son métodos de búsqueda adaptativos que intentan imitar los enfoques evolutivos (Saravanan et al., 2011). El objetivo de usar algoritmos genéticos es abordar problemas como actividades duplicadas, actividades ocultas, construcciones que no sean de libre elección, ruido e incompletitud; es decir, superar los problemas de algunos de los enfoques tradicionales (Gupta, 2014).

Los principales pasos del algoritmo genético son (Alves de Medeiros, s.f.): 1) Leer el registro de eventos, 2) Calcular las relaciones de dependencia entre actividades. Construir la población inicial, 3) Calcular el estado físico de los individuos, 4) Parar y/o devolver los individuos más aptos y 5) Crear la siguiente población utilizando los operadores genéticos.

La limitación más evidente que muestran los algoritmos de minería genética es que requiere una gran cantidad de potencia de computación, otras limitaciones se derivan de su enfoque evolutivo ya que no es muy eficiente para modelos y registros grandes lo que puede llevar a invertir mucho tiempo en descubrir un modelo que tenga un buen estado. Los algoritmos genéticos pueden lidiar con ruido, comportamiento infrecuente, tareas duplicadas y tareas invisibles. Permite mejoras y combinaciones incrementales con otros enfoques; de esta forma, es conveniente combinar heurísticas con algoritmos genéticos de MP para mejorar un modelo de procesos obtenido por medio de los algoritmos heurísticos (4.2.3), lo que ayuda al ahorro del recurso del tiempo y puede dar lugar a modelos que con los algoritmos convencionales no se podrían haber obtenido por su búsqueda de solamente dependencias locales (Van der Aalst, 2016).

4.1.3 Algoritmo Heurístico

Utiliza frecuencias y parametrización para manejar el ruido. Se puede usar para expresar el comportamiento principal pero no todos los detalles de un registro de eventos. Este algoritmo extiende el algoritmo alpha (4.2.1) considerando la frecuencia de las trazas en el registro de eventos. Explora la perspectiva de flujo de control de un modelo de proceso (Gupta, 2014).

Según (Alves de Medeiros, s.f.) el procedimiento para la aplicación del algoritmo heurístico es: 1) Leer el registro de eventos, 2) Obtener el conjunto de tareas, 3) Inferir las relaciones de ordenación en función de sus frecuencias, 4) Construir la red basada en relaciones inferidas y 5) Salida de la red.

El pseudocódigo del algoritmo heurístico se describe W un registro de eventos sobre T, $a, b \in T$, $|a >_W b|$ el número de veces que $a >_W b$ ocurre en W, y $|a \gg_W b|$ es el número de veces que un $\gg_W b$ ocurre en W (Weijters y Ribeiro, 2010):

$$a \rightarrow_w b = \left(\frac{|a >_w b| - |b >_w a|}{|a >_w b| + |b >_w a| + 1} \right) \text{ if } (a \neq b)$$

$$a \rightarrow_w a = \left(\frac{|a >_w a|}{|a >_w a| + 1} \right)$$

$$a \rightarrow_w^2 b = \left(\frac{|a \gg_w b| + |b \gg_w a|}{|a \gg_w b| + |b \gg_w a| + 1} \right)$$

T^* es el conjunto de todas las secuencias (es decir, trazas) que se componen de cero o más tareas de T. $W : T^* \rightarrow \mathbb{N}$ es una función de los elementos de T^* a \mathbb{N} (es decir, el número de veces un elemento de T^* aparece en el registro de proceso). En otras palabras, W es un conjunto de trazas. Debido a que el registro de eventos W es un conjunto, la misma traza puede aparecer más de una vez en el registro y los patrones pueden aparecer más veces en una traza (Weijters y Ribeiro, 2010).

4.1.4 Algoritmo Difuso

Los procesos de la vida real resultan ser un modelo menos estructurado de lo que las personas tienden a creer. Desafortunadamente, los enfoques tradicionales de MP tienen problemas para manejar procesos no estructurados. Los modelos descubiertos a menudo son confusos, mostrando todos los detalles sin distinguir lo que es importante y lo que no lo es. El algoritmo de minería

difusa es configurable y permite diferentes vistas fielmente simplificadas de un proceso en particular (Gupta, 2014).

La minería difusa es adecuada para la MP menos estructurados que exhiben una gran cantidad de comportamiento no estructurado y conflictivo. Aplica una variedad de técnicas, como eliminar bordes no importantes, agrupar nodos altamente correlacionados en un solo nodo y eliminar elementos aislados (Van der Aalst, 2016).

El algoritmo difuso se describe su entrada como un conjunto de N transacciones, cada una con un atributo n, mientras que su salida como Fase I: Fusionar atributos de comportamiento similares; Fase II: generar reglas meta; Fase III: Generar conjuntos de elementos difusos frecuentes y Fase IV: hacer reglas de asociación difusas. Después de terminar el algoritmo, se tendrá todas las reglas de asociación difusas resultantes de este nuevo conjunto de datos difusos. Las reglas obtenidas del algoritmo propuesto son un resumen exacto y semántico de todas las reglas creadas en algoritmos anteriores (Farzanyar y Kangavari, 2012).

4.1.5 Análisis comparativo de algoritmos de Minería de Procesos

El trabajo de Chamorro y Maturana (2013) ha proporcionado reseña de los algoritmos α , heurístico, difuso y genético para la selección de los algoritmos más adecuados en la implementación de la MP. Del mismo modo, Gupta, (2014) y Erdem et al. (2018) han realizado un análisis comparativo de algoritmos de MP considerando principalmente el genético, difuso y heurístico. Su análisis se basa en los campos de: descripción, estrategia, salida, cuándo usarlo, problemas a desafiar, comportamiento, formato, registro de eventos y resultados que ofrecen. En esta sección 4.1, se describieron los 3 algoritmos antes mencionados más el algoritmo α , bajo las características de: definición, pseudocódigo del algoritmo (alpha y heurístico), pasos para su ejecución y limitaciones que se puedan obtener.

En la Tabla 12 se presenta una extensión del análisis comparativo de algoritmos realizado por Gupta (2014), basándose de las características de Descripción, Formato de Salida, Problemas a desafiar y Registro de Eventos; agregando el algoritmo α y las características de Enfoque local, Enfoque global y el Plug-in de ProM donde se puede utilizar los 4 algoritmos.

Tabla 12. Comparativo de Algoritmos de MP y la herramienta ProM. Extendido de (Gupta, 2014)³.

Algoritmo	Alpha	Genético	Heurístico	Difuso
Característica				
Descripción	Proporciona una vista de descubrimiento de procesos con capacidad de descubrir una gran clase de modelos	Proporciona una vista de la frecuencia para ambas tareas y la sucesión entre ambas tareas, y descubre todas las estructuras comunes de control-flujo.	Proporciona una vista de los flujos de trabajo científicos considerando la dependencia de larga distancia.	Proporciona una vista ampliable del flujo de trabajo científico mediante el control del límite de significación para mostrar la tarea en diferentes niveles de importancia.
Formato de Salida	Red de Petri	Red de Petri	Red Heurística	Modelo Difuso
Problemas a desafiar	Puede extraer registros con limitaciones cuando se trata de bucles cortos o dependencias no locales.	Puede extraer registros con ruido, el manejo de nombres de tareas duplicados, construcciones locales y no locales de elección no libre y tareas invisibles.	Puede extraer registros que son menos sensibles al ruido, construcciones de elección local y no libres.	Puede extraer registros con ruido, pero no se puede convertir en red de Petri.
Registro de Eventos	El minado α tiene que manejar registros de eventos completos a un cierto punto.	El minado genético puede manejar fácilmente los registros incompletos.	El minado heurístico puede manejar registros incompletos a un cierto punto.	El minado difuso puede manejar registros incompletos
Enfoque local	x		x	
Enfoque global		x		x
Plug-in en ProM	Alpha miner	Genetic miner	Heuristic miner	Fuzzy miner

³ La definición conceptual de los términos de Registro de Eventos, Enfoque local y global se presenta en la Guía Técnica (Anexo A).

4.2 Herramientas de Minería de Procesos

Actualmente, existe un número creciente de sistemas de software disponibles comercialmente que ofrecen capacidades de MP. Algunos de estos productos integran la funcionalidad de MP en un sistema más grande y otros productos apuntan a simplificar la MP utilizando una interfaz de usuario intuitiva. Además de los sistemas comerciales, también hay varios grupos de investigación que desarrollan herramientas de descubrimiento de procesos independientes. Cabe de enfatizar que existen grandes diferencias entre las diferentes herramientas que se mencionaran por la falta de estandarización entre ellas.

En ocasiones, la selección o adquisición de una herramienta inadecuada se deriva de la existencia de diversos proveedores y herramientas de tipo comercial y uso libre que permiten realizar la tarea sin contar con un método objetivo de selección. Esto deriva en las siguientes consecuencias: una pérdida de tiempo, dinero y una alta probabilidad del riesgo por no cumplir con los objetivos de negocio (Pollo-Cattaneo et al., 2012).

Las herramientas de MP, aunque son eficientes en su campo, poseen alta complejidad, configuraciones técnicas y procedimientos intrincados, tanto para extraer registros de eventos como para obtener modelos. Por lo que la aplicación de esta tecnología en su estado actual es limitada para profesionales no expertos (Orellana-García et al., 2016).

En los trabajos de (Turner et al., 2012; Kebede, 2015) se muestra una evaluación entre las herramientas de MP ProM, Disco y Celonis, donde se comparan diferentes características tales como: tipo de archivo, capacidad de registro, licencia, modelos de salida, plataforma de soporte, filtrado de datos, descubrimiento de procesos, verificación de conformidad, minado de red social, minado de reglas de decisión, visualización de proceso, minado de la regla de discriminación, traza de agrupamiento (clusters), y análisis delta.

A continuación, y con base en la Figura 8 del Capítulo 2, se presentan las 7 herramientas de MP con mayor frecuencia de mención en la revisión de literatura (Urrea-Contreras, 2018b).

4.2.1 ProM

ProM es una de las principales herramientas de MP que ha pasado por varias versiones, no hay ningún producto dentro de la disciplina de MP que ofrezca un conjunto comparable de algoritmos de MP en donde, la verificación de conformidad, la MP de múltiples perspectivas y el soporte

operativo se han convertido en partes integrales de dicha herramienta. ProM es un marco extensible que admite una amplia variedad de técnicas de MP en forma de plug-ins. Es una plataforma independiente ya que se implementa en Java y se puede descargar de forma gratuita desde el sitio Web <http://www.processmining.org> (Van der Aalst, 2011).

Como plataforma académica, ProM está a la vanguardia de la investigación MP. Es adecuada para investigadores de MP y facilita el desarrollo y la prueba de nuevos algoritmos (Van der Aalst, 2011). Sin embargo, requiere un cierto nivel de experto para usarlo y no tiene soporte profesional. Por lo tanto, las herramientas de MP comerciales pueden ser una mejor opción para los profesionales que usan técnicas de MP como soporte para sus actividades principales (Ailenei, 2011).

En la Tabla 13 se presentan las características de ProM v. 6.8, la cual puede importar registros de eventos que cumplan con los formatos MXML o eXtensible Event Stream (XES) (IEEE, 2016) y puede cargar definiciones de modelo de proceso en diferentes estándares. Algunas de las funcionalidades principales proporcionadas por ProM son: utilizar la perspectiva de flujo de control de un proceso, permitir el minado de procesos menos estructurados, analizar el recurso y la perspectiva de datos de un proceso, verificar la conformidad de un registro de eventos con un modelo de proceso determinado e inspeccionar las métricas de rendimiento específicas de un proceso. Además, la herramienta admite aplicar una amplia variedad de técnicas de filtrado que se pueden utilizar en un registro de eventos en el paso de preproceso. ProM proporciona varios formatos de exportación, que van desde representaciones visuales de los resultados en PNG hasta diferentes formatos de registro como CSV (Ailenei, 2011).

Tabla 13. Características de la herramienta ProM v. 6.8.

ProM (v. 6.8)	
Desarrollador/ Proveedor	Universidad Tecnológica de Eindhoven
Sitio web	http://www.processmining.org
Tipo de herramienta	Académica / OpenSource
Tipo de formatos de entrada	MXML, XES
Tipo de formatos de salida	BPMN, Redes de Petri, PNG, CVS
Plataformas	Windows, MAC, Linux
Adquisición	Software Libre
Características	<ul style="list-style-type: none"> ✓ Descubrir la perspectiva de flujo de control. ✓ Minar procesos menos estructurados. ✓ Analizar el recurso y la perspectiva del proceso. ✓ Verificar la conformidad. ✓ Inspeccionar métricas de rendimiento de un proceso. ✓ Técnicas de filtrado.

Actualmente, la versión de ProM 6 (Figura 12) se distribuye en partes, lo que ofrece la máxima flexibilidad. Por ejemplo, el núcleo del ProM 6 se distribuye como un paquete que puede ser descargable e instalarse sin restricciones; mientras que los plug-ins de ProM 6 se distribuyen como paquetes separados.



Figura 12. Interfaz de la herramienta ProM 6.
Fuente: <http://www.processmining.org>

4.2.2 Disco

Disco por Fluxicon es una herramienta tipo comercial independiente para el análisis de MP, con un alto manejo de grandes volúmenes de información compleja y facilidad de uso. El algoritmo de MP de esta herramienta se basa en el enfoque de minería difusa. Además, contiene los algoritmos de extracción de procesos, el marco de administración y filtrado de registros más eficiente.

Esta herramienta cuenta con descubrimiento automatizado de procesos, puede crear mapas de procesos directamente desde los datos sin procesar, de forma automática. Cuenta con opciones de elegir el nivel de abstracción deseado, entre seis visualizaciones de métricas de proceso proyectadas directamente en su mapa y crear filtros directamente desde las actividades o las rutas

sin importar cuán grande o complejo sea su proceso. Otras dimensiones como el rendimiento se pueden analizar a través de avanzadas visualizaciones de los modelos de proceso minados.

Disco permite limpiar fácilmente los datos de proceso para enfocar su análisis, ya sea que desee desglosar por caso, periodo de tiempo o relaciones de eventos. Importa archivos CSV y MS Excel, maneja los formatos MXML de ProM y XES. Las principales características de Disco se presentan en la Tabla 14.

Tabla 14. Características de la herramienta Disco.

Disco	
Desarrollador/ Proveedor	Fluxicon
Sitio web	https://fluxicon.com/disco/
Tipo de herramienta	Comercial
Tipo de formatos de entrada	CSV, XLS, XLSX, MXML, XES, FXL, DSC
Tipo de formatos de salida	CSV, MXML, FXL, DSC, PDF, PNG, JPEG
Plataformas	Windows, MAC
Adquisición	Contacto al proveedor para recibir precio de venta
Características	<ul style="list-style-type: none"> ✓ Importación inteligente del archivo log. ✓ Gestión de proyectos. ✓ Descubrimiento automatizado de procesos. ✓ Animación de mapas de proceso. ✓ Estadísticas y gráficos. ✓ Explorador de casos. ✓ Filtrado de registro de eventos. ✓ Filtro de tiempo, variación, rendimiento, punto final, atributos, seguidor ✓ Actualización automática de software. ✓ Sistema de ayuda integrado. ✓ Retroalimentación integrada.

Se puede realizar la visualización de los procesos de manera animada, ayudando de esta forma a la detección de cuellos de botella (Figura 13) y otorgando una experiencia de usuario optimizada, que permite la usabilidad de la herramienta. También ofrece una vista de estadística sobre la descripción general de los datos a partir de gráficos interactivos y obtener información detallada sobre cada actividad, recurso y valor de atributo.

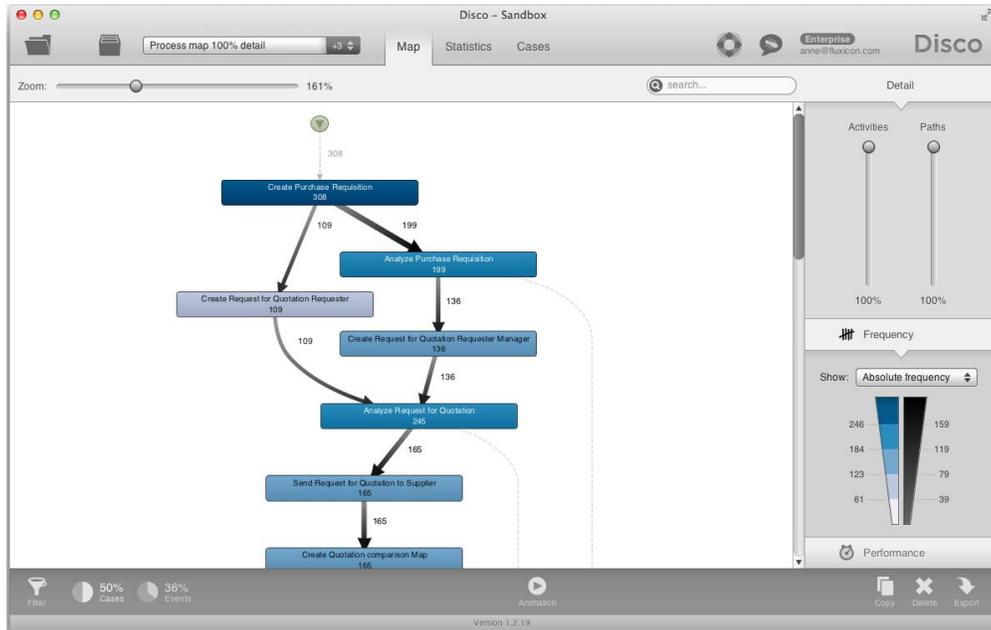


Figura 13. Interfaz de la herramienta Disco.
Fuente: <https://fluxicon.com/disco/>

4.2.3 Aris PPM

ARIS Process Performance Manager (PPM) de Software AG es una herramienta de MP tipo comercial que admite diversas técnicas de MP. La atención se centra principalmente en el análisis del rendimiento (exploración hasta el nivel de instancia, evaluación comparativa y cuadros de mandos). ARIS PPM también apoya a la minería organizacional como se observa en la Tabla 15 y no es compatible con la verificación de conformidad, la predicción y la recomendación (Van der Aalst, 2011).

ARIS PPM admite una amplia gama de formatos de entrada, como archivos CSV, registros de eventos extraídos de bases de datos y sistemas externos (por ejemplo, ERP, CRM, etc.), formato de gráfico PPM y formato de sistema de eventos PPM. Los modelos descubiertos se pueden exportar como archivos de imagen, XML y AML, mientras que los datos generados como resultado de diferentes tipos de análisis se pueden guardar en archivos de Microsoft Office Excel (Ailenei, 2011). La interfaz de Aris PPM se muestra en la Figura 14.

Tabla 15. Características de la herramienta Aris PPM.

Aris PPM	
Desarrollador/ Proveedor	Software AG
Sitio web	https://www.softwareag.com/it/products/process/process_minin_g/default.html
Tipo de herramienta	Comercial
Tipo de formatos de entrada	CSV, Gráfico PPM, eventos PPM
Tipo de formatos de salida	PNG, JPEG, XML, AML, XLS
Plataformas	Windows, MAC
Adquisición	Contacto al proveedor para recibir precio de venta
Características	<ul style="list-style-type: none"> ✓ Análisis de rendimiento. ✓ Minería organizacional. ✓ Descubrimiento de proceso. ✓ Análisis de proceso. ✓ Visualización de gráficos. ✓ Proceso de evaluación comparativa.

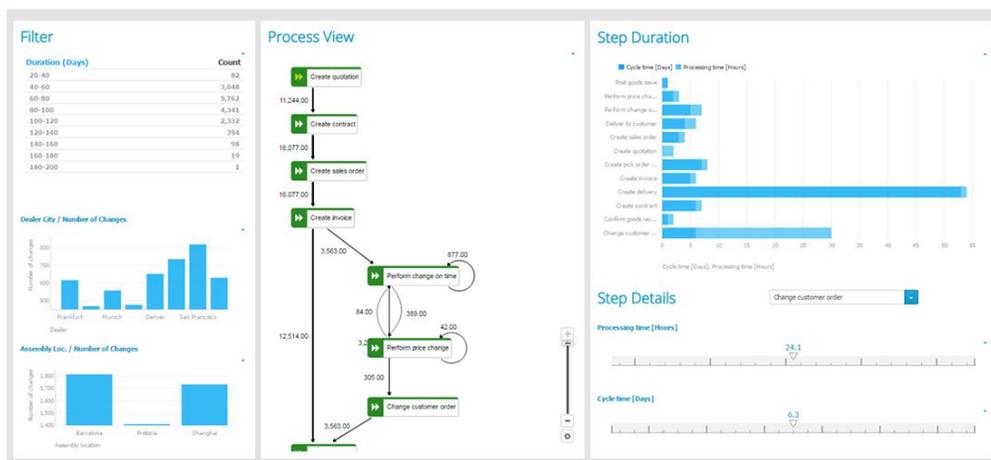


Figura 14. Interfaz de la herramienta Aris PPM.
Fuente: <https://www.softwareag.com>

4.2.4 Perceptive Process Mining (antes Futura Reflect)

Futura Reflect era una de las herramientas más maduras tipo comercial de Pallas Athena. Era posible aplicar técnicas de MP a modelos predefinidos. De esta forma, se podía respaldar todo el ciclo de vida de BPM. Tenía como objetivo la facilidad de uso y la escalabilidad. Utilizaba dos algoritmos de descubrimiento diferentes; uno se basa en el enfoque de minería genética y el otro

asume un modelo secuencial para facilitar el filtrado de un comportamiento infrecuente (Van der Aalst, 2011).

Posteriormente, Futura Reflect paso a ser Perceptive Process Mining (las características de esta herramienta se muestra en la Tabla 16) la cual es una solución de análisis de procesos e inteligencia de procesos de negocios que analiza automáticamente los procesos de negocios de un cliente mediante la extracción de datos de los registros de eventos para identificar cómo fluye el contenido dentro de un proceso, dónde existen cuellos de botella y dónde se producen excepciones.

Tabla 16. Características de la herramienta Perceptive Process Mining.

Perceptive Process Mining (Antes Futura Reflect)	
Desarrollador/ Proveedor	Pallas Athena/ Hyland
Sitio web	www.perceptivesoftware.com
Tipo de herramienta	Comercial
Tipo de formatos de entrada	CSV, XML
Tipo de formatos de salida	PNG, XLS
Plataformas	Windows
Adquisición	Se encuentra en fase de migración de funciones de soporte técnico
Características	<ul style="list-style-type: none"> ✓ Representación visual de los datos del proceso. ✓ Analizar cuellos de botella. ✓ Identificar recursos críticos. ✓ Evaluar la distribución de la carga de trabajo. ✓ Opciones de filtrado. ✓ Evaluar el flujo de información, la afectividad de los departamentos, personas y procesos.

Durante el tercer trimestre del 2017, Hyland adquirió la cartera de productos de Pallas Athena incluyendo Perceptive Process Mining (Figura 15). Actualmente, se encuentran en la primera fase de la migración de funciones de soporte técnico donde a los clientes anteriores a la adquisición tienen la opción de enviar los problemas de soporte técnico a través de la comunidad de Hyland.

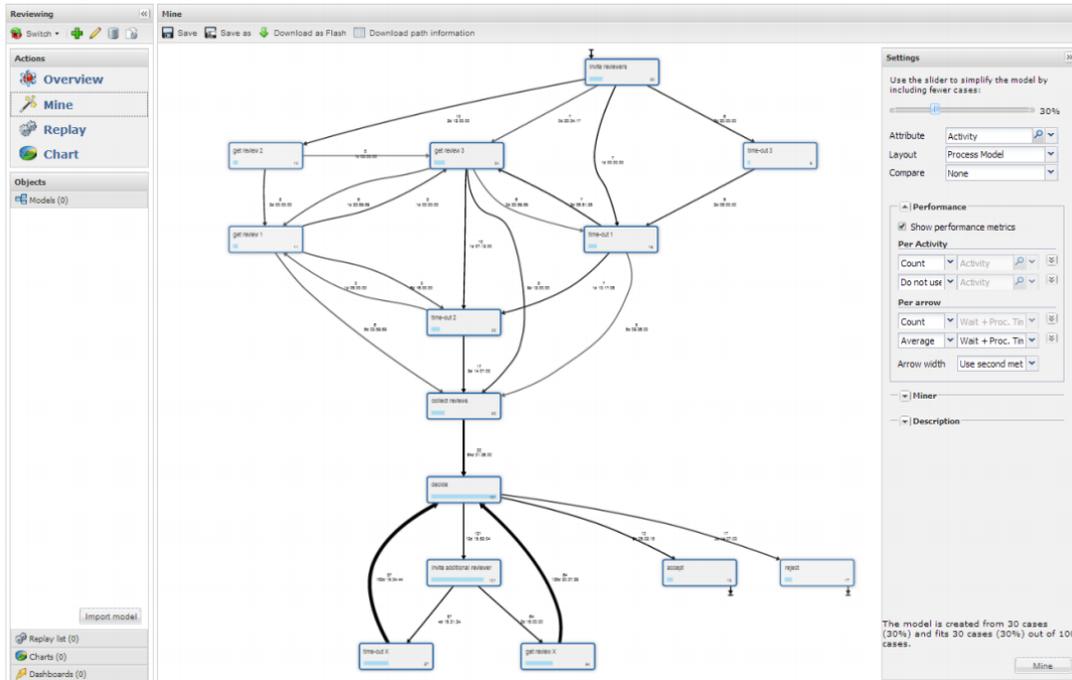


Figura 15. Interfaz de la herramienta Perceptive Process Mining (Van der Aalst, 2015).

4.2.5 CPN Tools

CPN Tools es un entorno de simulación gratuito basado en redes de Petri de color que se puede encontrar en <http://www.cpn-tools.org>. Dicha herramienta puede simular la fusión de las perspectivas flujo de control, dependencias de datos, características de desempeño y características organizacionales en un único modelo (Tabla 17) (Van der Aalst, 2011).

Tabla 17. Características de la herramienta CPN Tools.

CPN Tools	
Desarrollador/ Proveedor	Grupo AIS, Universidad Tecnológica de Eindhoven
Sitio web	http://cpn-tools.org/
Tipo de herramienta	Libre- Simulador de alto nivel de Redes Petri
Plataformas	Windows, Linux
Adquisición	Software libre y con licencia.
Características	<ul style="list-style-type: none"> ✓ Simular la fusión de las perspectivas de flujo de control. ✓ Dependencias de datos. ✓ Características de desempeño. ✓ Características organizacionales. ✓ Capacidad de simular redes Petri de una manera controlada por el usuario y automática.

Los tokens tienen tipos asignados y las redes definidas pueden extenderse con el tiempo y las jerarquías y manipularse utilizando el lenguaje de programación funcional subyacente, SML. Otra funcionalidad importante proporcionada por CPN Tools es la capacidad de simular redes de una manera paso a paso controlada por el usuario y de manera automática, que es más rápida y puede generar información estadística (Ailenei, 2011). La interfaz de CPN Tools se muestra en la Figura 16.

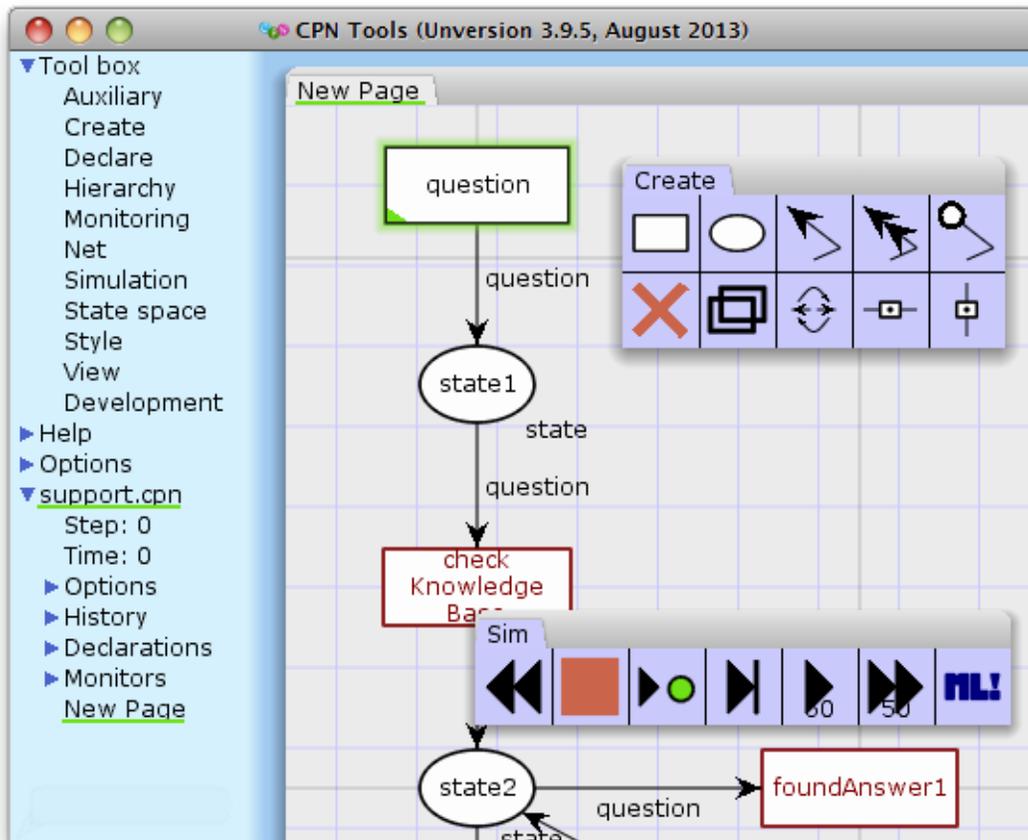


Figura 16. Interfaz de la herramienta CPN Tools.
Fuente: <http://cpntools.org/>

4.2.6 ProcessAnalyzer

ProcessAnalyzer es una herramienta tipo comercial que usa un algoritmo de descubrimiento de procesos inspirado en el algoritmo α y la minería heurística (Van der Aalst, 2011). Ofrece como característica (Tabla 18) la visualización automática de procesos, en donde durante el análisis de un diagrama de flujo muestra el diagrama de flujo del proceso real, donde puede ver cada tipo y

flujo de eventos, el tiempo que transcurre entre cada tipo de evento y la mediana y la duración promedio del proceso visualizado como un diagrama de flujo. Todas las variaciones, repeticiones, bucles y retrasos son fácilmente visibles proporcionando una visión general del proceso.

Tabla 18. Características de la herramienta ProcessAnalyzer.

ProcessAnalyzer	
Desarrollador/ Proveedor	QPR Software
Sitio web	https://www.qpr.com/products/qpr-processanalyzer
Tipo de herramienta	Comercial
Tipo de formatos de entrada	XLS, CSV, MS SQL
Tipo de formatos de salida	PNG, XLS
Plataformas	Windows
Adquisición	Contacto al proveedor para recibir precio de venta
Características	<ul style="list-style-type: none"> ✓ Visualización automática de procesos. ✓ Análisis de la causa raíz. ✓ Análisis del tiempo de espera. ✓ Análisis de conformidad. ✓ Análisis de caso, duración, eventos, diagrama de flujo. ✓ Filtrado de datos. ✓ Fuentes de datos

De igual manera se puede realizar un análisis de la causa raíz (Figura 17) donde el análisis de influencia permite a los usuarios identificar el impacto que tienen ciertos atributos para procesar las variaciones. Los usuarios pueden profundizar fácilmente en sus datos y ver qué factor contribuye más a las variaciones en el proceso. También puede evaluar los diferentes atributos para ver cuál es el más vulnerable a los retrasos en comparación con otros atributos.

ProcessAnalyzer permite el análisis del tiempo de entrega que muestra cuánto tardan los casos y cuántos de ellos caen en cada intervalo de duración. Los usuarios pueden ver de un vistazo cuál es la duración más común y cuántos casos se encuentran en la ranura de mayor duración. Esto es importante ya que la larga duración es una clara indicación de problemas en el proceso. Puede ver los cuellos de botella en el proceso y profundizar para encontrar factores contribuyentes.

Por último, se puede realizar un análisis de conformidad usando un editor BPMN, establece el estado objetivo para su proceso y luego monitorea qué tan bien la ejecución coincide con este objetivo. Al identificar los casos no conformes, sabe dónde dirigirse a las actividades de mejora y obtener mayores beneficios. ProcessAnalyzer proporciona conectores preparados para varias fuentes de datos como lo son Microsoft Dynamics, Oracle, SAP, entre otras, así como la posibilidad de crear conectores personalizados.

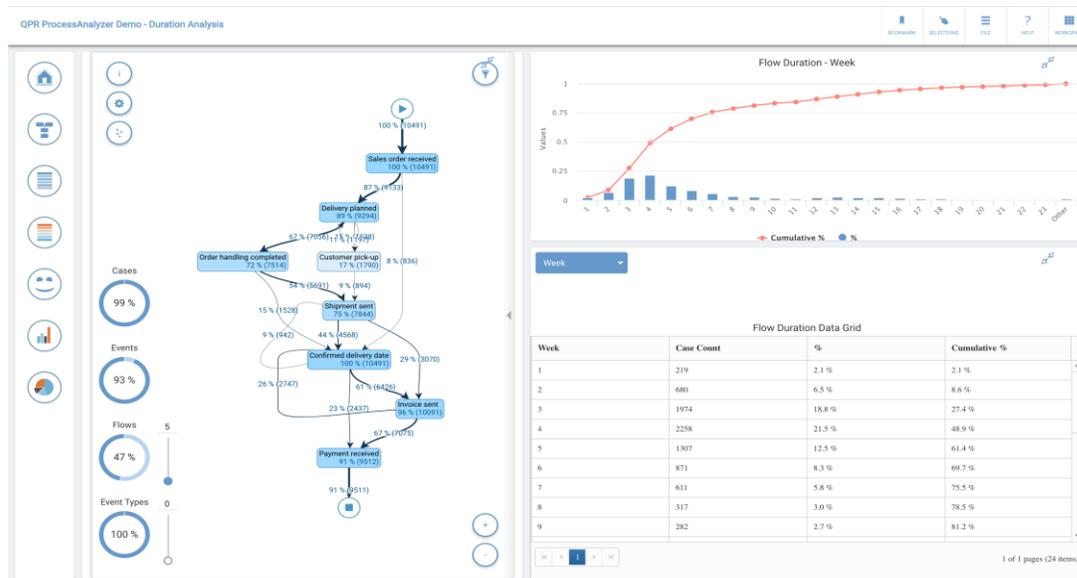


Figura 17. Interfaz de la herramienta ProcessAnalyzer.
Fuente: <https://www.qpr.com/products/qpr-processanalyzer>

4.2.7 Interstage Process Discovery

Interstage Process Discovery, fue desarrollado por Fujitsu como un servicio, es decir, las organizaciones no necesitan instalar un producto de software. La herramienta proporciona información sobre lo que realmente está sucediendo, puede abstraer a la perfección un comportamiento infrecuente pero no puede descubrir la simultaneidad, sus características se muestran en la Tabla 19.

Tabla 19. Características de la herramienta Interstage Process Discovery.

Interstage Process Discovery	
Desarrollador/ Proveedor	Fujitsu
Sitio web	http://www.fujitsu.com/global/products/software/middleware/application-infrastructure/interstage/solutions/bpmgt/bpm-services/apd/
Tipo de herramienta	Comercial
Tipo de formatos de entrada	CSV
Tipo de formatos de salida	BPMN, XPD, BPEL
Plataformas	Offline/Online
Adquisición	Contacto al proveedor para recibir precio de venta
Características	<ul style="list-style-type: none"> ✓ Descubrimiento y análisis de flujos de trabajo y procesos de negocios. ✓ Visualización de flujos de procesos. ✓ Identificación de cuellos de botella. ✓ Minado y análisis de procesos inteligentes. ✓ Uso de casos de procesos inteligentes.

Interstage Process Discovery (Figura 18) puede analizar el rendimiento utilizando indicadores como el tiempo de flujo. No se descubren perspectivas adicionales al tiempo y las funciones avanzadas como la predicción, la recomendación y la verificación de conformidad no son compatibles (Van der Aalst, 2011).



Figura 18. Interfaz de la herramienta Interstage Process Discovery.
Fuente: <http://www.fujitsu.com>

4.2.8 Análisis comparativo de herramientas de Minería de Procesos

Para poder realizar una selección asertiva de la(s) herramienta(s) de MP a utilizar dependerá de las necesidades potenciales con la que los usuarios quieran interactuar con dicha(s) herramienta(s). Por un lado existen usuarios con nivel de expertos que solo requieren responder preguntar particulares; y por otro lado, se encuentran los usuarios que solo buscan consultar descripciones generales de los procesos a través de la MP.

Dependiendo cual sea el fin por el cual se desee utilizar una herramienta de MP, se puede tener preguntas que solo necesiten la interacción de una hoja de cálculo para un análisis con mayor libertad donde la interpretación de resultados origine la extracción de datos adicionales. También existe el caso donde se ocupa un análisis donde los resultados deben ser personalizados y manejados con conocimiento de MP y comprensión de datos. Por último, de forma contraria al caso anterior, se puede tener usuarios que solo demanden los resultados que ofrecen las

herramientas de forma directa sin ninguna personalización. A continuación se presenta la Tabla 20 que concentra las ventajas y desventajas de las 7 herramientas descritas con anterioridad.

Tabla 20. Ventajas y Desventajas de herramientas de Minería de Procesos.

Herramienta	Ventajas	Desventajas
ProM	Ofrece verificación de conformidad, la minería de perspectivas, variedad de técnicas de filtrado y es una plataforma gratuita.	Requiere un cierto nivel de experto para usarlo y no tiene soporte profesional.
Disco	Manejo de grandes volúmenes de información, estadísticas y gráficos, filtrado del registro de eventos y gestión de proyectos.	El precio de venta se envía por correo según las necesidades del cliente.
Aris PPM	Apoya a la minería organizacional y acepta diversas técnicas de MP.	No es compatible con la verificación de conformidad.
Perceptive Process Mining	Utilizaba dos algoritmos de descubrimiento diferentes basados en la minería genética y asumiendo un modelo secuencial para facilitar el filtrado de información.	No se encuentra disponible a la venta.
CPN Tools	Simula la perspectiva de flujo de control y las redes de una manera controlada por el usuario, puede mostrar información estadística.	Solo simula redes de Petri.
ProcessAnalyzer	Usa un algoritmo de descubrimiento de procesos inspirado en el algoritmo α y la minería heurística	El precio de venta se envía por correo según las necesidades del cliente.
Interstage Process Discovery	Analiza el rendimiento utilizando indicadores como el tiempo de flujo.	No descubre perspectivas adicionales al tiempo y la verificación de conformidad no es compatible.

En esta sección, se describieron las características de 7 herramientas de MP (Tabla 13 a la 19), agregando 4 herramientas más en comparación a los trabajos de Turner et al. (2012) y Kebede (2015) quienes hicieron la descripción de ProM, Disco y Celonis. En este documento los elementos que se consideraron fueron: Desarrollador/Proveedor, sitio web de la herramienta, el tipo de herramienta comercial o de acceso libre, formatos de entrada y salida, plataformas en las que se pueden instalar, forma de adquisición y las características que poseen. Posteriormente, se

muestra una relación de ventajas y desventajas. A continuación, se describirán los algoritmos y técnicas principales de MP tomando de referencia los trabajos de Erdem et al. (2018) y Gupta (2014).

Por todo lo anterior, es importante resaltar que la falta de estandarización de las herramientas de MP así como el desconocimiento de los algoritmos existentes ha dificultado a los practicantes, líderes de proyectos, responsables de mejora de procesos de software, responsables de bases de conocimiento o tecnologías u otros, la toma de decisiones asertivas en la selección de herramientas, algoritmos y técnicas de MP adecuados para la solución de problemáticas identificadas. La aportación de este capítulo es presentar las principales características de herramientas, algoritmos y técnicas de MP que fueron resultado de una búsqueda en casos de estudio en artículos publicados internacionalmente (Urrea-Contreras et al., 2018b), como una base tónica y descriptiva para la implementación de la disciplina MP orientada a proyectos de mejora de procesos o desarrollo de software.

Capítulo 5. Evaluación del Marco de Trabajo

En este capítulo, se presenta el proceso de evaluación de MoProPEI-MP por medio de la técnica⁴ de Focus Group (sesión de grupo) y su ejecución para aplicarla en la disciplina de IS propuesta por (Mendoza-Moreno et al., 2013). Además, se exponen el análisis de los resultados empíricos los cuales están basados en la experiencia de los participantes, sus comentarios y hallazgos de del Marco de Trabajo para la identificación y selección de algoritmos y herramientas de MP logrando así una validación empírica.

5.1 Proceso de Evaluación (validación empírica)

Focus Group es un método empírico de investigación, fácil de comprensión y rápido para obtener información cualitativa y retroalimentación (desde un grupo específico), el cual se puede utilizar en varias fases y tipos de investigación (Kontio et al., 2004). Generalmente se realiza con 10 o menos participantes que comparten su experiencia y opinión para medir la viabilidad de lo propuesta a evaluación.

Para aplicar la técnica de Focus Group es necesario realizar una descripción y caracterización de los participantes y planificar las actividades a realizar en la ejecución. Esto significa que Focus Group al ser una técnica de recopilación de información primaria por medio de entrevistas grupales requiere la necesidad de una planificación adecuada al contexto que se necesite aplicar. Por tal motivo, en esta investigación se seguirá el proceso propuesto por (Mendoza-Moreno et al., 2013) dentro de la IS como un método útil para validar propuestas teóricas a partir del juicio de expertos, de quienes su experiencia promueven conceptos de alto valor para aceptarlas o desestimarlas.

En la siguiente sección, se presenta la estructura general y los procedimientos de Focus Group. Adicionalmente se describe la ejecución del proceso propuesto con el fin de obtener una evaluación de MoProPEI-MP por parte de la opinión de expertos.

⁴ Conjunto de procedimientos y recursos de que se sirve una ciencia o un arte (RAE, 2019).

5.1.1 Estructura General de Focus Group

La propuesta de aplicar de Focus Group en la disciplina de IS define un proceso que incluye los elementos de Fases (F), Actividades (A), Tareas (T), Productos de Trabajo (PT) y roles que guían su ejecución. Según lo planteado por (Mendoza-Moreno et al., 2013) el proceso posee cuatro fases (F1, F2, F3, F4) que están compuestas por ocho actividades (A1, A2, A3, A4, A5, A6, A7, A8) y estas a su vez por tareas (T1 – T13, para cada una de las tareas se vinculan roles (descritos en la Tabla 21) y PT (PT1 – PT17, PTF1) que pueden ingresar o derivarse (de tipo entrada o salida).

Tabla 21. Descripción de los roles de Focus Group.
Adaptado de (Mendoza-Moreno et al., 2013).

Roles	
Grupo investigador	Conjunto de investigadores que abordan una temática específica, soportados en la aplicación de Focus Group.
Moderador	Actor del Grupo Investigador que coordina el debate, sin influenciar las opiniones de los participantes, y controla la sesión para evitar dispersión y pérdida del contexto.
Relator	Actor encargado de ejecutar la captura y registro de información.
Participante	Actor que aporta su criterio en el debate. Preferentemente, deben ser expertos en la temática objeto de debate.

De manera general, la F1. Planteamiento de la investigación (Tabla 22) es desarrollada por los integrantes del grupo investigador quienes realizan dos actividades: A1. Definir el problema de investigación y A2. Preparar los materiales y métodos. Para su ejecución se requiere el producto PT1. “Recursos objeto de debate” relacionado a documentos para contextualizar al grupo de investigadores y generar los PTF1. “Recursos de planeamiento” que es una composición de subproductos resultantes de las tareas de la F1.

Tabla 22. Descripción de la Fase Planteamiento de la investigación.
Adaptado de (Mendoza-Moreno et al., 2013).

F1. Planteamiento de la investigación	
Establece los elementos de contenido y el procedimiento que se aplicará a los participantes.	
Rol Involucrado	Grupo investigador
Actividades	<ul style="list-style-type: none"> • A1. Definir del problema de investigación. • A2. Preparar los materiales y métodos.
PT de entrada	PT1. Recursos objeto de debate
PT de salida	PTF1. Recursos de planeamiento

La F2. Definición de grupos de discusión (Tabla 23) es desarrollada por los integrantes del grupo investigador al igual que la fase anterior, se realizan dos actividades: A3. Seleccionar los participantes y A4. Segmentación de grupos. Para su ejecución no se requiere un PT de entrada pero sí se genera el PT13. “Reporte de participantes” que se refiere a la clasificación de los participantes por grupos de discusión.

Tabla 23. Descripción de la Fase Definición de grupos de discusión.
Adaptado de (Mendoza-Moreno et al., 2013).

F2. Definición de grupos de discusión	
Identificar los participantes y constituir los grupos de debate.	
Rol Involucrado	Grupo investigador
Actividades	<ul style="list-style-type: none"> • A3. Seleccionar los participantes. • A4. Segmentación de grupos.
PT de entrada	<i>No se requiere</i>
PT de salida	PT13. Reporte de participantes

La F3. Conducción de la sesión de debate (Tabla 24) es coordinada por el moderador, sintetizada por el relator y desarrollada por los participantes, se realizan dos actividades: A5. Secuencia básica y A6. Capturar información. Para su ejecución se requieren los productos generados por la F2, el PT13. “Reporte de participantes” y la F1, los PTF1. “Recursos de planeamiento” y se genera el producto PT15. “Memorias del debate”.

Tabla 24. Descripción de la Fase Conducción de la sesión de debate.
Adaptado de (Mendoza-Moreno et al., 2013).

F3. Conducción de la sesión de debate	
Ejecuta los procedimientos establecidos en la Fase 1.	
Rol Involucrado	Coordinada por: Moderador Sintetizada por: Relator Desarrollada por: Participantes
Actividades	<ul style="list-style-type: none"> • A5. Secuencia básica. • A6. Capturar información.
PT de entrada	PT13. Reporte de participantes PTF1. Recursos de planeamiento
PT de salida	PT15. Memorias del debate

Por último, la F4. Análisis de información y reporte de resultados (Tabla 25), al igual que las F1 y F2, es desarrollada por los integrantes del grupo investigador, se realizan dos actividades: A7. Análisis de información y A8. Generar reporte de resultados. Para su ejecución se requieren los

PT de salida generados por la F3, el PT15. “Memorias del debate” y la F1, los PTF1. “Recursos de planeamiento” y se genera el producto PT16. “Análisis de resultados” que constituye una recopilación de las conclusiones y características del debate cumplido.

Tabla 25. Descripción de la Fase Análisis de información y reporte de resultados.
Adaptado de (Mendoza-Moreno et al., 2013).

F4. Análisis de información y reporte de resultados	
Obtener información y reporte de resultados.	
Rol Involucrado	Grupo investigador
Actividades	<ul style="list-style-type: none"> • A7. Análisis de información. • A8. Generar reporte de resultados.
PT de entrada	PT15. Memorias del debate PTF1. Recursos de planeamiento
PT de salida	PT16. Análisis de resultados

Con base en lo anterior, en la Figura 19 se puede observar las cuatro fases del proceso general de aplicación de Focus Group en IS a través de un diagrama de Rol – Actividad, el cual permite la identificación rápida de los roles involucrados y los PT necesarios para su aplicación satisfactoria.

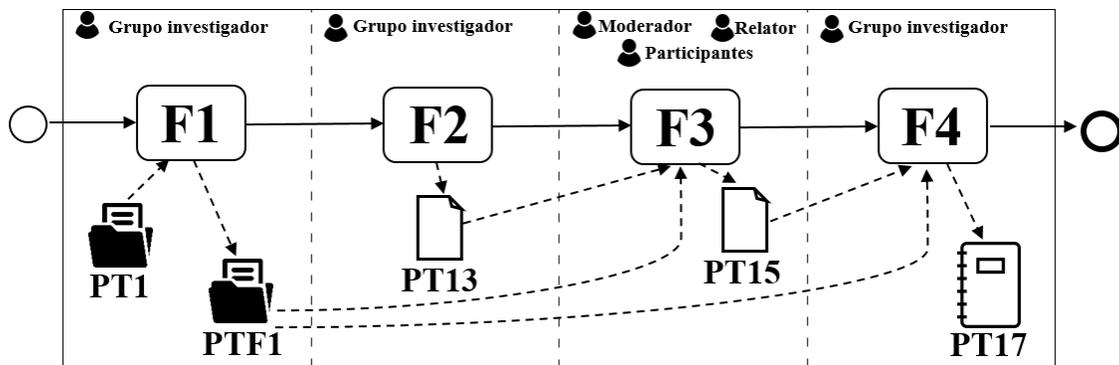


Figura 19. Diagrama de actividad para el proceso general de la aplicación de Focus Group.
(Mendoza-Moreno et al., 2013).

Siguiendo los lineamientos del proceso de Focus Group a un enfoque más detallado, en la Tabla 26 se muestra la relación de cada uno de los elementos mencionados (fases, actividades, tareas, PT de entrada y PT de salida) donde se puede observar la identificación específica de las tareas contenidas por las actividades, así como la totalidad de los PT generados por las tareas, esta misma información se representa en forma de un diagrama de actividad en la Figura 20.

Tabla 26. Relación de los elementos para el proceso de aplicación de Focus Group.

Fase	Actividad	Tareas	Producto de Trabajo (Entrada)	Producto de Trabajo (Salida)
F1	A1		PT1. Recursos objeto de debate	PT2. Reporte de contexto
	A2	T1. Definición de estructura	PT1. Recursos objeto de debate	PT3. Reporte de estructura
		T2. Definición de instrumentos		PT4. Instrumentos, materiales y métodos
		T3. Formalización de documentos		PT5. Compendio para los participantes
		T4. Definición métodos: captura y registro de información		PT6. Reporte de métodos de captura y registro
		T5. Definición métodos: Análisis de información		PT7. Reporte de métodos de análisis
T6. Definición estrategias publicación	PT8. Reporte de estrategias de publicación			
F2	A3	T7. Definición perfil de los participantes		PT9. Reporte perfil de participantes
		T8. Identificación potenciales participantes	PT9. Reporte perfil de participantes	PT10. Reporte potenciales participantes
		T9. Conformación grupo de participantes	PT10. Reporte potenciales participantes	PT11. Reporte selección de participantes
	A4	T10. Definición criterios de agrupamiento		PT12. Compilado criterios de agrupamiento
		T11. Aplicación criterios agrupamiento	PT12. Compilado criterios de agrupamiento PT11. Reporte selección de participantes	PT13. Reporte de participantes
F3	A5	T12. Ejecución del protocolo	PT3. Reporte de estructura PT13. Reporte de participantes	
		T13. Aplicación de técnicas, materiales y métodos		PT14. Conformidad del debate cumplido
	A6			PT15. Memorias del debate
F4	A7		PT15. Memorias del debate PTF1. Recursos de planeamiento	PT16. Métodos de análisis ejecutados
	A8		PTF1. Recursos de planeamiento	PT17. Reporte de análisis de resultados

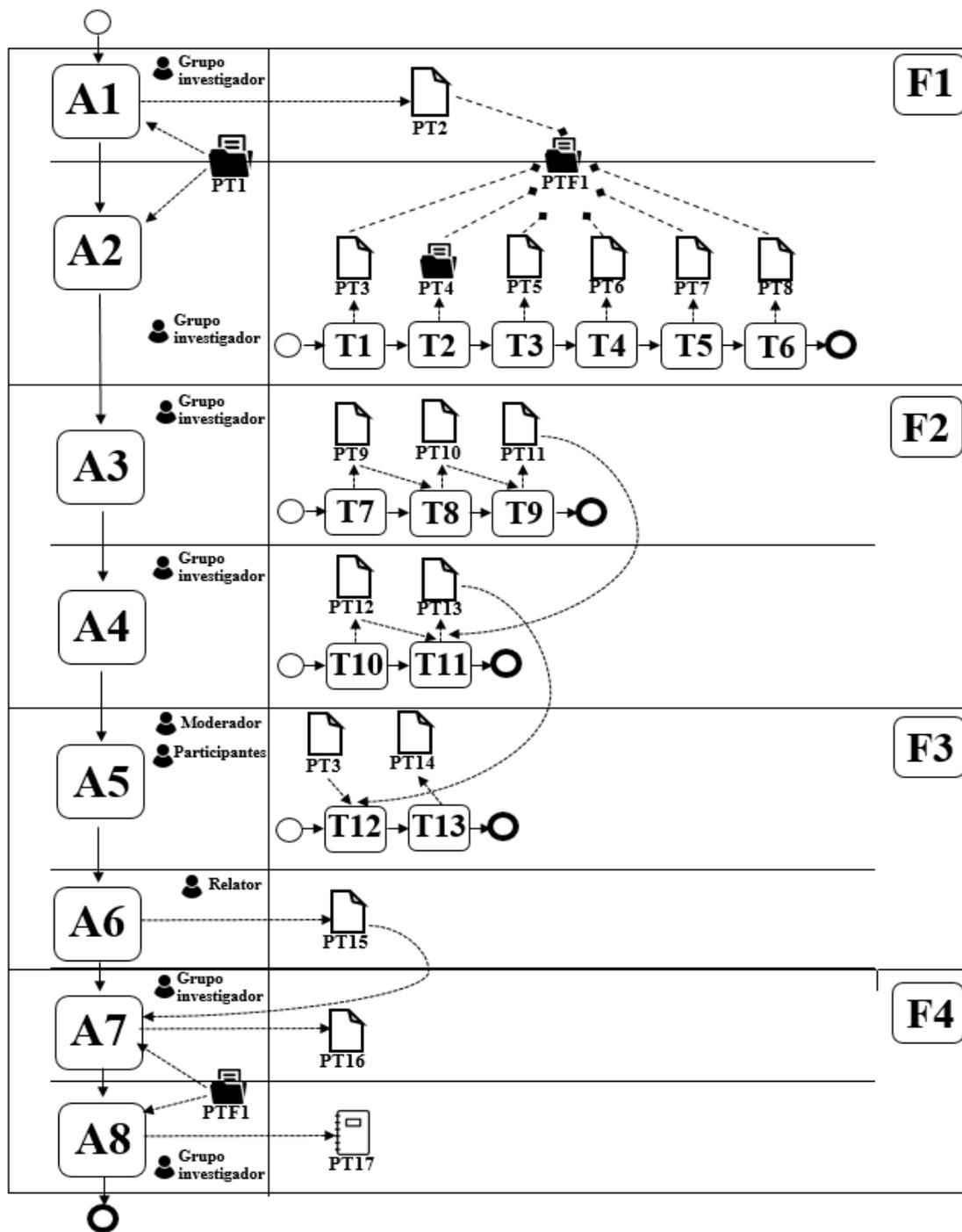


Figura 20. Diagrama de actividad para el proceso de la aplicación de Focus Group. Adaptado de (Mendoza-Moreno et al., 2013).

En la siguiente sección (5.1.2) se muestra la ejecución del proceso propuesto para Focus Group en donde los participantes evalúan el Marco de Trabajo para la identificación y selección de algoritmos y herramientas de MP con los elementos que lo conforman: MoProPEI-MP y la Guía Técnica.

5.1.2 Ejecución de la técnica Focus Group para la evaluación del Marco de Trabajo

La ejecución del proceso de Focus Group en la evaluación del Marco de Trabajo para la identificación y selección de algoritmos y herramientas de MP se realizó conforme a la estructura descrita en la sección anterior. En seguida, se presenta en detalle cada una de las fases realizadas.

5.1.2.1 F1. Planteamiento de la investigación

Como base para la F1 se requiere el PT1. “Recursos objeto de debate” por lo que se empleó un documento de síntesis de la propuesta para poder seguir con el cumplimiento de las siguientes actividades:

- **A1. Definición del problema de investigación:** La actividad derivó el PT2. “Reporte de contexto” (detallado en la en la Tabla 27) donde se especificó objetivos generales y específicos de la evaluación y alcances que se tendrían con la aplicación de Focus Group.

Tabla 27. Producto Reporte de contexto.

Reporte de contexto	
Objetivo General	Evaluar el Marco de Trabajo para la identificación y selección de algoritmos y herramientas de MP de acuerdo con el grado de aceptación o rechazo por parte de los participantes, quienes son expertos en el área de aplicación del Marco de Trabajo.
Objetivo Específicos	<ul style="list-style-type: none"> • Presentar la estructura general del Marco de Trabajo propuesto. • Conocer la viabilidad del modelo de procesos MoProPEI-MP. • Determinar la puntualización correcta del contenido de la Guía Técnica. • Verificar la aceptación del Marco de Trabajo por parte de la industria. • Refinar los elementos del Marco de Trabajo a partir de las observaciones obtenidas.
Alcances	Obtener retroalimentación cualitativa de alto nivel por expertos del área tomando los puntos de vista definidos como elementos fundamentales para la validación del Marco de Trabajo para la identificación y selección de algoritmos y herramientas de MP.

- **A2. Preparación de materiales y métodos:** La actividad derivó los PTF1. “Recursos de planeamiento”, que como se mencionó antes, es una composición de subproductos de la F1, en estos subproductos se encuentra el PT3. “Reporte de estructura” (Tabla 28) donde

se explica la estructura de Focus Group con los campos tema, duración y lugar de la sesión, fecha y hora, así las personas que desempeñaran los roles de Moderador y Relator.

Tabla 28. Estructura del Protocolo de Focus Group

Protocolo de Focus Group	
Tema	Evaluación del Marco de Trabajo para la identificación y selección de algoritmos y herramientas de Minería de Prcoesos.
Duración de la sesión	1 hora, 30 minutos
Lugar de la sesión	Presencial o virtual según la necesidad del participante.
Fecha y Hora	Según la disponibilidad de los participantes.
Moderador	Brenda Leticia Flores Ríos
Relator	Silvia Jaqueline Urrea Contreras

En el PT4. “Instrumentos, materiales y métodos” descrito en la Tabla 29 se muestra la metodología mixta (Hernández-Sampieri et al., 2014) que se siguió para el cumplimiento de la obtención de datos cualitativos y cuantitativos. Por otra parte, se especifica el PT5. “Compendio para los participantes” que corresponde a un documento que contiene una presentación describiendo los elementos del Marco de Trabajo.

El PT6. “Reporte de métodos de captura y registro” muestra a la entrevista dirigida semiestructurada y al cuestionario como métodos de recolección de datos cualitativos y cuantitativos respectivamente que permiten una posibilidad de que sus resultados puedan ser cualificados o cuantificados de igual manera; también el método de Focus Group permite esta posibilidad (Hernández-Sampieri et al., 2014), la entrevista dirigida semiestructurada visualiza los aspectos positivos, observaciones y aspectos a mejorar del Marco de Trabajo, así como el cuestionario se encarga de los aspectos generales.

El análisis de los resultados se representa en el PT7. “Reporte de métodos de análisis”, en el cual se utilizó el procedimiento de explicar resultados (profundizar) ya que se conducen a entrevistas y cuestionarios para explorar las diferencias o similitudes que se encuentren entre los resultados obtenidos. Por último, el PT8. “Reporte de estrategias de publicación” induce a la representación de las inferencias, comentarios y conclusiones obtenidas de los resultados, la estrategia de publicación que presenta en la sección (5.2) exterioriza primero los resultados cualitativos, posteriormente los resultados cuantitativos para finalizar los resultados de ambos métodos.

Tabla 29. Metodología Mixta: Diseño exploratorio secuencial (DEXPLOS).

Diseño EXPLORatorio Secuencial (DEXPLOS)	
Definición	La fase inicial del diseño DEXPLOS implica la recolección y análisis de datos cualitativos seguida de los datos cuantitativos (Hernández-Sampieri y Mendoza, 2008; Creswell et al., 2008)
Modalidades del Diseño	<ul style="list-style-type: none"> ○ Derivativa: la recolección y el análisis de los datos cuantitativos se hacen sobre la base de los resultados cualitativos. ○ Comparativa: la recolección y el análisis de los datos cualitativos, y de los datos cuantitativos. Los hallazgos descubiertos de ambas etapas se comparan e integran en la interpretación y elaboración del reporte de resultados
Esquema del Diseño Exploratorio Secuencial (DEXPLOS)	
<pre> graph LR A[Recolección de datos cualitativos] --> B[Análisis cualitativo] B --> C[Recolección de datos cuantitativos] C --> D[Análisis cuantitativo] D --> E[Interpretación del análisis completo] </pre>	
Justificación	<ul style="list-style-type: none"> ○ Los datos cualitativos y cuantitativos tienen la misma prioridad. ○ Los datos se recopilan de manera secuencial (un tipo de datos primero y luego otro) por lo que se toma la modalidad comparativa. ○ Se trabaja con una metodología mixta para la transformación de datos, cualificar datos cuantitativos y cuantificar datos cualitativos. ○ La metodología mixta se integra al proceso de Focus Group a partir de la F3.

5.1.2.2 F2. Definición de grupos de discusión

La F2 no requiere un producto para poder ser ejecutada, en esta fase se obtiene el PT13. “Reporte de participantes” que se convierte en producto de entrada para la F3, por lo que se empleó directamente al cumplimiento de las siguientes actividades:

- **A3. Selección de los participantes:** La actividad derivó el PT9. “Reporte perfil de participantes” donde se adoptaron los siguientes criterios para formar el grupo de participantes: 1) Formación académica relacionada con la IS, 2) Adscripción a una institución académica o a empresas de desarrollo de software y 3) Experiencia en IS y/o

en SPI. De esta manera se identificaron a seis expertos (PT10. “Reporte potenciales participantes”) que posteriormente fueron seleccionados y registrados bajo un código del participante en el PT11. “Reporte de selección de participantes” que se encuentra representado en la Tabla 30. Para cada participante se realizó una ficha donde se caracteriza el perfil del participante (Tabla 31).

Tabla 30. Reporte de selección de participantes.

Código del participante	Nombre	País	Institución/Empresa
P1	Angélica Astorga	México	UABC
P2	Alejandro Quiñones	México	Monobits
P3	Isaac García	México	Monobits
P4	Karla Aguilar	México	Consultor Privado
P5	Sandra Buitrón	Colombia	UNICAUCA
P6	Carlos Ardila	Colombia	UNICAUCA

Tabla 31. Caracterización de los participantes.

Ficha de participante	
Perfil de Formación Académica	1. Nivel de licenciatura Ingeniería en Computación (Trunca) Ingeniería en Computación Ingeniería de Sistemas Licenciado en Sistemas Computacionales 2. Nivel de posgrado Maestría en Ciencias de la Computación Magister en Computación Maestría en Administración
Academia/ Empresa	Organización educativa e Iniciativa privada
Actividad que desempeña	Mejora de Procesos Analista de Datos Desarrollador de Software Docente de tiempo completo Consultor Privado
Experiencia en el área de IS	Scrum Master- Metodologías Ágiles Evaluador de CMMI Mejora de Procesos de Software Pruebas de Software Calidad de Producto y de Proceso

- **A4. Segmentación de grupos:** La actividad derivó el PT13. “Reporte de participantes” donde se refleja la segmentación de dos subgrupos (Tabla 32), el primer subgrupo

corresponde a las personas que serán participantes presenciales mientras que el segundo subgrupo son los participantes de manera virtual.

Tabla 32. Reporte de participantes segmentados en grupos.

Tipo de sesión	Código del participante
Presencial	P1, P2, P3 y P4
Virtual	P5 y P6

5.1.2.3 F3. Conducción de la sesión de debate

Para la F3 se requiere el PT13. “Reporte de participantes” obtenido de la F2 y los PTF1. “Recursos de planeamiento” resultado de la F1. Para poder obtener su producto de salida PT15. “Memorias del debate” se dio cumplimiento de las siguientes actividades:

- **A5. Secuencia básica:** La actividad derivó el PT14. “Conformidad del debate cumplido” donde se puntualiza los lineamientos para dar conformidad del cumplimiento satisfactorio del protocolo de la sesión de Focus Group, dichos lineamientos se ven representados en la Tabla 33.

Tabla 33. Lineamientos de Conformidad del debate cumplido.

Lineamientos de Conformidad		
Lineamientos	Opciones	
	Si	No
1. ¿La sesión se llevó a cabo en la duración de tiempo establecido?	X	
2. ¿La sesión se llevó a cabo conforme a la modalidad de la segmentación de los participantes?	X	
3. ¿La sesión se llevó a cabo en las fechas y horas establecidas dependiendo la modalidad del grupo?	X	
4. ¿La sesión fue interrumpida?		X
5. ¿Se realizaron todas las preguntas de la entrevista?	X	
6. ¿Todos los cuestionarios fueron contestados?	X	
7. ¿El rol de moderador fue cumplido conforme a su especificación?	X	
8. ¿El rol de relator fue cumplido conforme a su especificación?	X	
9. ¿Fue recopilada toda la información necesaria?	X	
10. ¿Se cumplieron los objetivos específicos planteados?	X	

- **A6. Captura de información:** La actividad derivó el PT15. “Memorias del debate” que consta de la información obtenida por la capturada de información de aspectos positivos, observaciones y aspectos a mejorar, así como un cuestionario para evaluar aspectos generales por parte de la metodología mixta DEXPLOS.

5.1.2.4 F4. Análisis de información y reporte de resultados

Para la F4 se requiere el PT15. “Memorias del debate” obtenido de la F3 y los PTF1. “Recursos de planeamiento” resultado de la F1. Para poder obtener su producto de salida PT17 “Análisis de resultados” se dio cumplimiento de las siguientes actividades:

- **A7. Análisis de información:** La actividad derivó el PT16. “Métodos de análisis ejecutados” donde se dio seguimiento al análisis del PT15. “Memorias del debate” y los PTF1. “Recursos de planeamiento” en el cual se obtuvo la recolección de información por medio de ejecución de la metodología mixta DEXPLOS.
- **A8. Reporte de resultados:** La actividad derivó el PT17. “Análisis de resultados” donde fueron establecidos los formatos de representación de los resultados obtenidos de los datos cualitativos, cuantitativos y mixtos.

En la siguiente subsección (5.1.3) se muestra la información obtenida de las actividades A7 y A8 de la ejecución de la sesión de Focus Group en donde se determina la evaluación del Marco de Trabajo para la identificación y selección de algoritmos y herramientas de MP con base en un análisis cualitativo y cuantitativo (análisis mixto).

5.2 Análisis y Reporte de Resultados Cualitativos, Cuantitativos y Mixtos

El implementar un método mixto es utilizar las fortalezas del enfoque cualitativo y cuantitativo, combinándolos y tratando de minimizar sus debilidades potenciales (Hernández-Sampieri et al., 2014). En este proyecto el análisis y reporte de resultados se complementan con los resultados cualitativos y cuantitativos obtenidos de las técnicas de entrevista y cuestionario respectivamente, que al utilizarlas de forma conjunta generaron resultados mixtos permitiendo conocer, de una mejor forma, la opinión de los expertos hacia la evaluación del Marco de Trabajo.

Conforme al esquema de DEXPLOS (Tabla 29), primero se presentará el análisis cualitativo, posteriormente el análisis cuantitativo y por último, el análisis mixto para poder presentar los resultados.

5.2.1 Análisis Cualitativo

Para el enfoque cualitativo la recolección de datos resulta fundamental. Lo que se busca es obtener datos (que se convertirán en información) de procesos en las propias “formas de expresión” de cada uno. Al tratarse de información cualitativa, los datos que interesan son conceptos, percepciones y experiencias de los participantes, ya sea de manera individual, grupal o colectiva. Se recolectan con la finalidad de analizarlos y comprenderlos, y así generar conocimiento (Hernández-Sampieri et al., 2014).

Como método de recolección de datos cualitativos se utilizó la entrevista semiestructurada, la cual se basa en una guía de preguntas y el entrevistador tiene la libertad de introducir preguntas adicionales para precisar conceptos u obtener mayor información. (Hernández-Sampieri et al., 2014). El establecimiento de la entrevista semiestructurada para los resultados cualitativos propicia conocer la perspectiva del experto a través de la comunicación y la construcción de los datos para convertirlos en información.

En la Tabla 34 se presentan las preguntas aplicadas a los participantes de la sesión de Focus Group, con el fin de conocer la perspectiva de los mismos hacia el Marco de Trabajo para la identificación y selección de algoritmos y herramientas de MP.

Una vez aplicada la entrevista, se manejaron los datos recopilados en las categorías aspectos positivos, observaciones y aspectos a mejorar requeridos en la elaboración del reporte de resultados cualitativos del Marco de Trabajo.

Tabla 34. Preguntas de la Entrevista Semiestructurada (Análisis Cualitativo).

Preguntas Entrevista Semiestructurada	
1.	¿Cuál es su opinión acerca del Marco de Trabajo propuesto y su necesidad para las empresas dedicadas al desarrollo de software?
2.	¿Cómo visualiza usted el esquema de los componentes del Marco de Trabajo propuesto y los componentes que posee?
3.	¿Cómo visualiza usted que las actividades de MoProPEI-MP así como los objetivos de cada una de ellas sean implementadas por las empresas dedicadas al desarrollo de software?
4.	¿Cómo observa el cumplimiento establecido del flujo de trabajo fijo de las tareas de MoProPEI-MP?
5.	¿El establecimiento de Productos de Trabajo propios de MoProPEI-MP que ventajas/desventajas podrían originar?
6.	¿Qué tan útil es la descripción así como la cantidad de herramientas identificadas de Minería de Procesos para las empresas dedicadas al desarrollo de software?
7.	¿Cuál es la oportunidad en que la descripción de los algoritmos de Minería de Procesos pueda ser requerida para autoevaluaciones en los procesos de desarrollo de software y/o proyectos de mejora de procesos de software?
8.	¿Para manejar el contenido de la Guía Técnica, considera usted que se deberá poseer conocimientos del tema o puede ser considerado como una oportunidad para aprender de ellos?

Aspectos Positivos

1. El Marco de Trabajo no representa en estos momentos una utilidad fundamental pero su implementación puede agregar valor a futuro.
2. Puede ayudar a la maduración de los procesos de una empresa.
3. El Marco de Trabajo propuesto es homogéneo
4. Las actividades y objetivos de Marco de Trabajo se quedan en la fase de planeación a la cual pueden agregar valor.
5. Para realizar las actividades e implementar sus objetivos se requiere una fuerte capacitación sobre MP, puede costar trabajo al principio pero los beneficios pueden ser considerables.
6. A medida que se usen y refinen los PT de MoProPEI-MP representaran una ventaja.
7. Las organizaciones de software podrán empezar a ampliar y especializar los PT para sus dominios.

8. La descripción de los algoritmos puede servir como base formal, de modo que se pueda determinar de manera objetiva si los procesos se ajustan a lo deseado por cada empresa.
9. Los algoritmos tienen alta oportunidad para ser requeridos en los procesos de análisis.

Observaciones/ Hallazgos

1. Detallar de manera más precisa el esquema de los componentes.
2. Contextualizar el uso de los componentes del Marco de Trabajo mediante un ejemplo.
3. Las líneas punteadas del esquema de los componentes del Marco de Trabajo hacen ruido al momento de su visualización y entendimiento.
4. Eliminar las líneas que conectan los elementos de la Guía Técnica para evitar interpretaciones de un flujo definido de consulta.
5. Establecer un tipo de ciclo entre tareas para poder alimentar al problema inicial.
6. Implementar el flujo de trabajo con procesos claramente no definidos puede impactar en el nivel de calidad de los datos.
7. Unificar las tareas de identificación y selección tanto de los algoritmos como de las herramientas para mayor rapidez.
8. Representar las primeras 3 herramientas ya que son las de mayor documentación significativa y las 4 siguientes documentarlas en un tipo anexo.
9. Determinar qué mecanismos se utilizarán para la actualización de la Guía Técnica.
10. Explicar los algoritmos con base a sus características, manejar como un concepto de “Caja negra” para sintetizar información.
11. Replantearse la Guía Técnica al ofrecer ejemplos de uso y no solo un compendio sobre algoritmos y herramientas.
12. Incluir un breve ejemplo del flujo de MoProPEI-MP con ayuda de la Guía Técnica.
13. Explicar de manera más liviana la contextualización del problema y el valor de la MP para las empresas de software.

Aspectos a Mejorar

1. Reestructurar la información del Marco de Trabajo para no confundir MoProPEI-MP con el modelo de procesos del cual está basado.
2. Mejorar el esquema de los componentes tanto en el flujo de las líneas como su estilo para un entendimiento más conciso de los componentes del Marco de Trabajo.
3. Establecer perfiles claros para realizar la implementación del Marco de Trabajo.

4. Establecer precisiones de acuerdo con el flujo de trabajo para indicar cuando se complete una tarea y poder avanzar a la siguiente.
5. Agregar puntos de control al cumplimiento de las tareas.
6. Representar el flujo del Marco de Trabajo a través de un ejemplo de implementación.

5.2.2 Análisis Cuantitativo

Para el enfoque cuantitativo la recolección de datos se fundamenta en la medición, se lleva a cabo al utilizar procedimientos estandarizados y aceptados. Debido a que los datos son producto de mediciones, se presentan mediante números (cantidades) y se deben analizar con métodos estadísticos (Hernández-Sampieri et al., 2014).

Como método de recolección de datos cualitativos se utilizó el cuestionario de preguntas cerradas, el cual consiste en un conjunto de preguntas que contienen categorías u opciones de respuesta que han sido previamente delimitadas. Es decir, se presentan las posibilidades de respuesta a los participantes, quienes deben acotarse a éstas (Hernández-Sampieri et al., 2014). El establecimiento de un cuestionario de preguntas cerradas para el enfoque cuantitativo propicia una fácil codificación y preparación del análisis de los resultados.

En la Tabla 35 se presentan las preguntas cerradas aplicadas a los participantes de la sesión de Focus Group retomadas del trabajo de Ardila (2013), así como los resultados obtenidos en función de porcentaje con el fin de establecer un análisis cuantitativo del Marco de Trabajo. Del análisis de los resultados de los datos cuantitativos se puede evidenciar que:

- a. El 100% de los participantes consideran al Marco de Trabajo como necesario para las empresas dedicadas al desarrollo de software, de igual forma las actividades propuestas de MoProPEI-MP, la cantidad de herramientas identificadas de MP así como su implementación en proyectos de mejora de procesos. Esto último también involucrando los algoritmos de MP.
- b. La opinión de los participantes estuvo dividida con respecto al esquema de los componentes así como la falta de elementos del Marco de Trabajo.
- c. Las demás preguntas fueron respondidas a diferente porcentaje pero siempre teniendo el mayor aprobatorio a lo positivo.

Tabla 35. Cuestionario de Evaluación de Aspectos Generales (Análisis Cuantitativo).

Preguntas	Opciones		
	Sí	No	NS/NR
1. ¿Considera que el Marco de Trabajo propuesto es necesario para las empresas dedicadas al desarrollo de software?	100%		
2. ¿Considera que el esquema de los componentes del Marco de Trabajo propuesto para la Minería de Procesos se muestran de manera clara y detallada?	50%	50%	
3. ¿Considera que las actividades propuestas en MoProPEI-MP son apropiadas para las empresas dedicadas al desarrollo de software?	100%		
4. ¿Considera que los objetivos de las actividades de MoProPEI-MP se muestran de forma clara y concisa?	66.7%	33.3%	
5. ¿Considera que la cantidad propuesta de tareas es suficiente para abordar las actividades en MoProPEI-MP?	66.7%	33.3%	
6. ¿Considera que los Productos de Trabajo son requeridos por la tarea en cada actividad identificada en MoProPEI-MP?	100%		
7. ¿Considera que los Productos de Trabajo generados (de salida) son adecuados para las empresas dedicadas al desarrollo de software?	83.3%	16.7%	
8. ¿Considera que la cantidad de herramientas identificadas son suficientes para que las empresas dedicadas al desarrollo de software aborden las actividades de descubrimiento de procesos, verificación de conformidad y mejora de procesos?	100%		
9. ¿Considera que las herramientas de Minería de Procesos son parte de los requerimientos para que las empresas dedicadas al desarrollo de software implementen proyectos de mejora de proceso?	100%		
10. ¿Considera que la descripción de los algoritmos son requeridos para las evaluaciones internas en los procesos de desarrollo de software?	83.3%	16.7%	
11. ¿Considera que las descripciones de definiciones, tipos de minería y perspectivas de la Minería de Procesos presentadas en la Guía Técnica muestran de forma clara una auxiliar de información para la comprensión del Marco de Trabajo?	66.7%	33.3%	
12. ¿Considera que utilizar herramientas y algoritmos existentes y probados en Minería de Procesos, les permita a las pequeñas organizaciones dedicadas al desarrollo de software identificar y verificar la evidencia en la ejecución de las actividades o de Productos de Trabajo involucrados en los proyectos de mejora?	100%		
13. ¿Considera que el Marco de Trabajo propuesto le hacen falta elementos?	50%	50%	
14. ¿Considera que hay debilidades en el Marco de Trabajo propuesto?	33.3%	66.7%	
15. ¿De acuerdo con su experiencia, considera que el Marco de Trabajo propuesto podría ser de fácil aplicación?	66.7%		33.3%

Para obtener una representación de porcentajes destinada a conocer la aprobación de los participantes, se promediaron los resultados del cuestionario, clasificando las preguntas en cinco categorías: 1) Marco de Trabajo, 2) MoProPEI-MP, 3) Guía Técnica, 4) Herramientas y 5) Algoritmos, donde cada categoría se muestra bajo las etiquetas: Aprueban, Desaprueban y NS/NR (No sé/ No respondió). A continuación en la Tabla 36 se muestra la clasificación por categoría de las preguntas del cuestionario aplicado.

Tabla 36. Clasificación de las preguntas del cuestionario por categoría.

Número de pregunta (Cuestionario)	Categoría
1, 2, 13, 14 y 15	Marco de Trabajo
3, 4, 5, 6 y 7	MoProPEI-MP
11	Guía Técnica
8, 9 y 12	Herramientas
10 y 12	Algoritmos

Los resultados de los promedios se pueden observar en las Tablas 37 – 41 en las que se representan los valores obtenidos (porcentaje) de cada pregunta y el promedio obtenido por cada etiqueta. Los resultados de los promedios se representaron por medio de graficas circulares en la Figura 21.

Tabla 37. Valoración del Marco de Trabajo.

MARCO DE TRABAJO			
# Pregunta	Aprueban	Desaprueban	NS/NR
Pregunta 1	100	0	0
Pregunta 2	50	50	0
Pregunta 13	50	50	0
Pregunta 14 ⁵	66.7	33.3	0
Pregunta 15	66.7	0	33.3
Promedio	66.68%	26.66%	6.66%

⁵ Los resultados de la pregunta 14 se invirtieron por su connotación negativa.

Tabla 38. Valoración de MoProPEI-MP.

MoProPEI-MP			
# Pregunta	Aprueban	Desaprueban	NS/NR
Pregunta 3	100	0	0
Pregunta 4	66.7	33.3	0
Pregunta 5	66.7	33.3	0
Pregunta 6	100	0	0
Pregunta 7	83.3	16.7	0
Promedio	83.34%	16.66%	0

Tabla 39. Valoración de la Guía Técnica.

GUÍA TÉCNICA			
# Pregunta	Aprueban	Desaprueban	NS/NR
Pregunta 11	66.7	33.3	0
Promedio	66.7%	33.3%	0

Tabla 40. Valoración de las herramientas de MP.

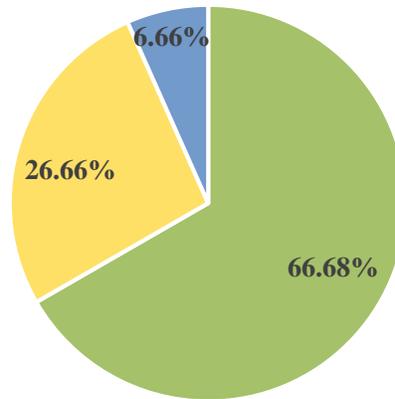
HERRAMIENTAS DE MP			
# Pregunta	Aprueban	Desaprueban	NS/NR
Pregunta 8	100	0	0
Pregunta 9	100	0	0
Pregunta 12	100	0	0
Promedio	100%	0	0

Tabla 41. Valoración de los algoritmos de MP.

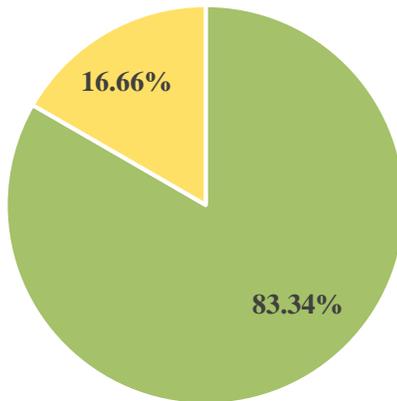
ALGORITMOS DE MP			
# Pregunta	Aprueban	Desaprueban	NS/NR
Pregunta 10	83.3	16.7	0
Pregunta 12	100	0	0
Promedio	91.65%	8.35%	0

■ APRUEBAN ■ DESAPRUEBAN ■ NS/NR

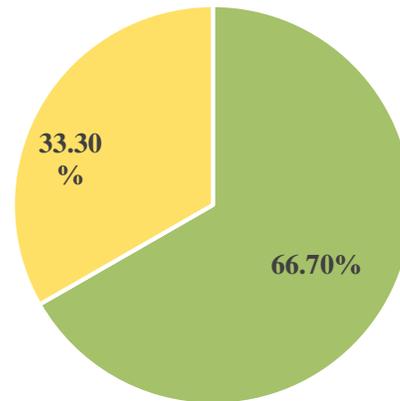
MARCO DE TRABAJO



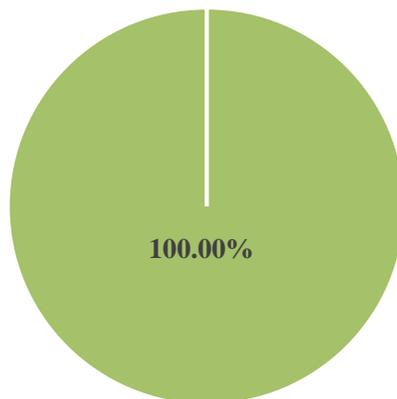
MoProPEI-MP



Guía Técnica



Herramientas



Algoritmos

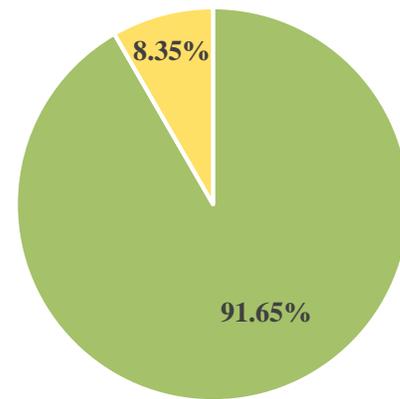


Figura 21. Resultados de las categorías en porcentaje.

5.2.3 Análisis Mixto

Los métodos mixtos representan un conjunto de procesos sistemáticos, empíricos y críticos de investigación e implican la recolección y el análisis de datos cuantitativos y cualitativos, así como su integración y discusión conjunta, para realizar inferencias producto de toda la información recabada y lograr un mayor entendimiento (Hernández-Sampieri et al., 2008).

Con el diseño DEXPLOS primero se recabaron y analizaron los datos cualitativos (subsección 5.2.1), posteriormente los datos cuantitativos (subsección 5.2.2) para comparar e integrar la interpretación de resultados. Esto permite que los resultados de la entrevista semiestructurada propicien conocer el porqué de los resultados cuantitativos logrando un descubrimiento más detallado de las distintas perspectivas que los participantes poseían.

En la Tabla 42 se presentan la trazabilidad de las preguntas aplicadas tanto en la entrevista como en el cuestionario, junto con la categoría identificada a la que pertenecen para originar el análisis mixto. Dichas categorías representan un ámbito del Marco de Trabajo que se sometió a apreciación; conocer la visión del participante hacia cada categoría es necesario para su evaluación.

Tabla 42. Trazabilidad de los instrumentos cualitativos y cuantitativos para el análisis mixto.

Número de pregunta (Entrevista)	Número de pregunta (Cuestionario)	Categoría	
1	1, 13, 14 y 15	Marco de Trabajo	
2	2	Esquema de Componentes	
3	3 y 4	Actividades y Objetivos	MoProPEI-MP
4	5	Tareas y Flujo de Trabajo	
5	6 y 7	Productos de Trabajo	
6	8, 9 y 12	Herramientas	
7	10 y 12	Algoritmos	
8	11	Guía Técnica	

Marco de Trabajo

El Marco de Trabajo fue aprobado por todos los participantes como necesario para las empresas dedicadas al desarrollo de software dentro del cuestionario, durante las entrevistas se identificó que su uso en estos momentos no es vital, pero al ser implementado puede llegar a agregar valor. Con respecto a la opinión de los participantes (50%) de la falta de componentes para el Marco de

Trabajo se debió a la falta de caracterización de los componentes a través de un ejemplo de caso de registro de eventos (eventos ficticios o reales) para su mayor comprensión técnica. Este punto también se vio relacionado a las respuestas del cuestionario correspondientes con las debilidades del Marco de Trabajo y la consideración de su fácil aplicación.

Esquema de Componentes

Con respecto a la claridad y forma detallada del esquema de los componentes, la opinión estuvo dividida (50%) debido a que en algunos participantes las líneas punteadas de los componentes le hacían un poco de ruido al momento de su interpretación, también la conexión del contenido de Guía Técnica se recomendó eliminarlas y representarlas de manera similar a como se representó MoProPEI-MP. Por último, se mencionó que el Marco de Trabajo se detecta como homogéneo (33.3%) y se sugiere demostrar una contextualización de su uso en términos de ejemplos.

Actividades y Objetivos

Todos los participantes detectaron que las actividades propuestas de MoProPEI-MP son apropiadas para las empresas dedicadas al desarrollo de software, en principio puede costar trabajo su aplicación pero los beneficios pueden ser considerables al concepto de mejora de procesos. El 33.3% de los participantes sugiere que para los objetivos se especifique marcar cuales objetivos van vinculados con que tareas y saber así su origen.

Tareas y Flujo de trabajo

Las tareas y su representación a través del flujo de trabajo propuesto obtuvieron un 66.7% de aprobación cuando los participantes detectaron la viabilidad de un flujo de trabajo fijo, aconsejando incluir puntos de control con nomenclatura de toma de decisión para verificar cuando una tarea cumple con éxito su PT y de esa forma poder pasar a su tarea posterior. Igualmente, se mencionó incluir un ciclo de la actividad 2 (A2) hacia la actividad 1 (A1) para poder alimentar el problema a resolver con la identificación de las características de los algoritmos de MP. Finalmente, también se analizó la mantenibilidad fija con respecto a la identificación y selección de las herramientas de MP y poder establecer un flujo alternativo desde la identificación y selección de los algoritmos de MP hacia el término del flujo de trabajo.

Productos de Trabajo

Los PT se consideraron requeridos por el 100% de los participantes en la tarea de cada actividad identificada ya que se reconoció que los PT una vez que se usen y refinen podrán convertirse en

un referente para las empresas de software puesto que también existirá la oportunidad de empezar a ampliarlos y especializarlos para los dominios en que se enfoquen. Por otro lado, un participante mencionó como desventaja la integración de los nuevos PT para las empresas que al principio pasarían por un periodo de adaptabilidad.

Herramientas

El 100% de los participantes estuvieron de acuerdo en considerar la utilización de las herramientas de MP como parte del proceso de identificar y verificar la evidencia en la ejecución de las actividades. De igual manera, que pueden llegar a tomarse en cuenta como parte de los requerimientos de las mismas y ser útiles para ejercicios de referenciación del tema de MP; con respecto a la cantidad de herramientas, concordaron que la cantidad identificada (7) es suficiente para las empresas de software al momento de abordar las actividades de descubrimiento de procesos, verificación de conformidad y mejora de procesos pero consideran que las primeras tres herramientas de MP se describan completamente y las restantes manejarlas como un tipo anexo dentro de la Guía Técnica.

Algoritmos

Asimismo que las herramientas de MP, todos los participantes (100%) consideraron que los algoritmos de MP pueden llegar a permitir la identificación y verificación de la evidencia generada de la ejecución de las actividades involucradas en los proyectos de mejora. De la misma manera, concordaron que las descripciones de los algoritmos de MP tienen alta probabilidad de ser requeridos en los procesos de análisis pero ajustando su descripción a mencionar características y no involucrar tanto la descripción del algoritmo como tal, sino manejar algo relacionado el concepto de “Caja negra”.

Guía Técnica

Con el contenido de la Guía Técnica, se recomendó estructurar su contenido para mejorar el sentido de su entendimiento ya que primero se presentaban las herramientas y posterior a ellas los algoritmos ocasionando una confusión al mostrar de forma clara la información. Se indicó que las guías técnicas van orientadas a personal con entrenamiento o formación en el área correspondiente por lo que representarla como una oportunidad para aprender de ella está marcada como fuera de contexto al no ser diseñadas como tutoriales. Tomando en cuenta lo anterior, también se expuso la duda de cuales mecanismos se utilizarán para mantener actualizado el contenido de la Guía Técnica.

5.3 Modificaciones realizadas a partir de la evaluación

A partir de las recomendaciones realizadas por los participantes en la sesión de Focus Group y la obtención del análisis de resultados, se procedió a modificar aquellos aspectos a mejorar que se identificaron; así como las observaciones indicadas por los participantes para beneficiar el Marco de Trabajo. En la Tabla 43 se señalan los ámbitos que se modificaron y los cambios que se integrarán a trabajo futuro como resultado del proceso de evaluación.

Tabla 43. Modificaciones y Trabajo Futuro con base a los resultados de la evaluación.

Categoría	Modificación/ Trabajo Futuro
Esquema de Componentes	<i>[MODIFICACIÓN]</i> Se eliminaron las líneas punteadas de los componentes del Marco de Trabajo y se incluyó los tipos de minería que se pueden realizar con la ayuda de la propuesta. El resultado se demuestra en la Figura 22.
Objetivos	<i>[TRABAJO FUTURO]</i> Se describirá la relación de los objetivos y las tareas con las cuales están directamente relacionadas.
Tareas	<i>[TRABAJO FUTURO]</i> Se analizará la propuesta de manejar en conjunto las tareas de Identificar y Seleccionar algoritmos así como herramientas de MP para mayor rapidez del cumplimiento del flujo de trabajo.
Flujo de Trabajo	<i>[MODIFICACIÓN]</i> Se tomó la recomendación de eliminar el flujo de trabajo fijo para dar origen a un flujo que incluye un punto de control entre la T1.1 y T1.2 para completar exitosamente el establecimiento del problema a resolver. También se optó poner una toma de decisión entre la T2.2 y la T2.3 para poder realizar un flujo alterno que solamente contenga la Identificación y Selección de algoritmo(s) de MP en caso de que ya se cuente con una herramienta a utilizar. El flujo de trabajo modificado de presenta en la Figura 23. <i>[TRABAJO FUTURO]</i> Integrar un flujo que permita alimentar el problema a resolver desde la A2 a la A1.
Productos de Trabajo	<i>[TRABAJO FUTURO]</i> Establecer plantillas de los PT para asegurar el cumplimiento de las tareas con las que interactúa.
Herramientas	<i>[TRABAJO FUTURO]</i> Actualizar la descripción y analizar la posibilidad de integrar o excluir herramientas de MP. Definir una priorización para manejar algunas herramientas como un anexo de la Guía Técnica.
Algoritmos	<i>[TRABAJO FUTURO]</i> Actualizar la descripción de los algoritmos de MP y manejar un contenido de características identificadas de cada uno para un mejor análisis. Definir una priorización de algoritmos de MP por facilidad de uso.
Guía Técnica	<i>[MODIFICACIÓN]</i> Se reestructuró el contenido de la Guía Técnica presentando primero los algoritmos y posteriormente las herramientas de MP, la modificación se puede observar en el Capítulo 4. <i>[TRABAJO FUTURO]</i> Establecer los mecanismos para mantener actualizada la Guía Técnica.

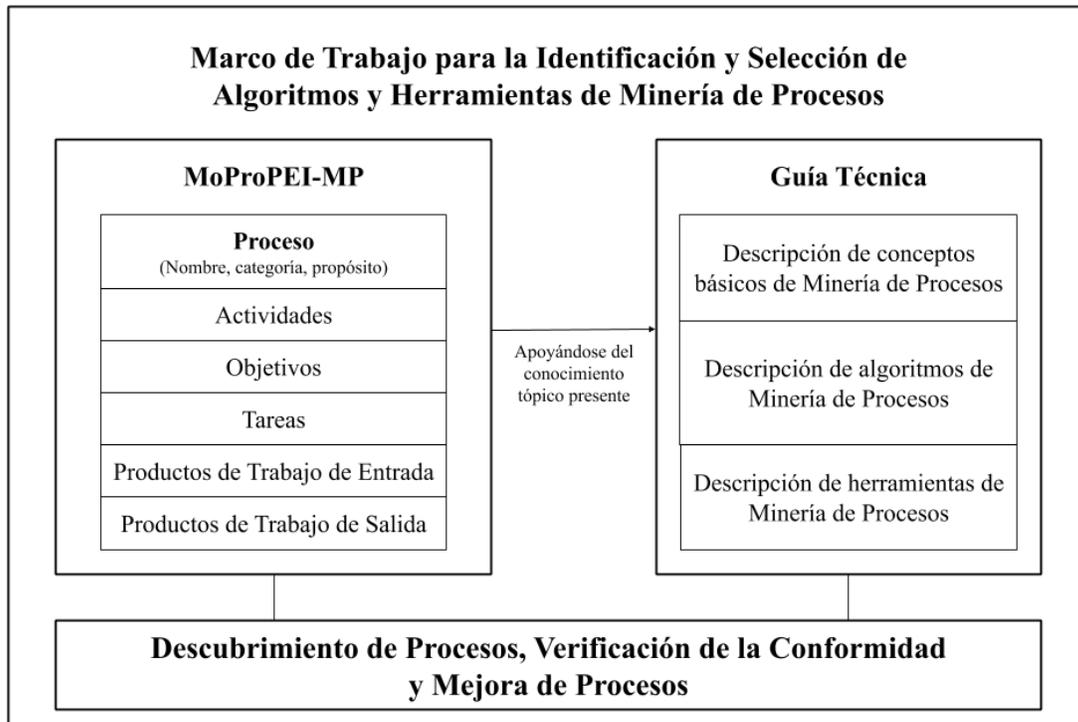


Figura 22. Esquema de los componentes del Marco de Trabajo para la Minería de Procesos (Modificado).

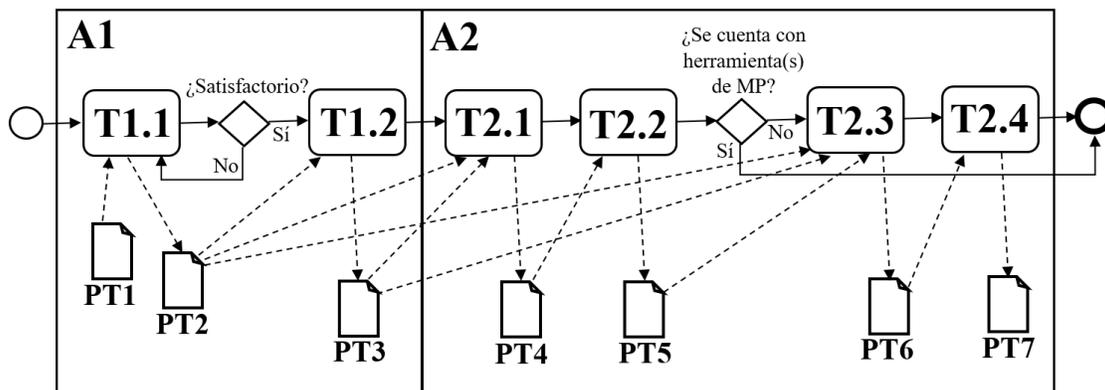


Figura 23. Flujo de trabajo de las tareas relacionadas a las actividades A1 y A2 para el proceso de Modelado de MoProPEI-MP (Modificado).

Capítulo 6. Conclusiones y Trabajo Futuro

6.1 Conclusiones

La MP al ser una disciplina joven (un poco menos de dos décadas) ha logrado grandes avances en la maduración de sus técnicas, algoritmos y herramientas como lo evidenció la revisión sistemática de literatura realizada. Por otro lado, de los casos de estudio encontrados en los 17 estudios primarios seleccionados de la revisión, se logró deducir que la identificación y selección de algoritmos y herramientas de MP se realizan de forma inductiva y en ocasiones sin fundamentos precisos de las características que poseen, estableciendo un desafío identificado que se abordó durante esta investigación.

Ante lo anterior, se presentó el Marco de Trabajo para la Identificación y Selección de Algoritmos y Herramientas de MP, el cual contempla como componentes un modelo de referencia MoProPEI-MP (Urrea-Contreras et al., 2018a) y una Guía Técnica. A partir del establecimiento de los componentes del Marco de Trabajo, dentro de los factores/criterios técnicos que se identificaron por parte de las herramientas de MP, el desarrollador/proveedor y el contacto inicial a través del sitio web para la adquisición de las herramientas tuvo más relevancia por parte de las comerciales debido a que el precio de venta se envía por correo según las necesidades del cliente, un punto que las empresas de software deben considerar relevante para la adquisición de las herramientas comerciales.

Por parte de la evaluación (validación empírica), el análisis de los datos cualitativos, cuantitativos y su trazabilidad a través de la metodología mixta (DEXPLOS) permitió conocer el conocimiento del porqué de los datos cuantitativos con base a los datos cualitativos. De los resultados obtenidos de dicha evaluación por parte de los expertos del Marco del Trabajo se concluye que:

1. El Marco de Trabajo fue aprobado por todos los participantes (100%) como necesarios para las empresas dedicadas al desarrollo de software, el cual puede enriquecer la etapa de planeación de un proyecto de software.

2. La falta de conocimiento técnico relacionado a la disciplina de MP ocasiona la solicitud de la caracterización de los componentes del Marco de Trabajo a través de un ejemplo de un registro de eventos.
3. La modificación del esquema de los componentes a partir de las observaciones de los participantes brindó una mejor claridad y contexto a los componentes del Marco de Trabajo.
4. Las actividades y objetivos son apropiados y posibles de implementar por partes de las empresas de software, teniendo la oportunidad de obtener los beneficios que la MP puede ofrecer.
5. El flujo de trabajo de las tareas al modificarlo con el punto de control entre las tareas de la actividad 1 (A1) y el flujo alterno de los algoritmos de MP representó un mecanismo de fortalecimiento para las desventajas detectadas con su flujo fijo.
6. Los productos de trabajo (PT) tienen la oportunidad de ser ampliados y especializados dependiendo del dominio del proyecto que se esté desarrollando.
7. Las herramientas de MP pueden brindar a las empresas de software ayuda en sus procesos internos como autoevaluaciones.
8. Los algoritmos de MP representan una alta oportunidad de ser utilizados para los procesos de análisis.
9. El contenido de la Guía Técnica a partir de su reestructuración ofrece un mayor entendimiento al contacto directo con la información que posee.

Los aspectos anteriores enriquecieron al Marco de Trabajo, debido a que ofrece una oportunidad de explorar aspectos que no se contemplaron previamente. La opinión de los participantes como expertos en IS brindó dos tipos de visiones: 1) Experiencia en institución académica y 2) Experiencia en la industria. La combinación de estas dos visiones origina un fortalecimiento para los dos componentes del Marco de Trabajo, de lo cual se establecieron las siguientes declaraciones:

- El establecimiento de MoProPEI-MP representa una oportunidad de 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.

- El contenido presente en la Guía Técnica permite emplearse como una herramienta descriptiva de conocimiento tópico que permita llevar la contextualización de la MP para consulta en la implementación del modelo de referencia.

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 emergente, se presentan desafíos, siendo uno de ellos la decisión de identificar y seleccionar algoritmos y herramientas adecuadas para el análisis y exploración de los registros de eventos (Urrea-Contreras et al., 2018a). Dado que las propuestas, tanto en casos prácticos como en los teóricos, en algoritmos como en herramientas orientadas a la MP son constantes, esta línea de investigación debe ser persistente para poder contrastar los factores que se pueden ofrecer. Mostramos cómo la MP puede integrarse en el proceso de software y resolver problemas relacionados con la eficiencia en el monitoreo y control y la mejora de los procesos.

6.2 Logros Académicos

Durante el desarrollo de la investigación se obtuvieron los siguientes productos académicos:

6.2.1 Productos Académicos

- Silvia Jaqueline Urrea-Contreras, Brenda Leticia Flores-Rios, María Angélica Astorga-Vargas, Jorge Eduardo Ibarra Esquer, Larysa Burtseva. Adopción de Herramientas de Minería de Procesos en la Ingeniería de Software, *Abstraction & Application*. Vol. 20. Pp. 24–37. 2018. México. ISSN: 2007-2635. Índice: Latindex.
- Silvia Jaqueline Urrea-Contreras, Brenda Leticia Flores-Rios, María Angélica Astorga-Vargas, Jorge Eduardo Ibarra Esquer, Félix Fernando González-Navarro. MoProPEI-MP an Adaptation in the Selection of Process Mining Tools for Software Processes. *Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería de Conocimiento (JIISIC 2018)*. Pp. 151-162. 2018. Chile. Índice: Scopus

6.2.2 Conferencias en Congresos Internacionales

- Adopción de Herramientas de Minería de Procesos en la Ingeniería de Software, 6to Congreso Internacional de Investigación e Innovación en Ingeniería de Software (CONISOFT, 2018). 25 de Octubre 2018, San Luis Potosí, México.

- MoProPEI-MP an Adaptation in the Selection of Process Mining Tools for Software Processes. Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería de Conocimiento (JIISIC 2018). 14 de Junio 2018. Copiapó, Chile.

6.2.3 Ponencia en Seminario de Investigación

- “5 cosas que debes de saber de la Minería de Datos y Minería de Procesos”, Ciclo de Conferencias: Jueves de Ciencia, Academia de Física de la Facultad de Ingeniería Campus Mexicali. 18 de Octubre 2018, Mexicali, B. C, México.

6.2.4 Registro de Modelo de Utilidad

- Urrea Contreras, S. J., Flores Rios, B. L., Astorga Vargas, M. A., Ibarra Esquer, J. E., González Navarro, F. F., Burtseva, L., Catálogo de algoritmos para la minería de procesos orientados a la Ingeniería de software. Número: 03-2019-030410242100-01. Instituto Nacional del Derecho de Autor. México. 2019.

6.3 Trabajo Futuro

Para continuar con la persistencia de la línea de investigación abordada y retomando el trabajo a futuro detectado en la Tabla 43 a partir de los resultados de la evaluación del Marco de Trabajo, se plantean las siguientes acciones:

1. Realizar la integración de las perspectivas de MP al Esquema de los componentes del Marco de Trabajo para la MP.
2. Extender el esquema de componentes integrando las perspectivas de MP para obtener una mejor contextualización del proceso minado.
3. Trazar la relación de las tareas con su respectivo objetivo que la derivó.
4. Buscar la rapidez del cumplimiento del flujo de trabajo analizando la posibilidad de la combinación de las tareas de Identificar y Seleccionar (Algoritmos o herramientas de MP).
5. Integrar al flujo de trabajo un flujo alternativo que permita poder alimentar el problema a resolver a partir de la visualización de la identificación de algoritmos a utilizar.
6. Establecer plantillas que se puedan adaptar dependiendo el tipo de proyecto de software que se esté realizando para la verificación exitosa de la tarea y de esa forma orientar en la realización del flujo de trabajo.

7. Actualizar el contenido de las herramientas de MP, analizando la alternativa de integrar o excluir herramientas a partir de una priorización de las mismas.
8. Definir una priorización de los algoritmos de MP por facilidad de uso para tener en cuenta sus características y agilizar la identificación y selección de los mismos.
9. Establecer los mecanismos y acciones que se tomaran para mantener actualizada la Guía Técnica.
10. Definir un ejemplo a través de un registro de eventos de la implementación del Marco de Trabajo para una mejor caracterización del contexto utilizando las perspectivas de MP.

Referencias

- Aguirre-Mayorga, H., & Rincón-García, N. (2015). Minería de procesos: desarrollo, aplicaciones y factores críticos. *Cuadernos De Administración*, 28(50), 137-157. doi:10.11144/Javeriana.cao28-50.mpda
- Aguirre-Mayorga, S., & Parra-Rodriguez, C. (2014). Metodología para la aplicación de minería de procesos. *Jornadas Chilenas de Computación*.
- Ailenei, I. M. (2011). Process Mining Tools: A Comparative Analysis. *Partial fulfillment of the requirements for the degree of Master of Science in Business Information Systems*.
- Alves de Medeiros, A. K. (s.f). Process Mining-Control flow mining algorithm. *Eindhoven University of Technology*. Obtenido de http://www.processmining.org/_media/courses/processmining/lecture3_controlflowminingalgorithms.pdf
- Ardila, C. (2013). Marco de Trabajo para Gestión Cuantitativa de Procesos de Desarrollo de Software en Pequeñas Empresas. *Tesis de Maestría en Computación*, Universidad del Cauca.
- Astorga-Vargas, M., Morales-Bustamante, J., Flores-Rios, B., & Ibarra-Esquer, J. (2014). 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)*. doi:10.1109/CISTI.2014.6876898
- Biolchini, J., Mian, P., Natali, A., & Travassos, G. (2005). Systematic Review in Software Engineering. *System Engineering and Computer Science Department COPPE/UFRJ*, 679(05), Technical Report ES, 1-30.
- Chamorro, M., & Maturana, S. (2013). Método para aplicar minería de procesos a la distribución de bebestibles no alcohólicos. *esis de Magíster en Ciencias. Universidad Católica. Santiago de Chile*. Obtenido de <https://repositorio.uc.cl/bitstream/handle/11534/2934/623146.pdf?sequence=1>
- Erdem, S., Demirörs, O., & Rabhi, F. (2018). Systematic mapping study on process mining in agile software development. *International Conference on Software Process Improvement and Capability Determination*, 289-299.
- Farzanyar, Z., & Kangavari, M. (2012). Efficient mining of fuzzy association rules from the pre-processed dataset. *Computing and Informatics*, 31(2), 331-347.
- Flores-Rios, B., Astorga-Vargas, M., Rodríguez-Elias, O., Ibarra-Esquer, J., & Andrade, M. (2014). Interpreting the mexican standards for software process implementation and capacity assessment under a knowledge management approach. *Revista de la Facultad de Ingeniería. Universidad de Antioquia*, 71, 85-100.

- García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., . . . Vanrell, J. (2011). Towards an Information Mining Engineering. *Software Engineering, Methods, Modeling and Teaching*, 83-99.
- Gupta, E. P. (2014). Process mining a comparative study. *International Journal of Advanced Research in Computer and Communications Engineering*, 3(11). doi:10.17148/ijarccce
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2014). Metodología de la investigación. vol. 3, México: McGraw-Hill.
- Kebede, M. (2015). Comparative Evaluation of Process Mining Tools. *Master's Thesis. University of Tartu*.
- Kitchenham, B. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. *EBSE Technical Report EBSE-2007-01*, Keele University and University of Durham.
- Kontio, J., Lehtola, L., & Bragge, J. (2004). Using the focus group method in software engineering: obtaining practitioner and user experiences. *2004 International Symposium on Empirical Software Engineering*, 271-280. IEEE.
- Lemos, A., Sabino, C., Lima, R., & Oliveira, C. (2011). Using Process Mining in Software Development Process Management: A Case Study. *Systems, Man and Cybernetics (SMC)*, 1181-1186. doi:10.1109/ICSMC.2011.6083858
- Martins, S., Pesado, P., & García-Martínez, R. (2014). Propuesta de Modelo de Procesos para una Ingeniería de Explotación de Información: MoProPEI. *Revista Latinoamericana de Ingeniería de Software*, 313-332. doi:10.18294/relais.2014.313-332
- Mendoza-Moreno, M., González-Serrano, C., & Pino, F. (2013). Focus group como proceso en ingeniería de software: una experiencia desde la práctica. *Dyna*, 80(181), 51-60.
- Nakatumba, J. (2010). Analyzing Resource Behavior Using Process Mining. *BPM 2009 Workshops, Proceedings of the Fifth Workshop on Business Process Intelligence*, 43, 69-80.
- Negash, S., & Gray, P. (2008). Business intelligence. Handbook on decision support systems 2. *International Handbooks Information System*, 175-193. doi:10.1007/978-3-540-48716-6_9
- Oktaba, H., Esquivel, C., Ramos, A., Martínez, A., Osorio, G., López, M., . . . Lemus, M. (2005). Modelo de Procesos para la Industria de Software. *México: Secretaría de Economía*.
- Orellana-García, A., Larrea-Armenteros, O., & Pérez-Alfonso, D. (2015). Generador de Registros de Eventos para el análisis de procesos en el Sistema de Información Hospitalaria xavia HIS. *Convención Salud 2015*.

- Orellana-García, A., Pérez-Alfonso, D., & Estrada-Sentí, V. (2016). Revisión de los principales modelos para aplicar técnicas de Minería de procesos. *Revista Internacional de Gestión Del Conocimiento y la Tecnología*, 12-21.
- Pérez-Alfonso, D., Yzquierdo-Herrera, R., & Lazo-Cortés, M. (2012). Utilización de técnicas de minería de proceso en el entorno empresarial Cubano. *VII Conferencia Internacional de Ciencias Empresariales*. Obtenido de <https://www.researchgate.net/publication/237843306>
- Pollo-Cattaneo, M. F., García-Martínez, R., Britos, P., Pesado, P., & Bertone, R. (2012). Elementos para una ingeniería de explotación de información. *Proyecciones*, 10, 67-84.
- Poncin, W., Serebrenik, A., & Van den Brand, M. (2011). Process mining software repositories. *15th European Conference on Software Maintenance and Reengineering*, 5-13. doi:10.1109/CSMR.2011.5
- Poncin, W., Serebrenik, A., & Van Der Brand, M. (2011). Process mining software repositories. *In Software maintenance and reengineering (CSMR)*, 5-14.
- Pressman, R. (2014). *Software Engineering: A Practitioner's Approach*. McGraw Hill, 8a. edición, EUA.
- Pyle, D. (2003). *Business Modeling and Business intelligence*. Morgan Kaufmann Publishers.
- Rubin, V., Lomazova, I., & Van der Aalst, W. (2014). Agile Development with Software Process Mining. *2014 International Conference on Software and System Process (ICSSP 2014)*, 70-74. doi:10.1145/2600821.2600842
- Saravanan, M., & Sree, R. (2011). A Role of Heuristics Miner Algorithm in the Business Process System. *International Journal of Computer Technology and Applications*, 2(2).
- TFFPM – IEEE Task Force on Process Mining. (2012). Process Mining Manifesto. In F. Daniel (Ed.), *Business Process Management Workshops. BPM 2011. Lecture Notes in Business Information Processing, vol 99*, pp. 169-194. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-28108-2_19
- Turner, C. J., Tiwari, A., Olaiya, R., & Xu, Y. (2012). Process mining: from theory to practice. *Business Process Management Journal*, 18(3), 493-512.
- Urrea-Contreras, S. J., Flores-Rios, B. L., Astorga-Vargas, M. A., Ibarra-Esquer, J. E., & Burtseva, L. (2018a). Adopción de Herramientas de Minería de Procesos en la Ingeniería de Software. *Abstraction & Application*, 20, 24-37.
- Urrea-Contreras, S. J., Flores-Rios, B. L., Astorga-Vargas, M. A., Ibarra-Esquer, J. E., & González-Navarro, F. F. (2018b). 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*, 151-162.

- Valenzuela-Ruiz, L., & Flores-Rios, B. (2008). Especificacion formal de elementos MoProSoft a partir del modele de referencia de flujos de trabajo. *VII Jornadas Iberoamericanas de Ingeniería de software e Ingenieria del conocimiento*, 197-203.
- Valle, A., Santos, E., & Loures, E. (2017). Applying process mining techniques in software process appraisals. *Information and Software Technology*, 87, 19-31. doi:10.1016/j.infsof.2017.01.004
- Van der Aalst, W. (2011). Process Mining: Discovery, Conformance and Enhancement of Business Processes. *Springer*.
- Van der Aalst, W. (2012). Process Mining: Overview and Opportunities. *ACM Transactions on Management Information Systems*, 3(2). doi:10.1145/2229156.2229157
- Van der Aalst, W. (2015). Extracting Event Data from Databases to Unleash Process Mining. *Management for Professionals, BPM - Driving Innovation in a Digital World*, 105-128. doi:10.1007/978-3-319-14430-6_8
- Van der Aalst, W. (2016). Process mining: data science in action. *Springer*. doi:10.1007/978-3-662-49851-4
- Van der Aalst, W., & Weijters, A. (2004). Process mining: a research agenda. *Computers in industry*, 53(3), 231-244. doi:10.1016/j.compind.2003.10.001
- Vanrell, J., & Bertone, R. (2010). Modelo de Proceso de Operación para Proyectos de Explotación de Información. *XVI Congreso Argentino de Ciencias de la Computación*, 674-682. Obtenido de <http://hdl.handle.net/10915/19328>
- Velazquez-Solis, P., Flores-Rios, B., Astorga-Vargas, M., Ibarra-Esquer, J., Gonzalez-Navarro, F., & Hernández-Castro, C. (2016). Process Mining in Software Process Improvement, a Systematic Literature Review. *IEEE 11th Colombian Computing Conference (IEEE 11CCC 2016)*. doi:https://www.researchgate.net/publication/309398450
- Wang, J., Wong, R., Ding, J., Guo, Q., & Wen, L. (2013). Efficient Selection of Process Mining Algorithms. *IEEE Transactions on Services Computing*, 6(4), 484-496. doi:10.1109/TSC.2012.20
- Weijters, A., & Ribeiro, J. (2011). Flexible heuristics miner (FHM). *Computational Intelligence and Data Mining (CIDM)*, 310-317. doi:10.1109/CIDM.2011.5949453
- Yzquierdo-Herrera, R., Silverio-Castro, R., Lazo-Cortes, M., & Torres-Grana, A. (2012). Diagnóstico de proceso basado en el descubrimiento de subprocesos. *Ingeniería Industrial* 33, 133-141.

Anexo A. Guía Técnica: Conceptos Básicos de la Minería de Procesos

La Guía Técnica es una herramienta descriptiva que permite llevar a la contextualización un término, desarrollando de esa forma una generación de conocimiento tomando en cuenta el contexto de MP que se expuso en el documento presentado.

Descripción de Términos Básicos

Descripción de términos básicos (Tabla 44) muestra un acceso más eficaz y complementado de la información hacia aquellos términos que se deben de interpretar adecuadamente. Además, añade estabilidad en la aclaración de alguna duda surgida durante la comprensión lectora.

Tabla 44. Descripción de Términos Básicos de la Minería de Procesos.

Términos Básicos	
Minería de Procesos	Disciplina encargada de descubrir, verificar la conformidad y mejorar los procesos identificados que las organizaciones ejecutan. Los procesos identificados se extraen de las fuentes de datos de las organizaciones para ser recopilados en un registro de eventos que será procesado a través de herramientas, algoritmos y/o técnicas propias de la MP.
Registro de eventos	Es la representación de los datos que se originan de la ejecución de las actividades de los procesos identificados.
Conocimiento tópicos	Se refiere al significado de las palabras, como las definiciones en el diccionario o libros de texto (Robillard, 1999).
Conocimiento tácito	Se refiere aquel conocimiento personal que no es fácil de plantear a través del lenguaje formal, por lo que resulta difícil de transmitir y compartir con otros (Nonaka & Konno, 1998). Este tipo de conocimiento tiene sus raíces en las acciones y experiencia individual, así como en los ideales, valores, emociones, intuición, ideas y aspectos subjetivos de cada persona (Flores-Rios, 2016).
Enfoque local	Estrategias locales basadas principalmente en la construcción paso a paso del modelo de proceso óptimo basado en información local (Van der Aalst y Weijters, 2003).
Enfoque global	Estrategias globales basadas principalmente en una búsqueda única del modelo óptimo (Van der Aalst y Weijters, 2003).

Tipos de Minería de Procesos

La MP cuenta con tres tipos de minerías (Tabla 45); por su parte, el Descubrimiento de Procesos se origina a través del análisis de los registros de eventos, por otra parte, la Verificación de la Conformidad y la Mejora de Procesos son posibles al conectar los registros de eventos con modelos de procesos.

Tabla 45. Tipos de Minería de Procesos.

Tipos de Minería de Procesos	
Descubrimiento de Procesos	Una técnica de descubrimiento toma un registro de eventos y produce un modelo sin usar ninguna información a-priori. El descubrimiento de procesos es la técnica de MP más destacada (TFPM – IEEE, 2011).
Verificación de Conformidad	Se compara un modelo de proceso existente con un registro de eventos del mismo proceso. La verificación de conformidad puede ser usada para chequear si la realidad, tal como está almacenada en el registro de eventos, es equivalente al modelo y viceversa. Distintos tipos de modelos pueden ser considerados: la verificación de conformidad puede ser aplicada a modelos procedurales, modelos organizacionales, modelos de procesos declarativos, políticas/reglas de negocio, regulaciones, etc. (TFPM – IEEE, 2011).
Mejora de Procesos	Se busca extender o mejorar un modelo de proceso a-priori existente usando la información acerca del proceso real almacenada en algún registro de eventos (TFPM – IEEE, 2011).

Al aplicar algún tipo de MP y obtener un modelo de proceso, este se puede extender para agregar perspectivas adicionales que aunque su análisis no es tan profundo, ofrecen una caracterización buena para el posterior análisis del modelo de proceso.

Perspectivas de la Minería de Procesos

Las perspectivas de MP (Tabla 46) ofrecen una descripción de los aspectos del proceso de minado, pueden superponerse parcialmente lo que resulta conveniente en el momento del análisis al poder ver las perspectivas de flujo de control, organizacional, del caso y del tiempo desde un mismo punto para su comprensión.

Tabla 46. Perspectivas de la Minería de Procesos.

Perspectivas de la Minería de Procesos	
Perspectiva del flujo de control	Se centra en el flujo de control, es decir, el orden de las actividades. El objetivo de minar esta perspectiva es encontrar una buena caracterización de todos los caminos posibles, por ejemplo, expresados en términos de una red de Petri o alguna otra notación (Van der Aalst, 2016).
Perspectiva organizacional	Se centra en la información sobre los recursos ocultos en el registro, es decir, qué actores (por ejemplo, personas, sistemas, roles y departamentos) están involucrados y cómo están relacionados. El objetivo es estructurar la organización clasificando a las personas en términos de roles y unidades organizativas o para mostrar la red social (Van der Aalst, 2016).
Perspectiva del caso	Se centra en las propiedades de los casos. Un caso se puede caracterizar por su camino en el proceso o por los creadores que trabajan en eso. Sin embargo, los casos también se pueden caracterizar por los valores de los elementos de datos correspondientes (Van der Aalst, 2016).
Perspectiva del tiempo	Se refiere al tiempo y la frecuencia de los eventos. Cuando los eventos llevan marcas de tiempo, es posible descubrir cuellos de botella, medir los niveles de servicio, supervisar la utilización de los recursos y predecir el tiempo de procesamiento restante de los casos en ejecución (Van der Aalst, 2016).