



UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS  
VICERRECTORÍA DE FORMACIÓN  
DIRECCIÓN DE FORMACIÓN POSTGRADUADA

## TÉCNICA PARA EL DIAGNÓSTICO DE VARIANTES DE PROCESOS DE NEGOCIO

Tesis presentada en opción al título de Máster en Informática Aplicada

Autor: **Damián Pérez Alfonso**

Tutor: **Dr.C. Raykenler Yzquierdo Herrera**

Ciudad de la Habana, 9 de Julio de 2014

A mi mamá, por todo su sacrificio.

A mi tía Chichi, por su apoyo y su ejemplo de superación constante.

A mi papá, por su espíritu inquebrantable ante las adversidades.

A mis abuelos, por enseñarme el valor de lo que hemos construido.

A Maya, por seguir siendo mi mango interminable.

Al Fuser, por todo.

A todos los que han hecho y siguen haciendo Revolución...

## DECLARACIÓN JURADA DE AUTORÍA Y AGRADECIMIENTOS

Yo Damián Pérez Alfonso, con carné de identidad 86062314227, declaro que soy el autor principal del resultado que expongo en la presente memoria titulada “Técnica para el diagnóstico de variantes de procesos de negocio”, para optar por el título de Máster en Informática Aplicada.

El trabajo fue desarrollado durante el período comprendido entre el 2013-2014 en colaboración del Dr.C. Raykenler Yzquierdo Herrera, los estudiantes Eudel Pupo Hernández y Reynaldo López Jiménez, quienes me reconocen la autoría principal del resultado expuesto en este trabajo.

En especial deseo agradecer a mi tutor, Raykenler Yzquierdo, quien me involucró y ha guiado en la investigación en minería de proceso, convirtiéndose para mí más que tutor, en un educador y un verdadero amigo. A su apoyo, confianza y los retos constantes que me impone, debo gran parte de mi formación profesional.

Además, deseo agradecer a Pupo, Reynaldo y Osiel, por su colaboración en el desarrollo de esta investigación. A Maya, Jorge, Mairelis, Yanet, Yadira, Juan Carlos, mis colegas del grupo de investigación, del proyecto, del departamento y del Seminario Científico de la Facultad, por escucharme hablar tanto y durante tanto tiempo de minería de proceso. Gracias a todos por sus críticas y comentarios, que de una u otra forma han contribuido a esta investigación.

Finalmente declaro que todo lo anteriormente expuesto se ajusta a la verdad, y asumo la responsabilidad moral y jurídica que se derive de este juramento profesional.

Y para que así conste, firmo la presente declaración jurada de autoría en la Ciudad de la Habana a los 9 días del mes de Julio del año 2014.

---

Firma del Maestrante

## Resumen.

Las técnicas de minería de proceso extraen conocimiento de los registros de eventos generados por los sistemas de información. El diagnóstico de proceso es una de las áreas de la minería de proceso y comprende el análisis de rendimiento, la detección de anomalías y la identificación de patrones de control de flujo. La comprensión del proceso en la etapa de diagnóstico es fundamental, debido a que es la primera dentro de la gestión de procesos de negocio. De forma general, las técnicas disponibles para el diagnóstico suelen ser afectadas por el ruido, el número de tareas involucradas, las dimensiones y/o la complejidad del registro de eventos. Además solo permiten una comprensión fragmentada del proceso, desde la perspectiva única de cada herramienta y brindan poca información sobre los comportamientos infrecuentes del proceso o la incompletitud del registro de eventos. Como parte de esta investigación se propone una nueva técnica para el diagnóstico del proceso, que posibilita la obtención de información relevante del mismo y la identificación de características del registro de eventos como el ruido y la ausencia de información. La técnica, denominada *Minería de Variantes*, tiene como objetivo construir diferentes variantes de modelos de proceso, a partir de la descomposición en subprocesos. La implementación de la *Minería de Variantes* en una herramienta informática, permitió la experimentación con registros de eventos que combinan aleatoriamente situaciones de ruido y ausencia de información. Adicionalmente, se analizaron los resultados de la aplicación de la *Minería de Variantes* a un registro de eventos proveniente de un entorno real. Estos resultados fueron comparados con los obtenidos por otras técnicas para el diagnóstico de procesos. Tanto la evaluación en un entorno real, como el experimento realizado, permiten concluir que la técnica propuesta mejora la comprensión del proceso.

**Palabras claves:** diagnóstico de proceso, minería de proceso, comprensión del proceso, variantes de modelos de proceso.

# Índice general

<b>Introducción</b>	<b>1</b>
Formulación del problema . . . . .	2
Hipótesis . . . . .	3
Objetivo general . . . . .	4
Métodos y técnicas a aplicar . . . . .	4
Aporte teórico-práctico . . . . .	5
Estructura del documento . . . . .	5
<b>1. Fundamentación teórica</b>	<b>7</b>
1.1. Conceptos asociados al dominio del problema . . . . .	7
1.1.1. Minería de proceso . . . . .	7
1.1.2. Diagnóstico de Procesos . . . . .	10
1.2. Comprensión del proceso . . . . .	13
1.2.1. Comprensión desde la modelación de procesos . . . . .	13
1.2.2. Comprensión desde la minería de proceso . . . . .	16
1.2.3. Comprensión del proceso en la etapa de diagnóstico . . . . .	17
1.3. Técnicas para el diagnóstico de procesos . . . . .	18
1.3.1. Análisis de Diagramas de Puntos . . . . .	18
1.3.2. Arreglos Tándem . . . . .	19
1.3.3. Alineación de Trazas . . . . .	20
1.3.4. Minería Difusa . . . . .	21
1.3.5. Descomposición en Bloques de Construcción . . . . .	22
1.3.6. Análisis comparativo . . . . .	24
1.4. Marco tecnológico . . . . .	25
1.5. Conclusiones parciales . . . . .	26

<b>2. Propuesta de solución</b>	<b>28</b>
2.1. Minería de Variantes . . . . .	28
2.2. Pre-procesamiento del registro de eventos . . . . .	32
2.3. Extracción de comportamiento . . . . .	33
2.4. Obtención de variantes de descomposición . . . . .	34
2.4.1. Descomposición en subprocesos . . . . .	34
2.4.2. Identificación de patrones de control de flujo . . . . .	35
2.5. Generación del perfil de diagnóstico . . . . .	41
2.6. Conclusiones parciales . . . . .	41
<b>3. Validación de la solución propuesta</b>	<b>42</b>
3.1. Aplicación informática desarrollada . . . . .	42
3.2. Diseño de la validación . . . . .	44
3.3. Identificación de patrones de control de flujo . . . . .	44
3.3.1. Diseño experimental . . . . .	45
3.3.2. Características de los registros de eventos generados . . . . .	46
3.3.3. Análisis de los resultados . . . . .	47
3.4. Comprensión del proceso en un entorno real . . . . .	48
3.4.1. Aplicación de <i>Minería de Variantes</i> a un registro de eventos del sistema <i>XAVIA HIS</i> . . . . .	48
3.4.2. Aplicación de técnicas existentes para el diagnóstico de proceso . . . . .	56
3.5. Conclusiones parciales . . . . .	61
<b>Conclusiones</b>	<b>62</b>
<b>Recomendaciones</b>	<b>63</b>
<b>Referencias bibliográficas</b>	<b>64</b>

# Índice de figuras

1.1. Representación de los tipos de técnicas de minería de proceso [van der Aalst et al., 2012]. . . . .	8
1.2. El ciclo de vida de BPM [van der Aalst et al., 2012] . . . . .	10
1.3. Análisis de Diagrama de Puntos [Song and Aalst, 2007]. . . . .	19
1.4. <i>Alineación de Trazas</i> [Bose, R.P.J.C. and van der Aalst, W.M.P., 2012]. . . . .	21
1.5. <i>Minería Difusa</i> [Günther and van der Aalst, W.M.P., 2007]. . . . .	22
1.6. <i>Descomposición en Bloques de Construcción</i> [Yzquierdo-Herrera et al., 2013a]. . . . .	23
2.1. Esquema representativo de la técnica Minería de Variantes. . . . .	29
2.2. Variantes de proceso [Li et al., 2011a] . . . . .	30
2.3. Variantes contenidas en un <i>Árbol de variantes</i> . . . . .	31
3.1. <i>Variants Miner</i> , complemento para ProM basado en la técnica propuesta. . . . .	42
3.2. Visualización del <i>Árbol de Variantes</i> y el <i>Perfil de Diagnóstico</i> . . . . .	43
3.3. Patrones identificados en el experimento. . . . .	47
3.4. Modelo del proceso <i>Solicitar Producto</i> . . . . .	49
3.5. Variante del proceso mediante la descomposición por <i>Selección Exclusiva</i> . . . . .	50
3.6. Modelo de proceso del subproceso <i>ver_detalles_sol_almacen</i> . . . . .	51
3.7. Subproceso <i>ver_detalles_sol_almacen</i> . . . . .	51
3.8. Modelo de referencia del subproceso <i>ver_detalles_sol_licitacion</i> . . . . .	52
3.9. Subproceso <i>ver_detalles_sol_licitacion</i> . . . . .	53
3.10. Modelo de proceso del subproceso <i>ver_detalles_sol_bq</i> . . . . .	54
3.11. Subproceso <i>ver_detalles_sol_bq</i> . . . . .	55
3.12. Alineación de trazas y su codificación, relativas al proceso <i>Solicitar Producto</i> . . . . .	56
3.13. Mapa del proceso <i>Solicitar Producto</i> . . . . .	58

3.14.Descomposición en Bloques de Construcción. . . . . 60



# Índice de tablas

1.1. Principios básicos para la modelación de procesos de negocio. . . . .	16
1.2. Características deseadas para la comprensión del proceso en el diagnóstico . .	17
1.3. Características de las técnicas de minería de proceso para el diagnóstico. . . .	24
3.1. Diseño experimental propuesto. . . . .	45
3.2. Características de los modelos de proceso construidos. . . . .	46
3.3. Características de los registros de eventos de cada grupo. . . . .	47

# Introducción

La ejecución de los procesos de negocio en una empresa, habitualmente es registrada por sistemas de información tales como los sistemas de Planificación de Recursos Empresariales (ERP<sup>1</sup>), sistemas de Gestión de la Relación con los Clientes (CRM<sup>2</sup>), sistemas de Gestión de Cadena de Suministros (SCM<sup>3</sup>) u otros de propósitos más específicos [Hendricks et al., 2007]. Estos registros almacenan las trazas de ejecución de los procesos de negocio y se denominan registros de eventos.

La minería de proceso es una disciplina que impulsa el desarrollo de técnicas y herramientas para analizar los procesos partiendo de los registros de eventos. Estos análisis tienen como objetivo proveer un mejor entendimiento de los procesos y los elementos para mejorar su rendimiento a partir de las trazas de ejecución de los mismos [van der Aalst, W.M.P., 2011]. Una traza contiene una secuencia de eventos ordenados según su ocurrencia. Estos eventos reflejan la ejecución de las actividades del proceso al cual pertenece la traza. Asociado a los eventos se registran el instante de tiempo en el cual ocurrió, el recurso que ejecutó la actividad y datos relevantes del negocio [Verbeek et al., 2011].

Las técnicas de minería de proceso analizan automáticamente la información contenida en los registros de eventos. Esto permite que dichas técnicas ofrezcan resultados en menor tiempo y con mayor fiabilidad que las técnicas tradicionales de análisis y mejora de procesos [Jans et al., 2011]. Debido a esto, dentro del área de gestión de procesos de negocio, ha aumentado sostenidamente el interés en la minería de proceso durante la última década [van der Aalst, 2012]. La consultora Gartner, por su parte, recomienda entre las buenas prácticas para organizaciones altamente adaptativas el análisis y descubrimiento de procesos y redes sociales en la organización [Betsy Burton, 2012], los cuales están ampliamente soportados por la minería de proceso. Adicionalmente, más de 100 casos de éxito apuntan

---

<sup>1</sup>Del inglés: Enterprise Resource Planning.

<sup>2</sup>Del inglés: Customer Relationship Management.

<sup>3</sup>Del inglés: Supply Chain Management.

a esta disciplina como una alternativa útil para la gestión de procesos y la auditoría empresarial [Depaire et al., 2013]. Importantes proveedores de software, universidades y empresas consultoras se han aliado junto al IEEE para la investigación y el desarrollo de esta área [van der Aalst et al., 2012].

## Formulación del problema

La minería de proceso comprende el descubrimiento de modelos del proceso, el chequeo de conformidad entre la ejecución real y sus modelos de referencia, así como la mejora de los procesos. Ha demostrado su utilidad en diferentes etapas de la gestión de procesos de negocio, incluyendo el diagnóstico, la ejecución, el ajuste y la (re)configuración. El diagnóstico del proceso ayuda a tener una visión general del proceso, de los aspectos más significativos del mismo y de las técnicas que pueden ser más útiles en su posterior análisis [Yzquierdo-Herrera et al., 2013a]. Esta etapa de la gestión de procesos de negocio abarca el análisis de rendimiento, la detección de anomalías, la identificación de patrones comunes y su contextualización [Bose, R.P.J.C. and van der Aalst, W.M.P., 2012].

La comprensión del proceso en la etapa de diagnóstico es fundamental, debido a que es la primera dentro de la gestión de procesos de negocio. La comprensión de los procesos, a través de los modelos que los representan, depende de las características del modelo y de las características del usuario que interpreta el modelo [Mendling et al., 2012].

La comprensión del modelo debe ser entendida como una escalera [Stamper, 1996]: la sintaxis (cómo se combinan los elementos gramaticales en un modelo de proceso) debe estar clara antes de que la semántica pueda ser discutida y la semántica (qué significan los elementos gramaticales del modelo de proceso) debe estar clara antes de discutir los elementos pragmáticos [Mendling et al., 2012]. Para facilitar la comprensión sintáctica debe considerarse la carga cognitiva extrínseca, la cual es determinada por la forma en la que la información es representada [Kirschner, 2002]. Por tanto, lograr la comprensión del proceso en la fase de diagnóstico requiere de:

- La representación adecuada de los aspectos relevantes (sintaxis).
- La veracidad de dichos aspectos (semántica).
- Que estos conduzcan a interpretaciones y acciones prácticas (pragmática).

Se han desarrollado un conjunto de técnicas de minería de proceso útiles en la etapa de diagnóstico. Sin embargo, determinadas características de los registros de eventos empleados para ello condicionan el impacto de estas técnicas en la comprensión del proceso. Las dimensiones y/o la complejidad del registro de eventos, así como el número de tareas involucradas provocan que las representaciones de las técnicas *Dotted Chart Analysis* [Molka et al., 2013] y *Trace Alignment* [Bose, R.P.J.C. and van der Aalst, W.M.P., 2012] sean complejas, dificultando la comprensión en su dimensión sintáctica. La *Descomposición en Bloques de Construcción* [Yzquierdo-Herrera et al., 2013a] al no considerar el ruido presente en el registro de eventos, genera descomposiciones incorrectas de los subprocesos, afectando la comprensión semántica del proceso en este tipo de situaciones. *Fuzzy Miner* [Günther and van der Aalst, W.M.P., 2007] maneja adecuadamente el ruido, sin embargo no considera la ausencia de información en el registro de eventos, lo cual puede conducir a representaciones incorrectas en el modelo que ofrece. Adicionalmente, las técnicas *Dotted Chart Analysis*, *Fuzzy Miner* y *Trace Alignment* no identifican los patrones de control de flujo presentes. Al no identificar los patrones se dificulta la ubicación en contexto de las anomalías y patrones comunes que ofrecen, limitando la comprensión en el ámbito pragmático.

De forma general, las técnicas disponibles para el diagnóstico suelen ser afectadas por el ruido, el número de tareas involucradas, las dimensiones y/o la complejidad del registro de eventos [López-Jiménez et al., 2014]. Además solo permiten una comprensión fragmentada del proceso, desde la perspectiva única de cada herramienta y brindan poca información sobre los comportamientos infrecuentes del proceso o la incompletitud del registro de eventos.

A partir de la problemática expuesta, se plantea como **problema a resolver**:

La presencia de ruido y/o ausencia de información en los registros de eventos, afecta la comprensión del proceso durante la fase de diagnóstico.

Se delimita como **objeto de estudio** la minería de proceso.

## Hipótesis

El desarrollo de una técnica para el diagnóstico de variantes de procesos de negocio, a partir de registros de eventos con ruido y/o ausencia de información, mejorará la comprensión durante la fase de diagnóstico del proceso.

## Objetivo general

Como **objetivo general** de este trabajo se plantea:

Desarrollar una técnica para el diagnóstico de variantes de procesos de negocio, a partir de registros de eventos con ruido y/o ausencia de información, para mejorar la comprensión durante la fase de diagnóstico del proceso.

Como **campo de acción** se delimita el diagnóstico de proceso y se desglosa el objetivo general en los siguientes **objetivos específicos**:

1. Construir el marco teórico referencial de la investigación, relacionado con el diagnóstico y la comprensión de procesos de negocio.
2. Desarrollar un algoritmo para la detección de patrones de control de flujo en registros de eventos con ruido y/o ausencia de información.
3. Desarrollar un algoritmo para la generación de un perfil de diagnóstico asociado a los patrones de control de flujo del proceso.
4. Desarrollar una herramienta informática basada en los algoritmos propuestos para el diagnóstico de variantes de procesos de negocio.
5. Validar la técnica desarrollada a partir de su aplicación.

## Métodos y técnicas a aplicar

Entre las estrategias de investigación que se utilizan están la exploratoria y la explicativa.

Se exploraron diferentes técnicas y tendencias en el diagnóstico de procesos basados en técnicas de minería de proceso, con vistas a desarrollar una técnica novedosa y que resuelva las deficiencias reportadas en la literatura respecto a la comprensión del proceso.

Los métodos teóricos empleados son: el histórico lógico, hipotético deductivo, analítico sintético y sistémico.

Se enfocaron las problemáticas asociadas a la comprensión en técnicas de minería de proceso para el diagnóstico desde un enfoque histórico lógico. En la primera parte de la investigación se desarrolló un estudio del estado del arte de la problemática analizada, detallando las bondades y deficiencias de cada uno de los métodos y las tendencias en la resolución de esta problemática.

La investigación sigue además un método hipotético deductivo porque a partir del problema concreto se plantean objetivos específicos e hipótesis que en el transcurso de la investigación son resueltas siguiendo métodos científicamente fundamentados.

El método analítico sintético se sigue al descubrir los distintos elementos que componen la naturaleza o esencia asociada a la comprensión del proceso de forma general y en el diagnóstico de forma particular. Se definen las causas y efectos para, posteriormente, integrar los elementos en una unidad nueva, en una comprensión total de la esencia de lo que ya se conoce en todos sus elementos y particularidades.

En cada caso se planteó el problema como un todo donde, las trazas utilizadas, la dinámica de aplicación de técnicas de minería de proceso en el diagnóstico y las técnicas computacionales desarrolladas para mejorar la comprensión del proceso, se funden en un sistema sostenible e integral.

Los métodos empíricos empleados son: experimentación y medición. Para ellos la investigación se basa en la experimentación con datos provenientes de situaciones reales, suministrados por sistemas utilizados para la gestión hospitalaria. Se aplicaron pruebas estadísticas para analizar la efectividad del modelo desarrollado y la calidad de las respuestas finales. Se establecieron indicadores adecuados que permiten realizar correctas mediciones de los resultados.

Los principales aportes de esta investigación se reflejan a continuación:

## **Aporte teórico-práctico**

- Un algoritmo para el diagnóstico de variantes de procesos de negocio.
- Una herramienta informática que implementa el algoritmo propuesto y permite su aplicación en entornos reales.

## **Estructura del documento**

El presente documento está estructurado en 3 capítulos cuyos contenidos se sintetizan a continuación.

**Capítulo 1:** se exponen los principales conceptos relacionados con la minería y el diagnóstico de proceso. Se realiza un estudio crítico de las técnicas de diagnóstico de proceso

existentes, destacándose sus fortalezas y debilidades. Se presentan las principales características del marco de trabajo para la minería de proceso ProM.

**Capítulo 2:** se describe el algoritmo diseñado como parte de una nueva técnica para el diagnóstico de proceso denominada Minería de Variantes. Se analizan cada uno de los patrones de control de flujo identificados. Se presenta la implementación de la técnica desarrollada como un componente para el marco de trabajo ProM.

**Capítulo 3:** se utiliza la técnica desarrollada en el diagnóstico de proceso a partir de registros de eventos provenientes de entornos reales. Se realiza un análisis comparativo de los resultados obtenidos con otras técnicas para el diagnóstico de proceso.

# Capítulo 1

## Fundamentación teórica

Con el objetivo de facilitar la comprensión del alcance de la investigación, en el presente capítulo se exponen conceptos fundamentales asociados a la minería y el diagnóstico de proceso. Posteriormente se analizan los trabajos publicados sobre comprensión de procesos de negocio, con el fin de sintetizar los elementos que pueden contribuir a la misma en la etapa de diagnóstico. Se realiza un análisis de las principales características de las técnicas de minería de proceso utilizadas para el diagnóstico, reflejando sus potencialidades y limitantes, respecto a la comprensión del proceso. Se presenta ProM, marco de trabajo para la minería de proceso.

### 1.1. Conceptos asociados al dominio del problema

#### 1.1.1. Minería de proceso

La minería de proceso es una disciplina científica cuyo fin es descubrir, monitorear y mejorar los procesos reales. Esto se realiza, a través de la extracción de conocimiento de los registros de eventos disponibles en los actuales sistemas de información. Esta se enmarca dentro de la Inteligencia de Negocios (*BI*<sup>1</sup>) formando una conexión entre la Minería de Datos y el Modelado y Análisis de Procesos [[van der Aalst, W.M.P., 2011](#)].

El proceso de negocio es un concepto básico tanto en la minería de proceso como en la *BI*. Se han ofrecido diversas definiciones sobre proceso de negocio [[Hammer and Champy, 1993](#), [Johansson et al., 1993](#), [Rummler and Brache, 2012](#), [Davenport, 2013](#)]. Para la presente investigación se asume la definición de Mathias Weske, por incluir los elementos de interés pa-

---

<sup>1</sup>Del inglés: Business Intelligence.



ra la *BI*: “un proceso de negocio es una colección de actividades que son realizadas coordinadamente en un ambiente técnico y organizacional. La conjunción de estas actividades logra un objetivo del negocio. Cada proceso de negocio es ejecutado por una simple organización, pero con él pueden interactuar procesos de negocios de otras organizaciones” [Weske and Heidelberg, 2007].

Los procesos, en las organizaciones, suelen ser ejecutados utilizando sistemas de información. Un registro de eventos es el resultado del almacenamiento por parte de los sistemas de información de las ejecuciones de un proceso, o sea, un registro de las actividades de un proceso ejecutadas en un período determinado de tiempo [Ma, 2012]. Las ejecuciones del proceso se denominan trazas. Estos elementos se formalizan según la Definición 1.1.1.

**DEFINICIÓN 1.1.1** Traza y registro de eventos: *Se denota  $A$  como el conjunto de las actividades del proceso  $P$ . Un evento  $e$  es una ocurrencia de una actividad:  $e \in A_P$ . Una traza  $t$  es una secuencia finita no vacía de eventos ejecutados en  $P$ :  $t \in A_P^+$ . Un registro de eventos  $L_P$  es un conjunto de trazas sobre  $P$ .*

La minería de proceso tiene como objetivo la extracción de información útil y relevante sobre los procesos, a partir de los registros de eventos. Existen tres tipos fundamentales de técnicas de minería de proceso (Figura 1.1): el descubrimiento de proceso, la verificación de conformidad y el mejoramiento de modelos [van der Aalst et al., 2012].

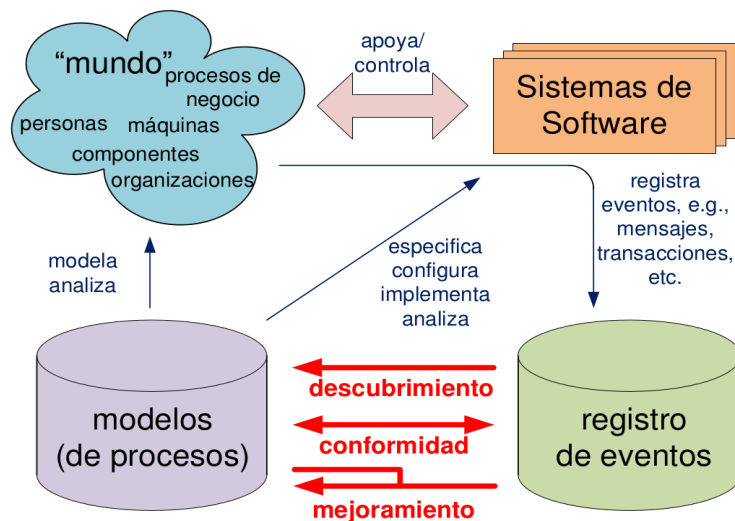


Figura 1.1: Representación de los tipos de técnicas de minería de proceso [van der Aalst et al., 2012].

La verificación de conformidad consiste en analizar un registro de eventos para comprobar que las restricciones del modelo que lo originó se cumplen en la ejecución real del proceso. Las técnicas de verificación se emplean para detectar posibles desviaciones en las ejecuciones de los procesos. El mejoramiento de modelos consiste en analizar la información almacenada en los registros de eventos, con el fin de ajustar o extender el modelo para que refleje mejor lo que ocurre durante su ejecución [van der Aalst et al., 2012]. Por su parte, el descubrimiento de proceso consiste en determinar, a partir de un registro de eventos, un modelo de proceso que explique el comportamiento presente en el mismo.

El descubrimiento es el área más atendida dentro de la minería de proceso, diseñándose varios algoritmos, con diferentes enfoques. Se entiende por algoritmo de descubrimiento una función que mapea un registro de eventos hacia un modelo de proceso [van der Aalst, W.M.P., 2011]. Existen varios algoritmos desarrollados tales como: Alpha [van der Aalst, W.M.P. et al., 2004], Genetic Miner [De Medeiros et al., 2007], Heuristic Miner [Weijters et al., 2006], Transition System Miner y Fuzzy Miner [Günther and van der Aalst, W.M.P., 2007]. Los modelos descubiertos pueden ser representados utilizando múltiples notaciones, dentro de las que se encuentran Redes de Petri y Redes de Flujo de Trabajo [van der Aalst, W.M.P., 2011].

A pesar de la aplicación de la minería de proceso en más de 100 organizaciones, aún es una disciplina emergente y existen importantes retos que deben ser enfrentados [van der Aalst et al., 2012]. Entre estos retos se encuentra el análisis sobre registros de eventos complejos, cuyas características afectan la aplicación de técnicas de minería de proceso y los análisis a partir de estas.

Los registros de eventos provenientes de entornos reales suelen poseer características como ruido, ausencia de información y/o presencia de patrones de control de flujo complejos [De Weerd et al., 2012]. Las técnicas de minería de proceso no manejan estas características en la misma medida, provocando, en algunos casos, la obtención de modelos de proceso de baja calidad [Pérez-Alfonso et al., 2013]. Para esta investigación se asumen las definiciones de ruido (Definición 1.1.2) y ausencia de información (Definición 1.1.3) propuestas por Aalst [van der Aalst, W.M.P., 2011] e Yzquierdo [Yzquierdo-Herrera et al., 2013b] respectivamente.

**DEFINICIÓN 1.1.2** Ruido: *Comportamiento reflejado en las trazas y que rara vez ocurre, que es excepcional o poco frecuente, es decir, que no se corresponde con el comportamiento típico observado en el proceso.*

DEFINICIÓN 1.1.3 Ausencia de información: *Ausencia en las trazas de una o varias actividades ejecutadas como parte de las instancias del proceso, debido a que estas no pueden ser registradas por los sistemas informáticos usados. A este tipo de actividad se le denominará actividad invisible.*

### 1.1.2. Diagnóstico de Procesos

El ciclo de vida de la Gestión de Procesos de Negocio (BPM <sup>2</sup>) abarca las siete fases de un proceso de negocio y los sistemas de información asociados al mismo. Primeramente se *diseña* un proceso, el cual es convertido en un sistema ejecutable en la fase de *implementación*. En la *fase de (re)diseño* se crea un nuevo modelo de proceso o se adapta un modelo de proceso existente. En la *fase de análisis* se analiza un modelo candidato y sus alternativas. Después de la *fase de (re)diseño*, se implementa el modelo (*fase de implementación*) o se (re)configura un sistema existente (*fase de (re)configuración*). En la fase de *ejecución* se ejecuta el modelo diseñado. Durante la *fase de ejecución* el proceso es monitoreado. Además, se podrán realizar pequeños ajustes sin rediseñar el proceso (*fase de ajuste*). En la fase de *diagnóstico* se analiza el proceso ejecutado y la salida de esta fase podría iniciar una nueva fase de *rediseño* del proceso [van der Aalst et al., 2012].

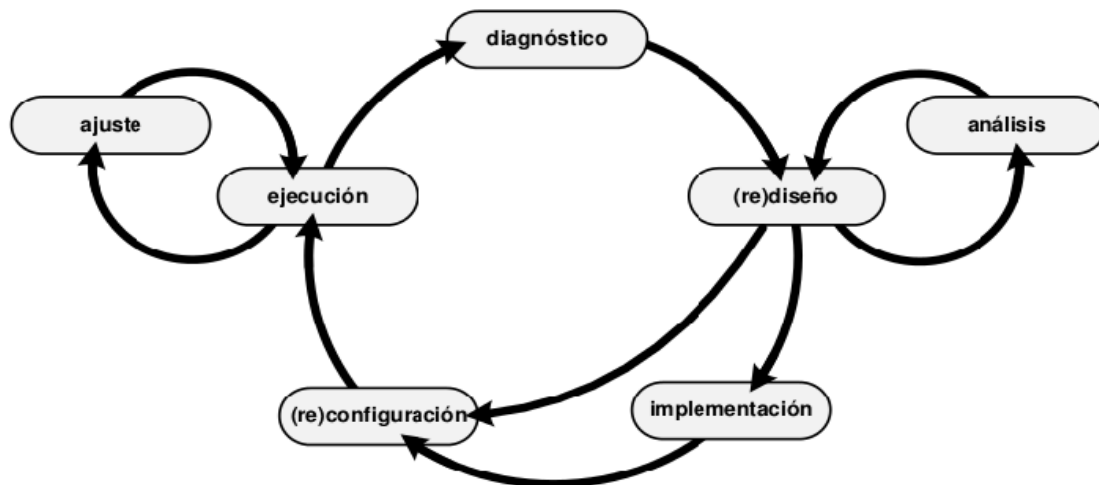


Figura 1.2: El ciclo de vida de BPM [van der Aalst et al., 2012]

El diagnóstico de procesos comprende el chequeo de conformidad, auditoría, análisis de rendimiento, la detección de anomalías y la inspección de patrones interesantes

<sup>2</sup>Del inglés: Business Process Management.

[Jagadeesh Chandra Bose and van der Aalst, W.M.P., 2010]. Por ende, durante la fase de diagnóstico resulta importante determinar los patrones de control de flujo del proceso. Para la presente investigación se presenta la Definición 1.1.4 para formalizar el concepto de patrones de control de flujo. La definición propuesta sintetiza los elementos planteados en [Yzquierdo-Herrera et al., 2013a, Buijs et al., 2012].

DEFINICIÓN 1.1.4 Patrones de control de flujo: *Un proceso se puede descomponer en múltiples subprocesos utilizando los siguientes patrones de control de flujo:*

- Secuencia ( $w_s$ ): *dos subprocesos se ejecutan secuencialmente si uno ocurre inmediatamente después del otro.*
- Selección exclusiva ( $w_x$ ): *dos subprocesos forman parte de una selección exclusiva si, en un punto de decisión, se puede ejecutar solamente uno de los dos.*
- Selección no exclusiva ( $w_o$ ): *dos subprocesos son opciones de una selección no exclusiva si, en un punto de decisión, pueden ejecutarse ambos o solamente uno de ellos.*
- Paralelismo ( $w_p$ ): *dos subprocesos se ejecutan en paralelo si ambos se ejecutan simultáneamente.*
- Lazo ( $w_l$ ): *dos subprocesos se encuentran en un lazo si se pueden repetir múltiples veces. Cada repetición comienza con la ejecución del primer subproceso (**Do**), continúa con el segundo (**Redo**) y termina con el **Do**. El **Redo** puede ser un subproceso vacío, por lo que el único repetido sería el **Do**.*

En este sentido, uno de los principales objetivos del diagnóstico es la determinación de patrones como el paralelismo y la selección, relacionados con la ejecución simultánea de actividades, debido al incremento de complejidad que significa su posterior análisis [Bose, 2012].

Se han desarrollado un conjunto de herramientas de minería de proceso que son útiles para el diagnóstico del proceso [Günther and van der Aalst, W.M.P., 2007, Molka et al., 2013, Bezerra and Wainer, 2013, Bezerra and Wainer, 2012]. Teniendo en cuenta las metas del diagnóstico, se plantea que las herramientas de minería de proceso para el diagnóstico deben brindar respuesta a las siguientes interrogantes [Bose, R.P.J.C. and van der Aalst, W.M.P., 2012]:

1. *¿Cuál es el comportamiento del proceso más común/probable según su ejecución?*
2. *¿Dónde se desvían las instancias del proceso?*

3. *¿Existen patrones de ejecución en las trazas?*
4. *¿Cuáles son los contextos en los cuales se ejecuta una actividad o un conjunto de actividades?*
5. *¿Cuáles son las instancias del proceso que comparten/capturan el comportamiento deseado ya sea exacta o aproximadamente?*
6. *¿Existen patrones particulares en el proceso (hitos, actividades concurrentes, etc)?*

Para dar respuesta a estas preguntas las técnicas para el diagnóstico de procesos se enfocan en identificar los comportamientos de mayor frecuencia de ocurrencia en el registro de eventos, la detección de ejecuciones anómalas y desviaciones en el proceso [Bose, 2012]. La presencia de desviaciones en el modelo constituye un indicador de posibles fraudes o violaciones en las políticas lo cual es útil a las organizaciones en la realización de auditorías. Las ejecuciones anómalas brindan información acerca de las posibles violaciones del modelo. La identificación de los comportamientos y patrones frecuentes posibilita dirigir las técnicas de mejora hacia los elementos más críticos del proceso.

En el diagnóstico también resulta relevante identificar los subprocesos en los que se descompone el proceso y la relación existente entre estos. Los subprocesos pueden ser utilizados para ubicar en contexto los patrones y anomalías identificadas, facilitando su análisis [Yzquierdo-Herrera, 2013]. En la presente investigación se asume el concepto de subproceso planteado en [Yzquierdo-Herrera, 2012] (Definición 1.1.5).

**DEFINICIÓN 1.1.5** Subproceso: *Un subproceso es una encapsulación de las actividades del negocio que representa una unidad de trabajo lógica y coherente. Los subprocesos tienen sus propios atributos y metas, pero contribuyen a alcanzar la meta general del proceso. Un subproceso es también un proceso y la mínima expresión de un subproceso es una actividad.*

El diagnóstico presenta una oportunidad para obtener una visión general de las principales características del proceso evidenciadas en el registro de eventos. Los resultados obtenidos a través del diagnóstico facilitan la selección de las técnicas a utilizarse en posteriores análisis del proceso [Yzquierdo-Herrera et al., 2013a]. Los elementos expuestos evidencian la necesidad de que las técnicas de minería de proceso faciliten una adecuada comprensión del proceso en la fase de diagnóstico.

## 1.2. Comprensión del proceso

La comprensión de un proceso pasa por la comprensión de sus representaciones, es decir, por la comprensión del o de los modelos que lo representan. Los modelos son representaciones explícitas de ciertas porciones de la realidad tal y como son percibidas por determinado actor [Krogstie et al., 2006], concepción que se asume en la presente investigación.

La comprensión del modelo del proceso (en términos de precisión y eficiencia de la comprensión) es una función de las características del modelo y de las características del usuario que interpreta el modelo [Mendling et al., 2012]. Se han investigado aspectos referentes a los usuarios, tales como la familiaridad con una técnica de modelación en particular [Recker and Dreiling, 2011] y los años de experiencia en la modelación de procesos [Mendling et al., 2007]. No obstante, la mayor parte de la investigación en el área de comprensión del modelo ha abordado este fenómeno desde las propiedades del modelo de proceso.

Para establecer los elementos que determinan la comprensión de un modelo de proceso se ha utilizado la semiótica y la teoría de carga cognitiva. Tal como se estableció anteriormente, la teoría semiótica establece que la comprensión debe ser entendida como una escalera, incluyendo la sintaxis, la semántica y la pragmática, en orden ascendente [Stamper, 1996]. Las principales asunciones de la teoría de carga cognitiva son la memoria de trabajo limitada y su interacción con una memoria de largo plazo prácticamente ilimitada.

La teoría de carga cognitiva distingue entre la carga cognitiva intrínseca y extrínseca. La intrínseca está determinada por la complejidad de la información, es decir la cantidad de elementos y sus relaciones e interacciones. En el dominio del proceso, la carga intrínseca se refiere a la complejidad del proceso modelado y por ende está fuera del control del analista que modela el proceso. Por otro lado, la carga extrínseca es determinada por la forma en la que la información es representada [Kirschner, 2002]. Incluso, para los mismos problemas o tareas, la dificultad relativa puede variar en dependencia su representación [Kotovsky et al., 1985]. Por ende, la carga cognitiva extrínseca está sujeta a las decisiones de diseño que se realizan al describir el proceso en un modelo [Figl et al., 2013].

### 1.2.1. Comprensión desde la modelación de procesos

La comprensión ha sido abordada en la literatura como un criterio determinante en la calidad de la modelación de procesos. El framework de calidad SEQUAL [Lindland et al., 1994], se basa en la semiótica, estableciendo las siguientes dimensiones de la calidad de un modelo:

sintáctica, semántica y pragmática.

SEQUAL define varios aspectos de calidad basados en relaciones entre el modelo, un cuerpo de conocimiento, un dominio, un lenguaje de modelado y las actividades de aprendizaje, adopción de medidas y modelación. La calidad sintáctica se asocia al modelo y al lenguaje de modelado; la calidad semántica relaciona el modelo, el dominio y el conocimiento y por último la calidad pragmática relaciona el modelo, la modelación y su habilidad para facilitar el aprendizaje y la acción [Mendling et al., 2007].

La calidad pragmática, según SEQUAL, es la relación entre el modelo y la interpretación de la audiencia del modelo, teniendo como meta la comprensión, atendiendo la habilidad del modelo para facilitar el aprendizaje y la acción. Se proponen, como medios para elevar la calidad pragmática, la capacidad de ejecución, simulación y animación del modelo, así como su transformación y filtrado para presentar abstracciones del modelo desde diferentes puntos de vista [Krogstie et al., 2006].

La Guía de Modelación (GOM<sup>3</sup>) es un marco de trabajo para evaluar la calidad de la modelación, que se basa en los principios de correctitud, claridad, relevancia, comparabilidad, eficiencia económica y diseño sistemático [Becker et al., 2000]. Utilizando el principio de claridad de GOM y la calidad pragmática de SEQUAL se ha determinado que, desde el punto de vista de las características del modelo, el tamaño constituye el factor primario que incide en la comprensión. Adicionalmente, se han identificado como características de interés el grado de conexión promedio entre los elementos y la densidad<sup>4</sup> del modelo [Reijers and Mendling, 2011].

Dentro del análisis de la modelación de procesos se ha estudiado cómo pueden ser diseñados los modelos de procesos para maximizar la comprensión [Recker et al., 2014]. Se han identificado factores que afectan el entendimiento del modelo de proceso tales como: propósito del modelo [Reijers and Mendling, 2011], notación de modelado [Sarshar and Loos, 2005, Hahn and Kim, 1999, Agarwal et al., 1999], presentación visual [Moher et al., 1993, Purchase, 1997, Reijers et al., 2011a] y complejidad del modelo de proceso [Canfora et al., 2005, Mendling, 2008]. Adicionalmente, se han analizado características del modelo que afectan su comprensión, tales como el diseño de los constructores gramaticales<sup>5</sup> [Figl et al., 2013], el resaltado de colores [Reijers et al., 2011a], el uso de diferentes constructores gramaticales [Recker et al., 2013], la utilización de la modularización [Reijers et al., 2011b]

---

<sup>3</sup>Del inglés: Guidelines of Modeling.

<sup>4</sup>Relación entre los arcos existentes con respecto al máximo número de arcos posibles en el modelo.

<sup>5</sup>Conjuntos de símbolos gráficos y reglas que describen cómo se conectan los símbolos gráficos [Wand and Weber, 2002].

y el estilo gramatical de las etiquetas de texto [Mendling et al., 2010a].

La comprensión sintáctica del modelo puede ser analizada desde dos ángulos: el rendimiento (en qué medida la interpretación del modelo contribuye a comprender el contenido formal del modelo) y la eficiencia (qué recursos son utilizados para comprender el modelo) [Mendling et al., 2012, Poels et al., 2011]. Considerando que la semántica del modelo se expresa en sus etiquetas textuales se ha demostrado que la información semántica adicional obstaculiza la comprensión sintáctica [Mendling et al., 2012]. Aunque se esperaría que una persona comprenda mejor un modelo con información textual, se obtiene un efecto diferente al esperado cuando se realizan preguntas que se refieren únicamente a la sintaxis. Esto se debe a una amplia activación de conceptos diferentes, La explicación teórica racional para esta expectativa parte de la teoría de carga cognitiva.

Cuando los individuos estudian un material nuevo, incrementan su carga cognitiva en la memoria de trabajo. Este tipo de memoria tiene la capacidad de procesar aproximadamente siete ítems de información en un instante de tiempo [Miller, 1956]. Una etiqueta de texto larga, implica una mayor carga cognitiva con respecto a una única letra. Las etiquetas textuales, por ende, pueden distraer a las personas de obtener las conclusiones adecuadas de un modelo debido a que una mayor parte de la memoria de trabajo se requiere para procesar la información textual y la información de dominio que representa.

En la comprensión semántica influye la selección del idioma utilizado para las etiquetas textuales presentes en el modelo. Si este coincide con la lengua nativa de los usuarios del modelo la comprensión semántica aumenta. El impacto de este factor sobre la comprensión semántica es significativamente superior al que genera la selección de la notación de modelado [Recker and Dreiling, 2011].

A partir del conocimiento obtenido en investigaciones como las citadas anteriormente, se ha propuesto una guía de siete principios básicos para la modelación de procesos de negocio: **7PMG** (Tabla 1.1). Cada principio está fundamentado en resultados de investigaciones cuantitativas que relacionan los estilos de modelación de procesos por una parte y el entendimiento del modelo y la propensión a errores por otra [Mendling et al., 2010b].



Tabla 1.1: Principios básicos para la modelación de procesos de negocio.

- |   |  |
|---|--|
| 1 | Usar tan pocos elementos en el modelo como sea posible     |
| 2 | Minimizar los caminos en los elementos de enrutamiento     |
| 3 | Utilizar un elemento de inicio y uno de fin                |
| 4 | Modelar lo más estructuradamente posible                   |
| 5 | Evitar los elementos de selección no exclusiva (OR)        |
| 6 | Usar etiquetas para las actividades conformadas por verbos |
| 7 | Descomponer cualquier modelo con más de 50 elementos       |

### 1.2.2. Comprensión desde la minería de proceso

Las investigaciones sobre comprensión, en el ámbito de la modelación, asumen que el modelo refleja fielmente el proceso, de tal forma que la interpretación del modelo conduce a la comprensión del proceso. Sin embargo, la experiencia en la aplicación de técnicas de minería de proceso indica que los modelos construidos con técnicas tradicionales de modelación suelen presentar diferencias notables con la ejecución real del proceso que representan [van der Aalst et al., 2012]. Por lo cual, es relevante considerar la veracidad de los modelos, o sea la relación entre la representación y la realidad, ya que esto tiene un impacto directo en los niveles semántico y pragmático de la comprensión del proceso.

En la minería de proceso, la relación entre un modelo de proceso y las trazas de su ejecución es considerada un indicador de su calidad. Este análisis suele utilizarse para la evaluación de la calidad de los modelos construidos a través de algoritmos de descubrimiento de procesos [vanden Broucke et al., 2013]. La calidad de los modelos descubiertos se establece a partir de un balance entre cuatro dimensiones [van der Aalst, W.M.P., 2011]:

**Aptitud:** el modelo debe permitir el comportamiento presente en el registro de eventos.

**Precisión:** el modelo no debe permitir comportamiento completamente sin relación con el que se puede apreciar en el registro de eventos.

**Generalización:** el modelo debe generalizar el comportamiento observado en el registro de eventos.

**Simplicidad:** el modelo debe ser tan simple como sea posible, reflejando el comportamiento del proceso.

Las dimensiones *Aptitud*, *Precisión* y *Generalización* se refieren a la alineación entre el modelo y la ejecución real del proceso y han sido consideradas en la evaluación de la comprensión del proceso [Yzquierdo-Herrera, 2012]. Por otra parte, la *Simplicidad* considera la carga cognitiva que implican los modelos complejos, producto de su densidad. En cada una de estas dimensiones han sido definidas métricas, que son aplicables fundamentalmente sobre modelos cuya notación sea Redes de Petri [De Weerd et al., 2013].

Se puede apreciar que el enfoque utilizado en la minería de proceso complementa el presentado en el epígrafe anterior, al incluir la alineación entre el modelo y la realidad. No obstante, aún cuando considera aspectos de representación en la dimensión *Simplicidad*, descarta otros elementos ya mencionados de formato, estructuración y presentación del modelo que también inciden en la comprensión del proceso.

### 1.2.3. Comprensión del proceso en la etapa de diagnóstico

Teniendo en cuenta los elementos teóricos planteados sobre los enfoques utilizados para analizar la comprensión en la modelación y la minería de proceso, se ha decidido utilizar, en esta investigación, un enfoque híbrido. Se propone un conjunto de características que deben poseer las herramientas de minería de proceso para mejorar la comprensión del proceso en la fase de diagnóstico. Esta propuesta se basa en la combinación de los objetivos del diagnóstico y los factores que influyen en la comprensión del proceso, ya sean relativos al modelo o a su alineación con la ejecución real del mismo. La propuesta se muestra en la Tabla 1.2.

Tabla 1.2: Características deseadas para la comprensión del proceso en el diagnóstico

<b>Características</b>	<b>Factores para la comprensión</b>
1 Identificación de patrones de control de flujo	Constructores gramaticales
2 Vista jerárquica del proceso	Modularización (subprocesos)
3 Vista global del proceso	Complejidad del modelo
4 Visualización y manejo del ruido	Complejidad del modelo
5 Visualización y manejo de la ausencia de información	Complejidad del modelo

La primera característica propuesta es la identificación de los patrones de control de flujo. Estos patrones constituyen elementos fundamentales dentro del proceso, ya que definen la forma en la que las actividades son ejecutadas. También constituyen puntos de referencia para la identificación de anomalías y desviaciones, el manejo del ruido y la ausencia de información. Adicionalmente, los patrones pueden ser utilizados para la descomposición en subprocesos y

la identificación de comportamientos comunes. Al ser constructores gramaticales, su identificación y diseño tienen un impacto en los niveles sintáctico y semántico de la comprensión del proceso. Su correcta identificación impacta además en la dimensión *Aptitud*.

La obtención de una vista jerárquica del proceso permite su análisis por niveles de abstracción y la ubicación en contexto de la ejecución de las actividades y los patrones interesantes. Se logra mediante modularización o separación en subprocesos, permitiendo realizar el análisis por partes, disminuyendo la carga cognitiva y mejorando de esta manera la comprensión.

La complejidad del modelo es otro elemento que impacta directamente en la comprensión. Obtener una vista global del proceso disminuye la complejidad del modelo porque permite la abstracción de los patrones específicos concentrándose en los comportamientos y patrones generales y de más alto nivel. La visualización y manejo del ruido y la ausencia de información define la cantidad de relaciones entre las actividades que se muestran, lo cual puede conducir a modelos con muchos nodos y alta densidad o a modelos sencillos pero que no representan adecuadamente al proceso. La identificación y manejo del ruido y la ausencia de información inciden directamente en las dimensiones de calidad *Precisión* y *Generalización*.

### **1.3. Técnicas para el diagnóstico de procesos**

El diagnóstico de procesos brinda información útil sobre determinados aspectos del proceso. Con el objetivo de extraer y mostrar dicha información se han desarrollado un conjunto de técnicas de minería de proceso útiles para su diagnóstico. A continuación se analizan las principales características de algunas de estas técnicas.

#### **1.3.1. Análisis de Diagramas de Puntos**

El Análisis de Diagramas de Puntos [Song and Aalst, 2007] permite tener una visión general del registro de eventos, lo que facilita la identificación de los aspectos significativos en el proceso. Dicha técnica, similar a los Diagramas de Gantt, presenta una “vista helicóptero” del registro de eventos. Muestra los eventos del proceso de forma gráfica, como puntos en un plano, donde una dimensión representa las trazas y la otra se refiere al instante en el que se ejecutaron los eventos. Cada uno de los eventos se representa en el diagrama con un color particular. En la Figura 1.3 se muestra un resultado de la aplicación de esta técnica.

El Análisis de Diagramas de Puntos presenta algunos indicadores de desempeño como

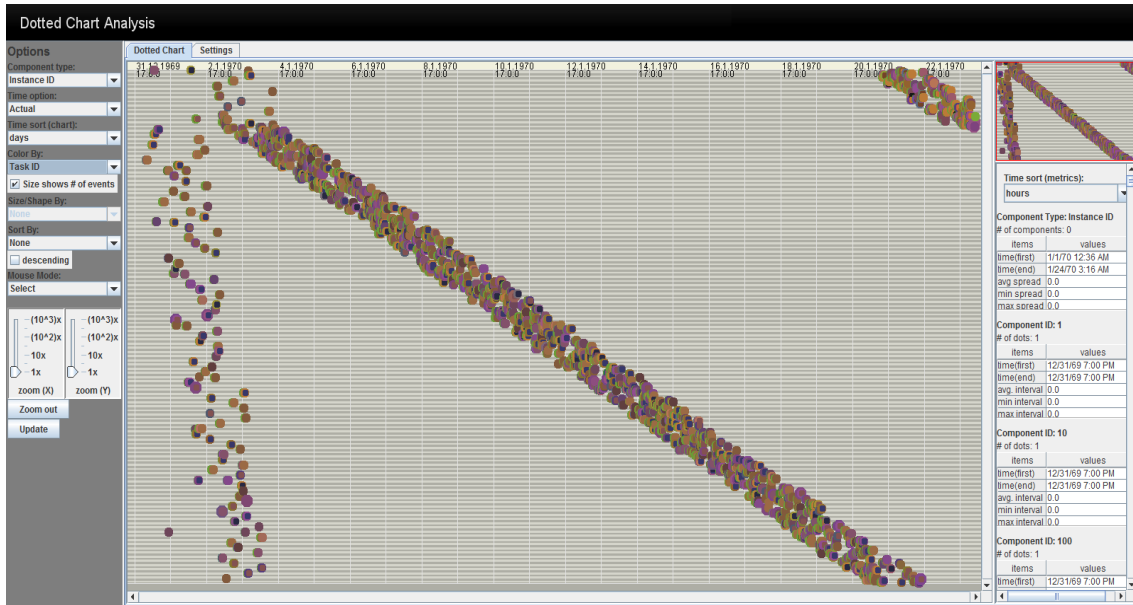


Figura 1.3: Análisis de Diagrama de Puntos [Song and Aalst, 2007].

el intervalo de tiempo mínimo, máximo y promedio entre la ejecución de los eventos, lo que facilita el análisis del rendimiento del proceso. Debido a que los eventos no se encuentran alineados con otras ejecuciones del proceso, se dificulta la detección de patrones que involucren varias trazas. Por otra parte, esta técnica es susceptible a la cantidad de actividades presentes en el proceso, pues al incrementarse el número de actividades al orden de las decenas, la inspección manual, la comprensión de la tabla de puntos y la identificación de patrones se torna compleja. Se considera que todos los eventos pertenecen al mismo nivel de abstracción en el registro de eventos, lo que impide la identificación de los subprocesos.

### 1.3.2. Arreglos Tándem

En [Jagadeesh Chandra Bose and van der Aalst, W.M.P., 2009] se propone el uso de Arreglos Tándem<sup>6</sup> y repeticiones máximas para capturar patrones recurrentes dentro y a través de las trazas. Esta técnica se emplea para procesar el registro de eventos, sustituyendo determinados eventos por los respectivos patrones a los que pertenecen. El registro de eventos una vez procesado es utilizado para aplicar algoritmos de descubrimiento de procesos.

La utilización de esta técnica permite la determinación de los patrones en tiempo de cómputo lineal, aunque el número de patrones descubiertos puede ser elevado. Para solventar esta

<sup>6</sup>Del inglés: Tandem Arrays

situación en [Li et al., 2011b] se han propuesto un conjunto de métricas que posibilitan el filtrado de los patrones insignificantes. Por otra parte, los patrones descubiertos son atómicos por lo que las dependencias entre estos deben ser descubiertas por separado.

### 1.3.3. Alineación de Trazas

La *Alineación de Trazas* es una técnica de diagnóstico que utiliza la programación dinámica para tabular las trazas, de forma que se simplifiquen los problemas asociados al entendimiento de las relaciones entre las actividades [Jagadeesh Chandra Bose and van der Aalst, W.M.P., 2010]. Esta técnica se basa en la Alineación de Múltiples Secuencias (MSA<sup>7</sup>), una herramienta empleada en la bioinformática para determinar las estructuras secundarias y terciarias de las proteínas y moléculas. Similar a la bioinformática, la minería de proceso se encuentra relacionada con las secuencias, aunque en este caso las secuencias se originan por el registro de la ejecución del proceso.

Para realizar la *Alineación de Trazas* se emplea el enfoque de alineación progresiva. El mismo se basa en la construcción iterativa de parejas de alineaciones, permitiendo la alineación entre: dos alineaciones, una traza y una alineación o dos trazas [Bose, R.P.J.C. and van der Aalst, W.M.P., 2012]. La calidad de la alineación final sobre un conjunto de trazas depende en gran medida del orden en que se realiza la alineación progresiva. La Figura 1.4 muestra un conjunto de trazas alineadas.

Una de las potencialidades de la *Alineación de Trazas* consiste en posibilitar la identificación de desviaciones en la ejecución del proceso, es decir, distinguir las ocurrencias anómalas del proceso, de las ocurrencias comunes. Dicha técnica permite la visualización de múltiples ejecuciones del proceso simultáneamente y facilita la determinación visual de los patrones de control de flujo.

Sin embargo, esta técnica no permite detectar los subprocesos que conforman al proceso analizado y enmarcar en estos las anomalías y patrones identificados. Esto dificulta la contextualización del aspecto detectado y el entendimiento de las causas que lo originaron [Yzquierdo-Herrera et al., 2013a]. Por otra parte, la presencia de ruido en el registro de eventos puede conducir a la obtención de alineaciones con baja calidad [Bose, R.P.J.C. and van der Aalst, W.M.P., 2012].

---

<sup>7</sup>Del inglés: Multiple Sequence Alignment.

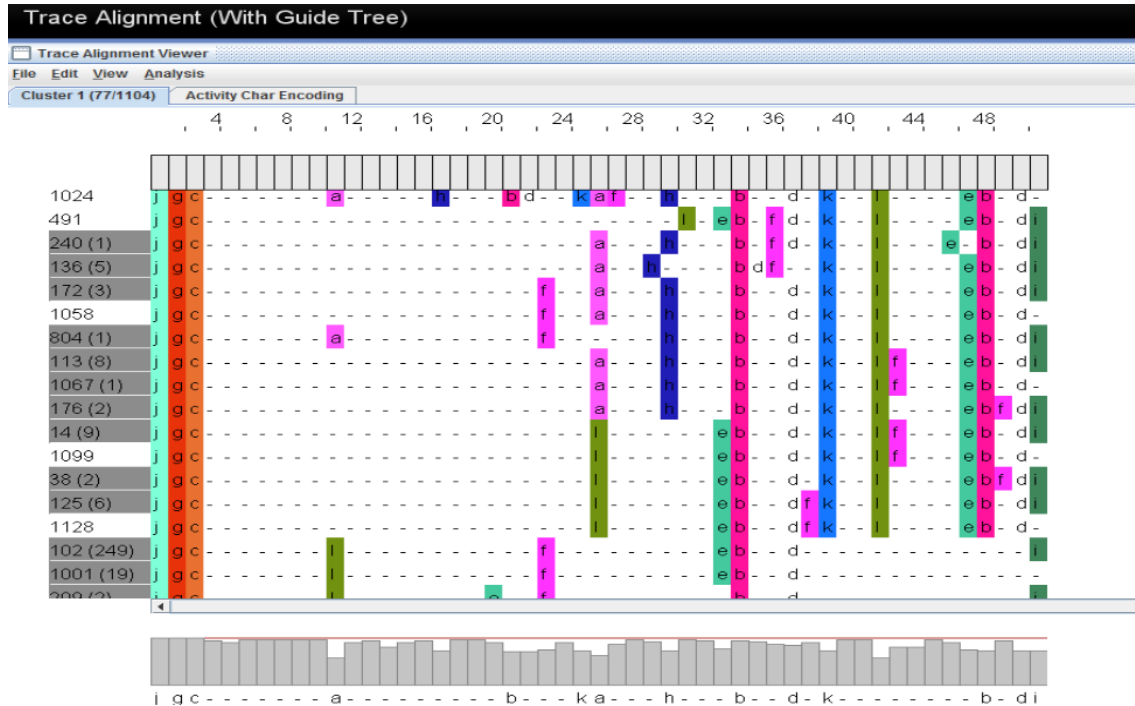


Figura 1.4: Alineación de Trazas [Bose, R.P.J.C. and van der Aalst, W.M.P., 2012].

### 1.3.4. Minería Difusa

La *Minería Difusa* es una técnica de diagnóstico que muestra las actividades y sus relaciones según diferentes niveles de abstracción. Utiliza similitudes con la forma de representar la información en la cartografía, aplicando conceptos como: abstracción, personalización, agregación y énfasis [van der Aalst, W.M.P., 2011]. Dicha técnica produce un modelo basado en grafos, donde existen dos tipos de nodos, los que representan una actividad y los que agrupan un conjunto de actividades denominados clústeres. Un modelo obtenido mediante esta técnica se presenta en la Figura 1.5.

La agregación se refiere a limitar la cantidad de elementos, nodos y relaciones que se muestran. En su lugar, varios elementos forman clústeres encapsulando la información y facilitando su visualización. La cantidad de información que se muestra es controlada mediante el empleo de umbrales de permisibilidad. La abstracción consiste en omitir de la visualización la información que en determinado contexto es insignificante. En determinados componentes se destacan características como color, contraste, saturación y tamaño para facilitar la visualización de la información relevante. Por su parte la personalización permite configurar la información que se muestra de acuerdo a determinados criterios

[Günther and van der Aalst, W.M.P., 2007].

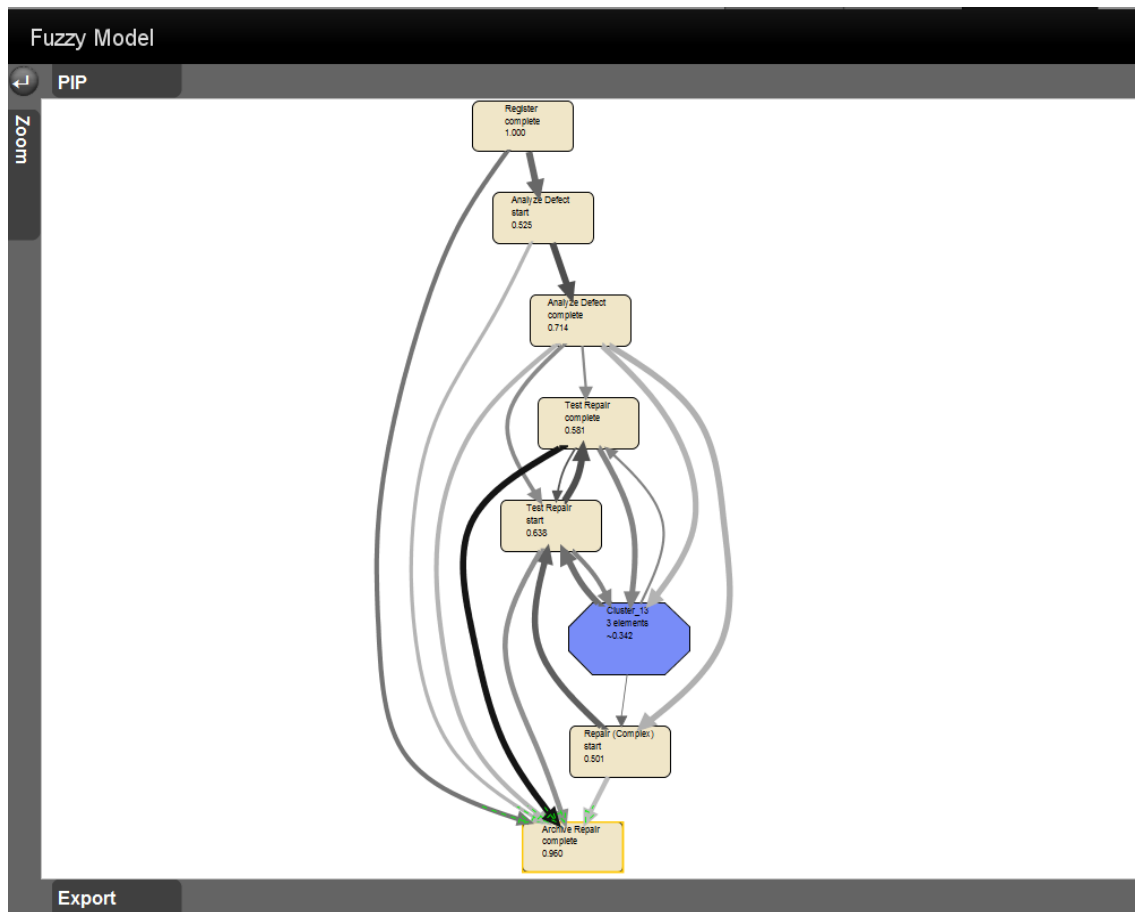


Figura 1.5: Minería Difusa [Günther and van der Aalst, W.M.P., 2007].

La *Minería Difusa* asume que cada actividad puede pertenecer solamente uno de los clústeres creados, sin tener en cuenta la presencia de relaciones de muchos a muchos entre actividades y clústeres. Además, los clústeres pueden no corresponderse con subprocessos reales. Esta técnica no permite la identificación de patrones de control de flujo ya que muestra solo las relaciones existentes entre actividades. Por último, es relevante señalar que, aunque maneja el ruido, no considera la ausencia de información.

### 1.3.5. Descomposición en Bloques de Construcción

La *Descomposición en Bloques de Construcción* se basa en la técnica de Alineación de trazas, explicada anteriormente. Consiste en obtener una visión jerárquica del proceso, descomponiendo cada nivel según el patrón de control de flujo identificado

[Yzquierdo-Herrera et al., 2013a]. En la Figura 1.6 se muestra el resultado de la aplicación de esta técnica.

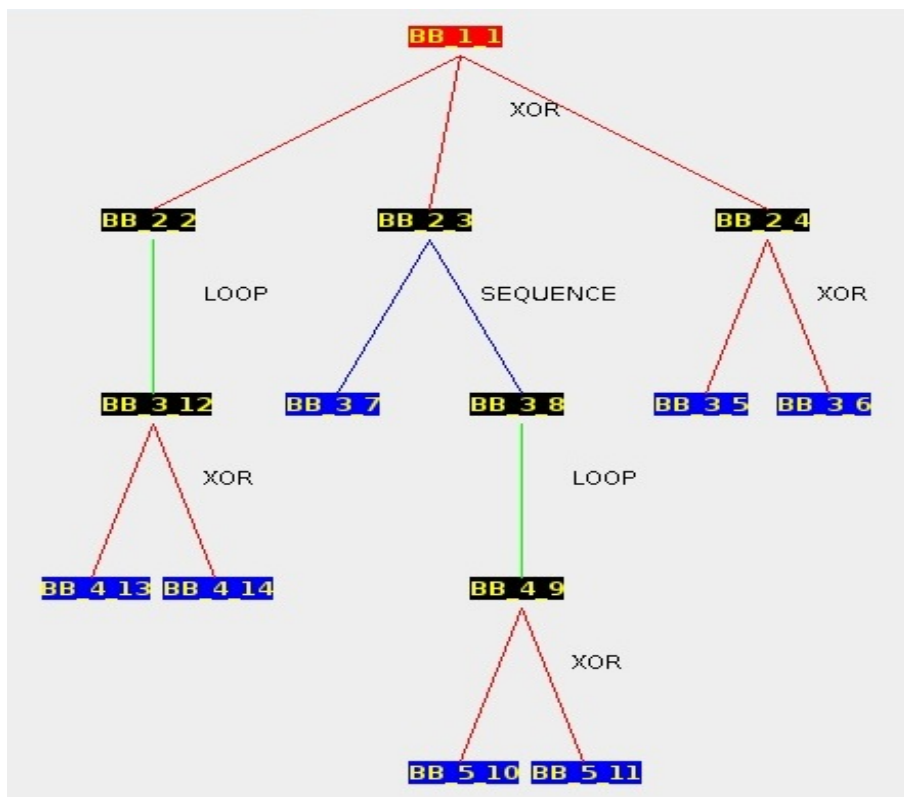


Figura 1.6: *Descomposición en Bloques de Construcción* [Yzquierdo-Herrera et al., 2013a].

Dicha técnica permite detectar los patrones de control de flujo presentes en el registro de eventos. Por otra parte, posibilita detectar los subprocesos que conforman el proceso analizado y enmarcar en éstos las anomalías y patrones detectados. Mediante la utilización de la *Descomposición en Bloques de Construcción*, se puede combinar el análisis de las frecuencias en que ocurren las actividades con el análisis de manera escalonada de la secuencia de acontecimientos estructurada en los subprocesos. Esto contribuye a la comprensión de las causas de las fallas y por tanto, a la posible mejora del proceso [Yzquierdo-Herrera et al., 2013a].

Al utilizar como base la *Alineación de Trazas*, la técnica se encuentra sujeta a las dificultades inherentes a ésta, además de agregarle un tiempo adicional a su aplicación. La principal debilidad de la técnica *Descomposición en Bloques de Construcción* es la afectación de sus resultados por la presencia de ruido y algunos casos de ausencia de información. Dichas características, presentes en los registros de eventos, dificultan la identificación de patrones de control de flujo.



### 1.3.6. Análisis comparativo

Se ha realizado un análisis comparativo considerando las características deseadas, en las técnicas de diagnóstico, para la comprensión del proceso. En la Tabla 1.3 se sintetizan las limitaciones y potencialidades de las técnicas analizadas.

Tabla 1.3: Características de las técnicas de minería de proceso para el diagnóstico.

	Patrones de control de flujo	Vista jerárquica	Vista global	Visualización y manejo del ruido	Visualización y manejo de la ausencia de información
Análisis de Diagramas de Punto			✓		
Arreglos Tándem	≈				
Alineación de Trazas	≈		✓		
Minería Difusa	≈	≈	✓	✓	
Descomposición en Bloques de Construcción	✓	✓	✓		✓

Se puede apreciar que la mayoría de las técnicas ofrecen una vista global del proceso. Adicionalmente, la característica de identificación de los patrones de control de flujo está presente al menos parcialmente en 4 de las técnicas analizadas. En este sentido es importante destacar que solamente la *Descomposición en Bloques de Construcción* identifica completamente los patrones. Además, la *Minería Difusa* permite conocer información parcial asociada a los patrones tal como las actividades que se ejecutan secuencialmente y en lazos. Por su parte la *Alineación de trazas* facilita la identificación de secuencias, lazos y selecciones exclusivas en la ejecución de las actividades, siempre que las dimensiones y calidad de la alineación obtenida permita su inspección a simple vista.

La vista jerárquica del proceso se logra parcialmente en la *Minería Difusa* ya que solo considera relaciones entre eventos y clústeres de uno a muchos, sin considerar el concepto de subproceso. En contraposición resalta en esta técnica el manejo del ruido a partir de umbrales configurables. Sin embargo, esta característica es el punto débil de la técnica que muestra los mejores resultados en el análisis realizado, la *Descomposición en Bloques de Construcción*.

Las técnicas existentes suelen descartar el ruido presente en el registro de eventos o no identificarlo. Al no identificar el ruido pueden obtenerse modelos demasiado complejos o que no representen adecuadamente el proceso, como sucede con la *Descomposición en Bloques de Construcción*.

La *Minería Difusa* utiliza el enfoque de descartar el ruido del modelo, usando umbrales de frecuencia configurables por el usuario. Sin embargo, esto puede obviar en la representación comportamientos del proceso que son infrecuentes por definición.

Ambas alternativas tienen un impacto estructural sobre el modelo obtenido ya que pueden influir en el tipo de patrón de control de flujo identificado o sugerido por la técnica, en las secciones del registro de eventos donde está presente el ruido. Aún cuando el usuario suele tener la posibilidad de establecer los umbrales de frecuencia, como en la *Minería Difusa*, no posee control sobre el impacto estructural de los comportamientos descartados en el modelo del proceso.

La ausencia de información es otra característica del registro de eventos a considerar en la construcción de un modelo. En la bibliografía han sido descritas ocho situaciones diferentes donde la ausencia de información genera un impacto estructural en el modelo [Yzquierdo-Herrera et al., 2013b]. La *Descomposición en Bloques de Construcción* es la única técnica de diagnóstico que considera todas estas situaciones. Además ofrece la oportunidad al usuario de insertar actividades artificiales en el modelo, para representar las actividades no registradas durante la ejecución del proceso.

Para resolver la afectación que provocan el ruido y la ausencia de información en la comprensión del proceso es necesario controlar el impacto estructural de estas características. Sin embargo, considerar ciertos comportamientos infrecuentes como ruido y/o asumir que faltan determinadas evidencias de la ejecución del proceso, son estimaciones que sólo pueden ser confirmadas a partir del contexto de ejecución particular del proceso analizado. Por ende, el usuario debe poder decidir cuáles comportamientos deben ser considerados como ruido y descartados del modelo resultante, así como seleccionar cuáles comportamientos están ausentes y cómo deben ser insertados en el modelo.

## 1.4. Marco tecnológico

ProM es un marco de trabajo para el desarrollo de herramientas de minería de proceso en un ambiente estandarizado [Verbeek et al., 2011]. Está desarrollado en Java y se encuentra

disponible bajo licencia GPL1.

Este marco de trabajo posee una arquitectura orientada a componentes que soporta la adición de complementos. De esta manera constituye un entorno tecnológico para el desarrollo de nuevos algoritmos y técnicas en el campo de la minería de proceso. Los complementos para ProM necesitan determinada cantidad de parámetros de entrada y producen uno o varios objetos de salida. Los parámetros de entrada pueden ser registros de eventos u objetos obtenidos a partir del procesamiento realizado por otros complementos. Mientras que los objetos de salida obtenidos pueden ser empleados como parámetros de entrada de otros complementos.

ProM cuenta con más de 550 complementos, cada uno de los cuales posibilita realizar diferentes análisis y visualizaciones [[Jan Claes and Geert Poels, 2012](#)]. Incluye además facilidades para la exportación, importación y visualización de modelos de procesos en múltiples formatos como Redes de Petri, BPMN, Redes Heurísticas, YAWL, Redes Causales, Mapas de Proceso, Árboles de Proceso, entre otros. Las técnicas de diagnóstico analizadas, en el epígrafe anterior, han sido implementadas como complementos para ProM.

Las herramientas incluidas en ProM han sido empleadas en el análisis de procesos provenientes de diferentes dominios entre los que se encuentran el gubernamental, el hospitalario y sistemas ERP [[van der Aalst, W.M.P., 2011](#)].

## 1.5. Conclusiones parciales

La comprensión del proceso ha sido abordada desde la modelación y la minería de proceso. Ambos enfoques son complementarios y deben ser considerados para un análisis holístico de la comprensión del proceso en la etapa de diagnóstico. La identificación de patrones de control de flujo, manejo y visualización del ruido y la ausencia de información así como la obtención de una vista global y jerárquica del proceso son características que impactan en la comprensión del proceso durante el diagnóstico.

La mayoría de las técnicas de diagnóstico no cubren totalmente las características antes mencionadas. Por tanto, se requiere combinar la aplicación de varias de las técnicas analizadas para mejorar la comprensión del proceso. La aplicación de múltiples técnicas para el diagnóstico del proceso representa un incremento en el costo temporal del análisis y lo hace más complejo.

El manejo del ruido mediante umbrales es la característica de la *Minería Difusa* con mejor impacto en la comprensión. Mientras que en la *Descomposición en Bloques de Construcción*

destacan el manejo de la ausencia de información, la identificación de los patrones de control de flujo y su uso para ofrecer una vista jerárquica. Es necesaria la combinación de estas características en una única herramienta para mejorar integralmente la comprensión del proceso durante la fase de diagnóstico.

El marco de trabajo para la minería de proceso ProM, ofrece un entorno favorable para la implementación de la solución al problema planteado.

## Capítulo 2

# Propuesta de solución

En este capítulo se presenta la técnica desarrollada, denominada *Minería de Variantes*. La técnica propuesta tiene como objetivo construir diferentes variantes de modelos de proceso, a partir de la descomposición en subprocesos. La descomposición en subprocesos se realiza utilizando patrones de control de flujo. Se obtienen diferentes variantes de descomposición, de cada subproceso, descartando el comportamiento poco frecuente y/o asumiendo la presencia de comportamiento, en un ámbito definido por parámetros de configuración.

### 2.1. Minería de Variantes

La Figura 2.1 muestra una representación de la técnica propuesta, nombrada *Minería de Variantes*. Al igual que el resto de las técnicas para el diagnóstico analizadas, utiliza como entrada un registro de eventos.

La *Minería de Variantes* está conformada por cuatro etapas que se ejecutan secuencialmente:

1. Pre-procesamiento del registro de eventos.
2. Extracción de comportamiento.
3. Obtención de variantes.
4. Generación del perfil de diagnóstico.

El objetivo perseguido por la *Minería de Variantes* es construir variantes de proceso a partir de diferentes descomposiciones en subprocesos considerando patrones de control de flujo.

Las variantes de descomposición obtenidas se muestran en un *Árbol de Variantes*, asociado al cual se obtiene un *Perfil de Diagnóstico*.

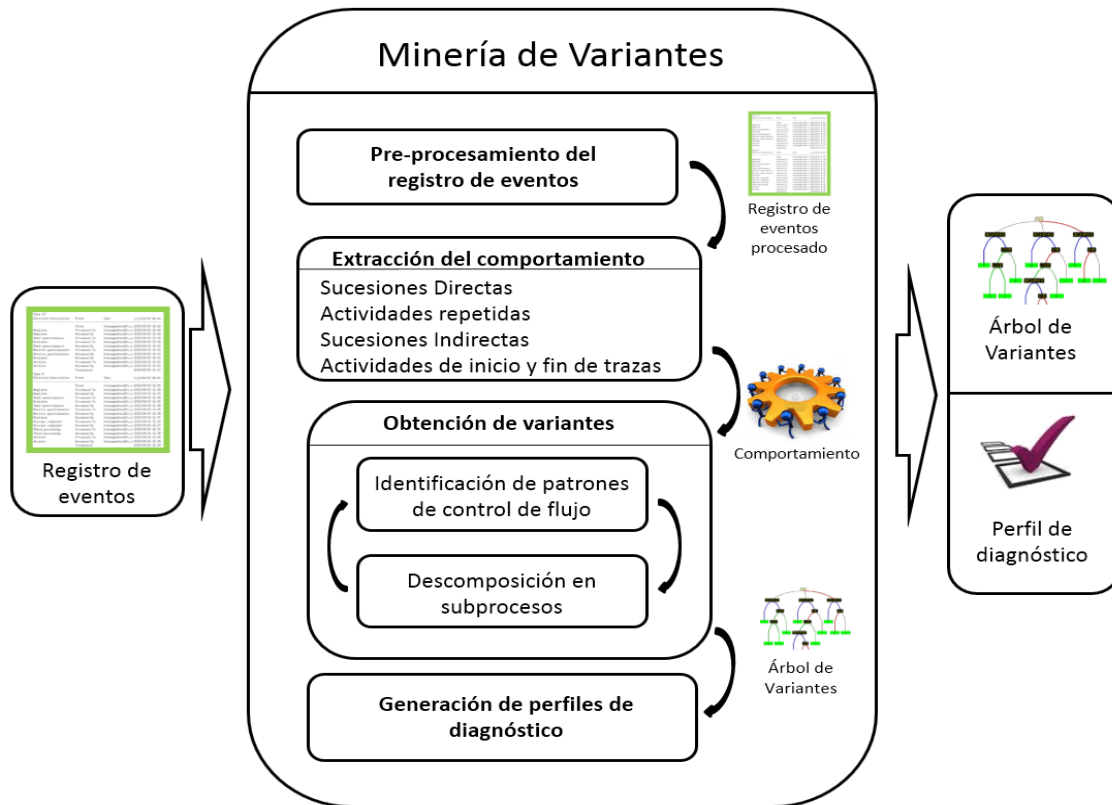


Figura 2.1: Esquema representativo de la técnica Minería de Variantes.

Para comprender el *Árbol de Variantes* que ofrece la técnica como resultado, es relevante conocer que la descomposición que se realiza en la técnica se basa en la Definición 2.1.1.

**DEFINICIÓN 2.1.1** Descomposición en subprocesos: Se denota con  $S_P$  al conjunto de los subprocesos del proceso  $P$ , y por  $W = \{w_s, w_x, w_o, w_p, w_l\}$  se denota al conjunto de los patrones de control de flujo. Un subproceso  $s_i \in S_P$  se puede descomponer a través de diferentes patrones de control de flujo,  $w_j \in W$ , en varios subprocesos,  $s_k \in S_P : k \neq i$ , hasta el nivel de actividad.

La solución propuesta sigue un enfoque diferente al de otras técnicas de diagnóstico, al proponer varias descomposiciones alternativas para el mismo subproceso, utilizando diferentes patrones de control de flujo. Esto permite controlar el impacto estructural del ruido y la ausencia de información en la construcción de las alternativas. Las alternativas se construyen,

descartando o no, determinados comportamientos presentes en el registro de eventos. También se pueden construir, considerando o no, determinados comportamientos ausentes. Las diferentes alternativas de descomposición que pueden existir en cada subproceso conforman variantes del proceso, concepción expuesta en la Definición 2.1.2.

DEFINICIÓN 2.1.2 Variantes de proceso: *Las variantes de un modelo de procesos o variantes de proceso, son modelos de un proceso que describen el mismo proceso de negocio, y poseen algunas diferencias estructurales. Las diferencias están dadas por los patrones de control de flujo que se utilizan en secciones equivalentes del proceso y la presencia de determinadas actividades.*

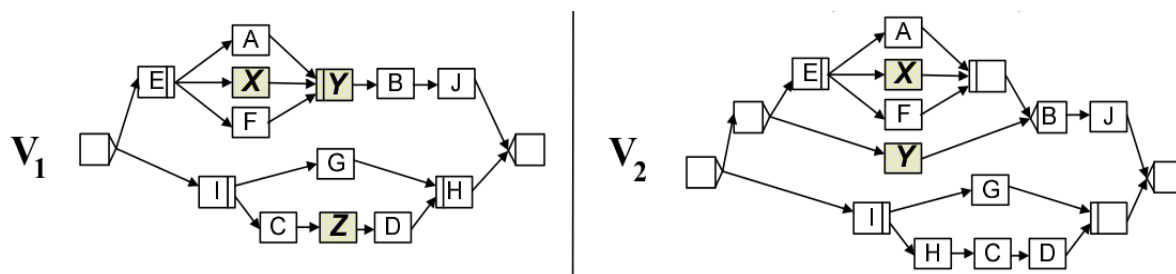


Figura 2.2: Variantes de proceso [Li et al., 2011a]

En la Figura 2.2, pueden apreciarse dos variantes (**V1** y **V2**) de un proceso de negocio. En **V1**, la actividad Y se muestra en secuencia con el subproceso conformado por las actividades E, A, X y F; por otro lado en **V2**, la actividad Y aparece en selección exclusiva con el subproceso antes mencionado. Además la actividad Z, aparece en **V1**, pero no en **V2**.

Los elementos antes planteados permiten que la técnica ofrezca un modelo jerárquico en forma de árbol, que representa diferentes variantes del proceso cuya ejecución está contenida en el registro de eventos. Para esto se propone un modelo de proceso denominado *Árbol de Variantes*, formalizado en la Definición 2.1.3. Este árbol está compuesto por dos tipos de nodos, los nodos subproceso y los nodos patrón. Un nodo subproceso representa un subproceso y posee tantos nodos patrones de hijos, como posibles descomposiciones se hayan identificado para el subproceso. Un nodo patrón representa una descomposición de su padre, de acuerdo a un patrón de control de flujo, por lo que un nodo patrón posee dos o más nodos subproceso como hijos. El nodo raíz es un nodo subproceso y se refiere a todo el proceso. Los nodos hojas son siempre nodos de tipo subproceso.

DEFINICIÓN 2.1.3 **Árbol de variantes:** *Un árbol de variantes ( $V_P$ ) es una representación de diferentes descomposiciones en subprocesos aplicadas al proceso  $P$  a partir de un registro de eventos  $L_P$ .*

*Denotando a  $N_V$  como el conjunto de todos los nodos de  $V_P$  y por  $\Gamma_V \subset N_V$  el conjunto de todos los nodos hoja. Siendo  $n_{s_i} \in N_V : 0 < i \leq |S_P|$ , un nodo subproceso que representa al subproceso  $s_i \in S_P$  y  $l_i$  la sección del registro de eventos asociada a  $s_i$ . Siendo  $n_{i w_j} \in N_V \wedge n_{i w_j} \notin \Gamma_V : 1 < j \leq |W|$ , un nodo patrón que representa una descomposición del subproceso de acuerdo a cierto patrón de control de flujo  $w_j \in W$ .*

*El nodo raíz de  $V_P$  es  $n_{s_1}$  ya que  $s_1 \in S_P \wedge s_1 = P$ . Un nodo subproceso  $n_{s_i}$  puede tener ninguno o varios nodos hijos  $n_{i w_j}$ . Un nodo  $n_{i w_j}$  tiene dos o más nodos hijo  $n_{s_i}$ .*

En la Figura 2.3 se muestran las variantes de descomposición para un mismo subproceso, que forman parte de un *Árbol de variantes*.

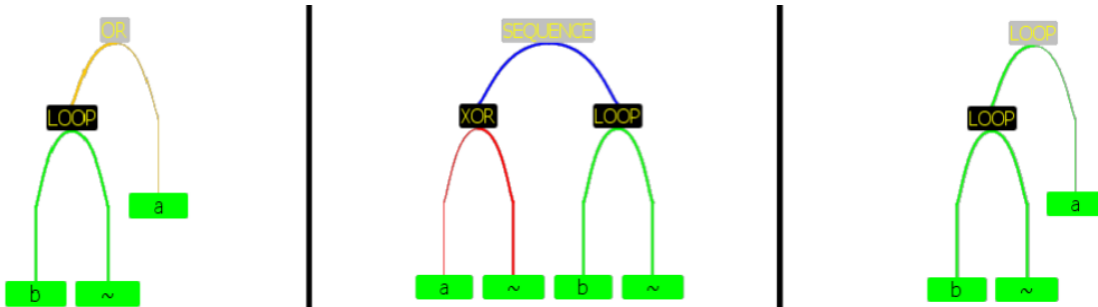


Figura 2.3: Variantes contenidas en un *Árbol de variantes*.

La primera variante corresponde al patrón *Selección no Exclusiva*. La segunda variante representa una descomposición por el patrón de control de flujo *Secuencia*. Por último, se muestra la descomposición del subproceso utilizando el patrón *Lazo*. Los nodos hojas representan las actividades que conforman el proceso. Los nodos patrón, muestran el nombre del patrón de control de flujo utilizado para descomponer el subproceso en la variante correspondiente. Cada nivel en el árbol, de la raíz a las hojas, representa un nivel de abstracción inferior.

Asociado al *Árbol de Variantes* se propone un *Perfil de Diagnóstico* (Definición 2.1.4), el cual agrupa la información de diagnóstico asociada a cada nodo de tipo patrón del árbol.

DEFINICIÓN 2.1.4 **Perfil de Diagnóstico:** *Se denota por  $D_V$  al perfil de diagnóstico del árbol  $V_P$  tal que:  $\forall n_{i w_j} \in N_V \exists d_{i w_j} \in D_V$ .*

*Denotando por  $f$  a la estimación de aptitud, por  $\rho$  a la estimación de precisión, el comportamiento descartado como  $\eta_{i w_j}$  y el comportamiento asumido como  $\theta_{i w_j}$ . La información de*



diagnóstico  $d_{iw_j}$  está constituida por la tupla  $\{f, \rho, \eta_{iw_j}, \theta_{iw_j}\}$ .

Los comportamientos  $\eta_{iw_j}$  y  $\theta_{iw_j}$  se refieren al subproceso  $s_i$  del nodo  $n_{s_i}$  padre de  $n_{iw_j}$ .

Se denota por  $\beta_{w_j l_i}$  al comportamiento del patrón  $w_j$  expresado en la sección del registro de eventos  $l_i$ . El valor de  $f$  expresa la relación entre  $\eta_{iw_j}$  y  $\beta_{w_j l_i}$ . De manera similar  $\rho$  expresa la relación entre  $\theta_{iw_j}$  y  $\beta_{w_j l_i}$ .

En el *Perfil de Diagnóstico* se incluyen los comportamientos descartados, ya que estos pueden ser considerados como ruido, anomalías o desviaciones del proceso. A su vez, los comportamientos asumidos durante la descomposición, reflejan situaciones de ausencia de información, las cuales pueden estar asociadas también a anomalías durante la ejecución del proceso. La presentación de ambos tipos de comportamiento como resultado de la técnica, responde directamente a los intereses del diagnóstico. Se pretende, de esta manera, contribuir a la comprensión del proceso en la dimensión pragmática.

A partir de los comportamientos anteriormente descritos, se realiza una estimación de la calidad del modelo propuesto, en términos de su *Aptitud* y *Precisión*. Dicha estimación se incorpora en el *Perfil de Diagnóstico* con la intención de mejorar la comprensión del proceso, en el sentido de ofrecer información sobre la alineación que existe entre el modelo propuesto y la ejecución real del proceso.

En las secciones siguientes se presentan las etapas que componen la técnica *Minería de Variantes*. Se describen además, los algoritmos fundamentales que se proponen como parte de la solución.

## 2.2. Pre-procesamiento del registro de eventos

Esta etapa tiene como objetivo preparar el registro de eventos para el diagnóstico de variantes. El algoritmo que se ejecuta en esta etapa realiza la codificación de los eventos presentes en el registro de eventos. La codificación consiste en la asignación de una letra a cada evento contenido en el registro. Esto tiene como objetivo disminuir la carga cognitiva durante la comprensión del proceso, eliminando la información semántica de las etiquetas de los eventos. Mediante la codificación se pretende mejorar la comprensión sintáctica del modelo, antes de pasar a la comprensión semántica del mismo. Adicionalmente, la conversión de las trazas en secuencias codificadas facilita el procesamiento subsecuente.

Posterior a la codificación se unifican las trazas cuyas secuencias coinciden, disminuyendo

el número de trazas a procesar y agrupando los comportamientos coincidentes. Las secuencias resultantes de la agrupación son ordenadas descendientemente, por su frecuencia de aparición, para expresar su importancia relativa dentro del proceso.

Por último se extraen características generales presentes en el registro de eventos, como la cantidad de trazas diferentes y la cantidad de eventos diferentes.

## 2.3. Extracción de comportamiento

El propósito de esta etapa es la extracción de comportamientos representativos de los patrones de control de flujo, a partir del registro de eventos preprocesado. Para la técnica se considera el comportamiento del proceso según la Definición 2.3.1.

**DEFINICIÓN 2.3.1** Comportamiento del proceso: *Se denota por  $l_i \subseteq L_P$  a la sección del registro de eventos relacionada con el subproceso  $s_i \in S_P$ . El comportamiento del patrón de control de flujo  $w_j$  para el subproceso  $s_i$ , denotado por  $\beta_{w_j s_i} : w_j \in W$  está compuesto por todas las posibles:*

- *Sucesiones directas (DS), se denomina sucesión directa entre la actividad  $a$  y la actividad  $b$ , a la existencia en alguna traza de la secuencia  $ab$ .*
- *Sucesiones indirectas (IS), se denomina sucesión indirecta entre la actividad  $a$  y la actividad  $b$ , a la aparición en alguna traza de la actividad  $a$  seguida, inmediatamente o no, por la actividad  $b$ .*
- *Eventos que inician y/o finalizan trazas.*
- *Eventos repetidos en las trazas.*

*Siendo  $l_i$  la sección del registro de eventos relacionada con el subproceso  $s_i$ , el subconjunto de este comportamiento observable en  $l_i$  es el comportamiento del proceso  $w_j$  en  $l_i : \beta_{w_j l_i} \subseteq \beta_{w_j s_i} | w_j \in W$ . Por tanto, el comportamiento presente en la sección del registro de eventos  $l_i$  es  $\beta_{l_i} = \beta_{w_s l_i} \cup \beta_{w_x l_i} \cup \beta_{w_o l_i} \cup \beta_{w_p l_i} \cup \beta_{w_l l_i}$ .*

Por cada comportamiento extraído se obtiene la frecuencia de su aparición en las trazas codificadas. Esta frecuencia es utilizada para establecer cuán representativo es un comportamiento del proceso que lo genera.

Los comportamientos extraídos en esta fase se combinan durante la obtención de variantes de descomposición. La extracción de comportamiento permite identificar, con mayor precisión que a nivel de traza, los elementos de ejecución del proceso, cuya baja frecuencia de ejecución sugiere que deban ser considerados como ruido.

## 2.4. Obtención de variantes de descomposición

En la tercera etapa se obtienen las variantes de descomposición del proceso para su representación en un *Árbol de Variantes*. Las variantes se obtienen a partir de la combinación de los comportamientos extraídos en la fase anterior y la asunción de ciertos comportamientos ausentes del registro de eventos. En esta fase se ejecutan dos algoritmos iterativamente, uno para la descomposición en subprocesos y otro para la identificación de patrones de control de flujo.

Las posibles variantes de descomposición de un subproceso se determinan dentro de un ámbito definido como parámetros de entrada. El ámbito se especifica en términos de umbrales para ruido ( $T_n$ ) y completitud ( $T_c$ ). A partir de un registro de eventos y un ámbito definido se genera el árbol de variantes.

Los umbrales para ruido y completitud se definen como un número entre 0 y 100 y pueden ser especificados individualmente para cada patrón de control de flujo. Los valores por defecto están basados en el Principio de Pareto [Leemans et al., 2013], siendo 20 % para ruido y 80 % para completitud. Por ejemplo, un umbral de 20 % para ruido en el patrón secuencia significa que hasta el 20 % del comportamiento expresado en el registro de eventos puede ser descartado para encontrar una descomposición mediante el patrón secuencia. De manera similar, un 80 % de completitud para paralelismo indica que el subproceso puede ser descompuesto mediante ese patrón conteniendo solamente el 80 % del comportamiento necesario para ser identificado.

### 2.4.1. Descomposición en subprocesos

El algoritmo de descomposición en subproceso (Algoritmo 1) utiliza un registro de eventos del proceso  $P$  y los umbrales definidos para construir un árbol de variantes  $V_P$ . Siendo  $S_P$  el conjunto de todos los subprocesos del proceso  $P$ , para cada nuevo subproceso  $s_i \in S_P$ , se busca una variante de descomposición según cada uno de los patrones de control de flujo.

Para encontrar una variante de descomposición para un subproceso  $s_i$ , se extrae el comportamiento del proceso del sublog  $l_i$  relacionado con  $s_i$ . Denotemos como  $\beta'_{w_j s_i}$  al comportamiento del proceso para el patrón  $w_j$  en  $s_i$  que no está en  $\beta_{l_i}$  pero puede ser asumido dentro del umbral de completitud. Sea  $\beta'_{l_i} \subset (\beta_{l_i} \setminus \beta_{w_j l_i})$  el comportamiento del proceso que está presente en  $\beta_{l_i}$  y puede ser descartado dentro del umbral de ruido. Para un patrón de control de flujo  $w_j$  es posible encontrar una variante si  $(\beta_{w_j l_i} \setminus \beta'_{l_i}) \cup \beta'_{w_j s_i}$  es equivalente a  $\beta_{w_j s_i}$ .

El nuevo subproceso generado por cada descomposición es añadido a  $S'_P$  con el objetivo de que sea analizado posteriormente. Un subproceso  $w_j$  es analizado solo si el conjunto de actividades que contiene, denotado por  $A_{s_i}$ , tiene más de una actividad.

---

### Algoritmo 1 Descomposición en subprocesos

---

**Entrada:** registro de eventos ( $L_P$ ), umbrales de ruido y completitud ( $T_P$ )

**Salida:** árbol variantes ( $V_P$ ), perfil de diagnóstico

```

1: Procedimiento DECOMPOSE( $L_P, T_P$ )  $W \leftarrow \{w_s, w_x, w_o, w_p, w_l\}$   $\triangleright$  Patrones de control de
   flujo para la descomposición
2:    $n_{s_1} \leftarrow$  CREATESUB-PROCESS( $L_P$ )  $\triangleright$  Crea el nodo raíz.
3:    $V_P \leftarrow n_{s_1}$ 
4:    $S'_P \leftarrow \{n_{s_1}\}$   $\triangleright$  Crea una lista para los subprocesos nuevos.
5:   while  $n_{s_i} \leftarrow$  extraer el primer nodo subproceso existente en  $S'_P$  do
6:     if  $s_i$  no es una actividad then
7:        $\beta_{l_i} \leftarrow$  EXTRACTBEHAVIOR( $n_{s_i}$ )  $\triangleright$  Extrae el comportamiento contenido en  $l_i$ 
8:       for all  $w_j$  en  $W$  do
9:          $V \leftarrow$  FINDVARIANTS( $w_j, \beta_{l_i}, T_P$ )
10:        if  $|V| > 0$  then  $\triangleright$  If  $n_{s_i}$  puede ser descompuesto por  $w_j$  bajo  $T_P$ 
11:          for all nodo subproceso  $k$  en  $V$  do
12:            adicionar el nodo subproceso  $k$  como hijo del nodo patrón  $n_{iw_j}$ 
13:            adicionar el nodo subproceso  $k$  to  $S'_P$ 
14:          end for
15:          adicionar el nodo  $n_{iw_j}$  como hijo del nodo subproceso  $n_{s_i}$ 
16:           $d_{iw_j} \leftarrow$  CREATEDIAGNOSISINFORMATION( $w_s, l_i, V$ )
17:          adicionar  $d_{iw_j}$  al perfil de diagnóstico.
18:        end if
19:      end for
20:    end if
21:  end while
22: end Procedimiento

```

---

#### 2.4.2. Identificación de patrones de control de flujo

La identificación de los patrones de control de flujo se realiza a partir de la combinación de comportamientos presentes en el registro de eventos y comportamientos asumidos bajo el

umbral de ausencia de información. Para encontrar la combinación de comportamientos que puede conducir a la descomposición de un subproceso por un determinado patrón de control de flujo se utiliza el procedimiento que se expone a continuación (Algoritmo 2).

### Búsqueda de variantes

Siendo  $\Upsilon_{w_j s_i}$  una posible descomposición para el subproceso  $s_i$  utilizando el patrón de control de flujo  $w_j$  y  $A_{s_i}$  el conjunto de actividades presentes en  $s_i$ , cualquier  $\Upsilon_{w_j s_i}$  contiene  $n$  conjuntos disjuntos de actividades, pertenecientes a  $A_{s_i}$ . Por cada patrón de control de flujo  $w_j$  y cada sección del registro de eventos  $l_i$ , cada elemento de comportamiento  $b \in \beta_{w_j l_i}$  es procesado, en una búsqueda en orden descendente según la frecuencia de  $b$  (denotada por  $b(\varphi(b))$ ).

La búsqueda de variantes está basada en la Búsqueda de costo uniforme [Russell et al., 1995]. Este método de búsqueda expande el nodo que posea el menor costo de camino. En caso de que los costos de todos los pasos de un nodo a otro sean iguales, es idéntico a una búsqueda a lo ancho. Si el costo de cada paso es menor o igual que una constante  $\varepsilon$ , se garantiza la completitud y optimalidad del método. La complejidad temporal y espacial del peor caso de este método de búsqueda pueden ser descritas por la ecuación  $r^{1+C/\varepsilon}$ , donde  $r$  es el factor de ramificación del árbol y  $C$  es el costo del camino de la solución óptima.

---

#### Algoritmo 2 Búsqueda de variantes

---

**Entrada:** patrón de control de flujo ( $w_j$ ), comportamiento del proceso de la sección del registro de eventos  $l_i(\beta_{l_i})$ , umbrales de ruido y completitud ( $T_P$ )

**Salida:** variante de descomposición en subprocesos a través del patrón  $w_j$

```

1: Procedimiento FINDVARIANTS( $w_j, \beta_{l_i}, T_P$ )
2:    $\beta'_{w_j l_i} \leftarrow \text{SORTBYFREQUENCY}(\beta_{l_i}[w_j])$            ▷ Ordenando comportamiento en orden
   descendente
3:    $\tau_{w_j} \leftarrow T_P[w_j]$                                    ▷ Umbrales para el patrón  $w_j$ 
4:    $\Upsilon_{w_j s_i} \leftarrow \text{CREATEFIRSTDECOMPOSITIONPROPOSAL}(w_j, \beta_{w_j l_i}, \tau_{w_j})$ 
5:   if  $\Upsilon_{w_j s_i} = \emptyset$  then ▷  $\beta_{w_j l_i}$  no es suficiente para crear una propuesta de descomposición
   bajo  $\tau_{w_j}$ 
6:     return  $\emptyset$ 
7:   end if
8:    $\psi_1 \leftarrow \text{CREATESearchNode}(\beta_{w_j l_i}, \Upsilon_{w_j s_i})$ 
9:    $\Psi \leftarrow \{\psi_1\}$                                        ▷ Cola con prioridad de nodos abiertos de acuerdo a  $c$ 
10:  while  $\psi_n \leftarrow \text{extract first node on } \Psi$  do
11:    if  $\psi_n$  is goal &  $\beta'_{w_j l_i} = \emptyset$  then

```

---

---

```

12:      $g \leftarrow \psi_n$ 
13:     else if  $\psi_n$  is not closed then
14:          $\Psi_n \leftarrow \text{SEARCHNEIGHBORS}(w_j, \psi_n, \beta'_{w_j l_i}, \tau_{w_j})$ 
15:          $\Psi \leftarrow \Psi + \Psi_n$  ▷ Adicionar vecinos como nodos abiertos
16:     end if
17: end while
18: if  $g$  is not null then
19:      $\Upsilon_{w_j s_i} \leftarrow$  descomposición potencial en  $g$ 
20:     return  $\text{CREATENEWSUB-PROCESSES}(w_j, \Upsilon_{w_j s_i}, \beta_{l_i})$ 
21: else
22:     return  $\emptyset$ 
23: end if
24: end Procedimiento

```

---

En la búsqueda de variantes cada nodo en el espacio de búsqueda está definido por  $\Upsilon_{w_j s_i}$ , el comportamiento no procesado ( $\beta'_{w_j l_i} \subseteq \beta_{w_j l_i}$ ), el comportamiento descartado ( $\eta_{iw_j}$ ) y el comportamiento asumido ( $\theta_{iw_j}$ ).

La creación de nuevos nodos de búsqueda, descrita en el Algoritmo 3, se realiza utilizando tres operadores. Se crea un nuevo nodo cuando, para  $\Upsilon_{w_j s_i}$ , una  $\Upsilon'_{w_j s_i}$  diferente puede ser creada al procesar un comportamiento  $b_x \in \beta'_{w_j l_i}$ . También un nuevo nodo es creado si el comportamiento  $b_x$  puede generar una  $\Upsilon'_{w_j s_i}$  diferente, pero descartando  $b_x$ , o algún comportamiento  $b_y \in \beta_{w_j}$  puede ser asumido con el objetivo de mantener  $\Upsilon_{w_j s_i}$ .

---

### Algoritmo 3 Buscar vecinos

**Entrada:** patrón de control de flujo ( $w_j$ ), nodo de búsqueda actual ( $\psi_0$ ), comportamiento por procesar ( $\beta'_{w_j l_i}$ ), umbrales de ruido y completitud para el patrón  $w_j$  ( $\tau_{w_j}$ )

**Salida:** variante de descomposición en subprocesos a través del patrón  $w_j$

```

1: Procedimiento  $\text{SEARCHNEIGHBORS}(w_j, \psi_0, \beta'_{w_j l_i}, \tau_{w_j})$ 
2:    $\Upsilon_{w_j s_i} \leftarrow$  descomposición potencial en  $\psi_0$ 
3:    $\Upsilon'_{w_j s_i} \leftarrow \Upsilon_{w_j s_i}$ 
4:    $\Psi_n \leftarrow \{\}$  ▷ Crea el conjunto vacío de nodos vecinos
5:   while  $|\beta'_{w_j l_i}| > 0$  &  $\Upsilon'_{w_j s_i} \neq \Upsilon_{w_j s_i}$  do
6:      $b_x \leftarrow \beta'_{w_j l_i}$ 
7:      $\beta'_{w_j l_i} \leftarrow \beta'_{w_j l_i} \setminus b_x$ 
8:      $\Upsilon'_{w_j s_i} \leftarrow \text{PROCESSBEHAVIORBYPATTERN}(\Upsilon'_{w_j s_i}, b_x)$ 
9:   end while
10:  if  $\Upsilon'_{w_j s_i} \neq \Upsilon_{w_j s_i}$  then ▷ Si  $\Upsilon'_{w_j s_i}$  es nuevo
11:    if  $|\Upsilon'_{w_j s_i}| > 1$  then ▷ Si  $\Upsilon'_{w_j s_i}$  es válido
12:       $\psi_1 \leftarrow \text{CREATESEARCHNODE}(\beta'_{w_j l_i}, \Upsilon'_{w_j s_i})$ 
13:       $(\psi_1 \rightarrow td) \leftarrow (\psi_0 \rightarrow td)$  ▷ El costo del camino permanece igual

```

---

---

```

14:      $\Psi_n \leftarrow \Psi_n + \psi_1$                                 ▷ Adicionar nodo hijo "ajustado"
15:     end if
16:     if  $(c(b_x) + \psi_0 \rightarrow td) \leq \tau_{w_j}$  then          ▷ Si  $b_x$  puede ser descartado
17:          $\psi_2 \leftarrow \text{CREATESEARCHNODE}(\beta'_{w_j l_i}, \Upsilon_{w_j s_i})$ 
18:          $(\psi_2 \rightarrow td) \leftarrow (\psi_0 \rightarrow td + c(b_x))$ 
19:          $(\psi_2 \rightarrow \eta_{i w_j}) \leftarrow b_x$                                 ▷  $\psi_2$  descartar  $b_x$ 
20:          $\Psi_n \leftarrow \Psi_n + \psi_2$                                 ▷ Adicionar nodo "ruidoso"
21:     end if
22:      $b_y \leftarrow \text{PROCESSBEHAVIORTAKINGONINCOMPLETENESS}(w_j, b_x, \Upsilon_{w_j s_i})$ 
23:     if  $b_y$  is not null &  $(c(b_y) + \psi_0 \rightarrow td) \leq \tau_{w_j}$  then
24:          $\psi_3 \leftarrow \text{CREATESEARCHNODE}(\beta'_{w_j l_i}, \Upsilon_{w_j s_i})$ 
25:          $(\psi_3 \rightarrow td) \leftarrow (\psi_0 \rightarrow td + c(b_y))$ 
26:          $(\psi_3 \rightarrow \theta_{i w_j}) \leftarrow b_y$                                 ▷  $\psi_3$  assume  $b_y$ 
27:          $\Psi_n \leftarrow \Psi_n + \psi_3$                                 ▷ Adicionar nodo "incompleto"
28:     end if
29: end if
30: return  $\Psi_n$ 
31: end Procedimiento

```

---

El costo del camino desde un nodo a alguno de sus vecinos es  $c = c(b_x) + c(b_y)$ , donde:

$$c(b_x) = \frac{\varphi(b_x)}{|\beta_{w_j l_i}|}, \text{ si } b_x \in \beta'_{w_j l_i} \text{ es descartado y } c(b_y) = \frac{\varphi(b_y)}{|\beta_{w_j}|}, \text{ si } b_y \in \beta_{w_j} \text{ es asumido. El}$$

$$\sum_{n=1} \varphi(b_n) \qquad \sum_{n=1} \varphi(b_n)$$

objetivo se alcanza cuando  $\beta_{w_j l_i}$  ha sido procesado completamente y la posible descomposición  $\Upsilon_{w_j s_i}$  posee más de un conjunto disjunto.

La optimalidad y completitud de la búsqueda se garantizan con el costo asignado a los nodos generados. El procesamiento de los comportamientos en orden descendente de su frecuencia garantiza que se procesen primero aquellos que no pueden ser descartados según el umbral. Esto también garantiza la exploración de los estados más prometedores primero: aquellos cuyos costos de camino son menores descartaron comportamientos poco frecuentes y poseen menos comportamiento por procesar. Las dimensiones del espacio de búsqueda pueden ser controladas por el usuario, a través de los umbrales de ruido y completitud.

En cada nodo de búsqueda se procesa paulatinamente el comportamiento del patrón presente en la sección del registro de eventos relativa al subproceso que se intenta descomponer. A continuación se exponen los principios básicos que guían este procesamiento para cada uno de los patrones de control de flujo que se intentan identificar.

### Secuencia:

El objetivo de este procedimiento es determinar si un subproceso puede ser descompuesto en subprocesos ordenados secuencialmente. El comportamiento procesado por este patrón son las sucesiones directas. Los subprocesos  $s_i$  y  $s_{i+i}$  están en secuencia si todas las  $DS$  dirigidas hacia actividades de  $s_{i+i}$  son solo de actividades de  $s_i$  o en  $s_{i+i}$  no existen relaciones directas hacia  $s_i$ .

La búsqueda se realiza partiendo de una posible descomposición, donde se encuentra un subproceso el cual contiene todas las actividades presentes en  $l_i$ . El conjunto de actividades es procesado para crear el primer subproceso en secuencia con otro subproceso que contiene el resto de las actividades. Este procedimiento se realiza iterativamente hasta que todos los subprocesos son analizados.

### Lazo:

El objetivo de este procedimiento es determinar si un subproceso puede descomponerse en dos subprocesos, uno denominado **Do** y el otro denominado **Redo** como parte de un lazo. Los comportamientos procesados para la identificación de este patrón son las  $DS$ , las actividades que inician y concluyen trazas y las actividades repetidas. Los subprocesos **Do** y **Redo** están en lazo si se cumplen las siguientes condiciones:

- Todas las actividades que inician o terminan trazas pertenecen al **Do** y deben ser actividades repetidas.
- En todas las  $DS$  cuya actividad predecesora está en el **Do** y sucesora aparece en el **Redo**, la actividad predecesora debe iniciar al menos una traza.
- En todas las  $DS$  cuya actividad predecesora está en el **Redo** y sucesora aparece en el **Do**, la actividad sucesora debe terminar al menos una traza.

La primera descomposición posible contiene como parte del **Do** a todas las actividades que inician o terminan trazas con una frecuencia por encima del umbral de ruido. Durante el procesamiento de las  $DS$  el resto de las actividades son adicionadas al **Do** o al **Redo**. Una nueva posible descomposición de lazo puede ser creada al asumir algunas actividades como inicio o fin de trazas dentro del umbral para completitud.



### **Paralelismo:**

El objetivo de este procedimiento es determinar si un subproceso puede ser descompuesto en subprocesos que se ejecutan concurrentemente. El comportamiento procesado son las sucesiones indirectas (*IS*). Los subprocesos *A* y *B* forman parte de un paralelismo si para cada actividad **a** que pertenece al subproceso *A* y para cada actividad **b** que pertenece a *B* existe una sucesión indirecta desde **a** hasta **b** y una desde **b** hasta **a**. Además, en cada una de las trazas debe aparecer una representación de los subprocesos *A* y *B*.

La primera descomposición posible de paralelismo contiene cada una de las actividades, como parte de un subproceso, en paralelo con el resto. El procesamiento de las *IS* en las cuales se violan las condiciones planteadas, provoca la unión de los subprocesos en los que se encuentran las actividades presentes en la *IS* procesada. La ausencia de representación de los subprocesos en paralelo, en algunas de las trazas, puede ser ignorada dentro del umbral de ruido. Una nueva descomposición puede ser generada asumiendo, dentro del umbral de completitud, la existencia de alguna *IS*.

### **Selección Exclusiva:**

Este procedimiento determina si un subproceso puede ser descompuesto mediante el patrón de Selección Exclusiva. Los subprocesos *A* y *B* están en selección exclusiva si no existen *IS* desde **a** hasta **b** o desde **b** hasta **a**, para cada actividad **a** que pertenece al subproceso *A* y cada actividad **b** que pertenece a *B*.

La primera descomposición posible contiene a cada una de las actividades como un subproceso. El procesamiento de las *IS*, en las cuales se violan las condiciones planteadas, provoca la unión de los subprocesos a los que pertenecen. Es posible descartar determinadas *IS* dentro del umbral de ruido.

### **Selección no Exclusiva:**

Este procedimiento determina si un subproceso puede ser descompuesto mediante el patrón de Selección no Exclusiva. Los subprocesos *A* y *B* están en selección no exclusiva si para cada actividad **a** perteneciente a *A* y para actividad **b** que pertenece a *B* existe una *IS* desde **a** hasta **b** y una *IS* desde **b** hasta **a**. Además, debe existir representación de solo uno de los subprocesos en al menos una traza.

La primera descomposición posible para este patrón contiene un subproceso por cada una de las actividades. El procesamiento de las *IS* en las cuales se violan las condiciones planteadas, provoca la unión de los subprocesos a los que pertenecen. La asunción de la existencia de alguna *IS* o de la presencia en una traza de solo uno de los subprocesos. Pueden ser creadas nuevas descomposiciones asumiendo la existencia de alguna *IS* o la presencia en una traza de solo uno de los subprocesos dentro el umbral de completitud.

## **2.5. Generación del perfil de diagnóstico**

Esta es la última etapa que se ejecuta dentro de la técnica. El algoritmo implementado en esta etapa recorre el árbol de búsqueda generado y recopila los comportamientos considerados como ruido y ausencia de información durante la identificación de los patrones. A partir de estos comportamientos asumidos o descartados se establecen la estimación de aptitud ( $f$ ) y precisión ( $\rho$ ). El perfil se conforma con la información de diagnóstico de cada nodo de tipo patrón en el *Árbol de Variantes* obtenido.

## **2.6. Conclusiones parciales**

Durante este capítulo se presentó una técnica para el diagnóstico de procesos denominada Minería de Variantes. La técnica permite la identificación de los patrones de control de flujo Secuencia, Lazo, Selección no Exclusiva, Selección Exclusiva y Paralelismo.

La Minería de Variantes posibilita la obtención de información relevante del proceso y la identificación de características del registro de eventos como el ruido y la ausencia de información. La identificación de los patrones de control de flujo presentes en cada nivel de jerarquización posibilita el análisis del comportamiento presente en el registro de eventos y ubicarlo en el contexto de su ejecución. Diferentes variantes de un proceso pueden ser obtenidas al asumir comportamiento por concepto de completitud o al descartar comportamiento por concepto de ruido.

## Capítulo 3

# Validación de la solución propuesta

En este capítulo se presenta la implementación de la técnica propuesta como complemento para el marco de trabajo ProM. Haciendo uso de la aplicación informática desarrollada se evalúan un conjunto de procesos y se analizan los resultados obtenidos.

### 3.1. Aplicación informática desarrollada

La técnica propuesta, *Minería de Variantes* ha sido implementada como complemento del marco de trabajo ProM. El complemento desarrollado se denomina *Variants Miner* (Figura 3.1). Utiliza como entrada un registro de eventos y genera como salida un *Árbol de variantes*, con su respectivo *Perfil de Diagnóstico*. Los umbrales de ruido y completitud necesarios para definir el ámbito de la búsqueda de variantes son configurados por el usuario al iniciar el complemento.

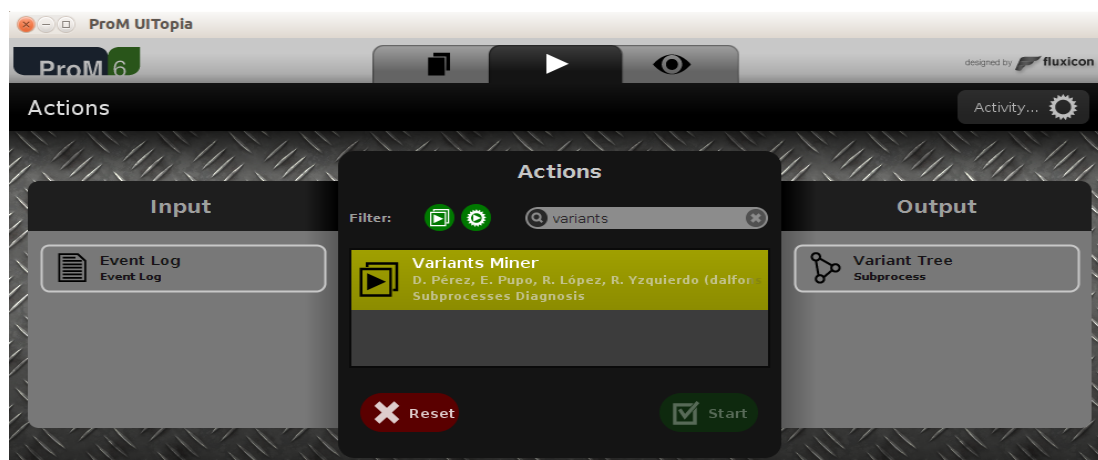


Figura 3.1: *Variants Miner*, complemento para ProM basado en la técnica propuesta.

La Figura 3.2 muestra el *Árbol de Variantes* en el panel central de la herramienta y la información del *Perfil de Diagnóstico* entre el panel de la derecha y el panel inferior.

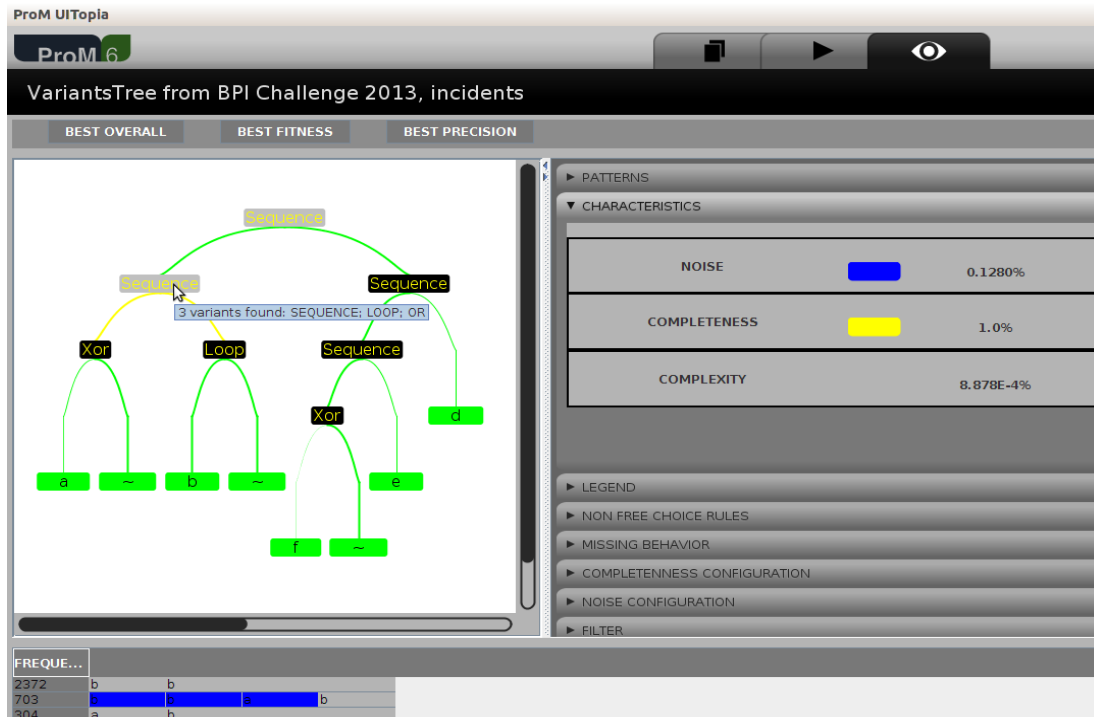


Figura 3.2: Visualización del *Árbol de Variantes* y el *Perfil de Diagnóstico*.

Para la representación del *Árbol de Variantes* en la herramienta, se muestra en cada momento una única variante del proceso. En cada nodo visible se combinan, el nodo subprocesso relativo al subprocesso del nivel correspondiente del árbol, con el nodo patrón relativo a la descomposición seleccionada. En un mensaje contextual se muestran las descomposiciones alternativas del subprocesso, utilizando otros patrones de control de flujo. Por ejemplo, en la Figura 3.2, se puede apreciar la variante del proceso, en la cual al subprocesso de la izquierda, del segundo nivel, se le ha aplicado una descomposición utilizando el patrón *Secuencia*. En el mensaje contextual se muestra que existen dos descomposiciones alternativas asociadas a ese nodo, una por *Lazo* y otra por *Selección*.

Las aristas del árbol representan caminos dentro del proceso. El grosor de estas aristas depende de la frecuencia con la cual aparece en el registro de eventos el camino que representan. Los nodos del árbol pueden ser contraídos o expandidos a voluntad del usuario. También pueden realizarse acercamientos sobre el árbol.

Por cada nodo patrón, se muestra la información de diagnóstico correspondiente que forma parte del *Perfil de Diagnóstico*. En el panel inferior se muestra la sección del registro de eventos relativa al nodo subproceso del cual dicho nodo patrón es hijo. En esta sección se señalan los comportamientos considerados como ruido en la descomposición y la trazas que poseen ausencia de información. A partir de los comportamientos descartados y asumidos se determina el porcentaje de ruido e incompletitud que posee la descomposición propuesta, para el nodo patrón seleccionado. Este porcentaje se visualiza en el panel de la derecha. En la Figura 3.2, para el subproceso antes mencionado, se pueden apreciar en el panel de la derecha los valores de ruido e incompletitud. Mientras que en el panel inferior, aparece marcada la sucesión indirecta b->a como un comportamiento considerado ruido para la descomposición por el patrón *Secuencia*.

### **3.2. Diseño de la validación**

La validación tiene como objetivo demostrar que la técnica propuesta mejora la comprensión del proceso durante la fase de diagnóstico. Para ello se utilizan las características deseadas para la comprensión, propuestas en el Epígrafe 1.2.3. Para la validación se emplean dos enfoques. El primer enfoque está orientado a comprobar la identificación de patrones de control de flujo en registros de eventos artificiales. El segundo está dirigido a comparar integralmente las técnicas de diagnóstico, a partir de las características que impactan en la comprensión del proceso. Para ambos enfoques se utiliza el complemento de ProM donde se implementó la solución propuesta.

### **3.3. Identificación de patrones de control de flujo**

La comprobación de la identificación de los patrones de control de flujo persigue verificar si la técnica propuesta identifica correctamente dichos patrones. Esta verificación complementa la validación de la herramienta en un entorno real, ya que los patrones de control de flujo influyen en la descomposición en variantes. Por ende, la alineación con la realidad tanto del *Árbol de Variantes* como del *Perfil de Diagnóstico* dependen de la identificación correcta de estos patrones. Como se ha analizado con anterioridad, este factor impacta en la dimensión pragmática de la comprensión.

### 3.3.1. Diseño experimental

Para comprobar si la solución propuesta identifica correctamente los patrones de control de flujo se conformaron dos grupos de registros de eventos artificiales y se definieron dos momentos.

El primer momento se refiere a la identificación de los patrones de control de flujo en el modelo original. En el segundo momento se realizó la identificación de patrones de control de flujo a partir del modelo generado por la *Minería de Variantes*. Para el experimento se utilizaron 10 modelos de procesos que combinan diferentes patrones de control de flujo. A partir de estos modelos de proceso se generaron aleatoriamente 20 registros de eventos, divididos en dos grupos de 10. El diseño experimental propuesto se resume en la Tabla 3.1, utilizando la siguiente simbología:

**G:** Grupo de participantes. Cada grupo está formado por 10 registros de eventos, uno de cada modelo. En el segundo grupo se introdujeron situaciones de ruido y ausencia de información.

**R:** Asignación al azar.

**X:** Tratamiento o estímulo.  $X_1$  corresponde con la extracción de eventos de los registros de eventos de  $G_1$ . Por otra parte  $X_2$  corresponde a la aplicación de la *Minería de Variantes*

**O:** Observación.

Tabla 3.1: Diseño experimental propuesto.

	Modelo original		Árbol de Variantes	
R $G_1$	$O_1$		$X_2$	$O_2$
R $G_2$	$O_3$	$X_1$	$X_2$	$O_4$

En el primer momento (Modelo original) las observaciones  $O_1$  y  $O_3$  representan la cantidad de patrones de control de flujo presentes en el modelo. En el segundo momento (Minería de variantes) las observaciones  $O_2$  y  $O_4$  están asociadas a la cantidad de patrones correctamente identificados por la *Minería de Variantes*, con respecto al modelo original. En este punto para cada observación se hacen 10 mediciones, una por cada modelo de proceso.

Para formar los grupos se realizó la extracción de eventos de los registros de eventos, generando ruido y ausencia de información en los mismos. El primer grupo ( $G_1$ ) está formado

por 10 registros de eventos generados aleatoriamente a partir de los modelos originales. A estos registros de eventos se les extrajo el 5 % de los eventos para conformar  $G_2$ .

Se realizaron comparaciones por pares en un grupo, utilizando el test no paramétrico de signos con rangos de Wilcoxon. Se analizaron los datos correspondientes al primer y segundo momento, por ejemplo  $O_1$  y  $O_2$ , buscando detectar un empate entre los valores, lo cual no revelaría diferencias significativas.

Se ejecutaron análisis transversales para comparar las observaciones de los diferentes grupos en un mismo momento, por ejemplo  $O_2$  y  $O_4$ . En este caso, se esperaba no encontrar diferencias significativas al aplicar el test de Mann-Whitney.

### 3.3.2. Características de los registros de eventos generados

Para la generación de los registros de eventos se generaron 10 modelos de proceso. Los modelos de proceso fueron generados con la herramienta *Process Log Generator*, en la versión 1.4 [Burattin and Sperduti, 2011]. Esta herramienta permite la construcción de modelos a partir de la combinación aleatoria de características como: cantidad de patrones anidados, probabilidad de *Secuencia*, *Lazo*, *AND Split/Join* y *XOR Split/Join*. Los modelos de procesos utilizados poseen las características descritas en la Tabla 3.2.

Tabla 3.2: Características de los modelos de proceso construidos.

Características	Modelos									
	1	2	3	4	5	6	7	8	9	10
Actividades	10	8	11	10	15	10	10	6	13	6
Secuencia	7	6	7	7	9	6	7	6	10	6
Lazo	0	0	0	1	0	1	0	0	0	0
Paralelismo	1	1	2	1	1	0	0	1	1	1
Selección exclusiva	1	0	0	0	2	1	2	0	0	0
Total de patrones	9	7	9	9	12	8	9	7	11	7

Por cada uno de los modelos de proceso descritos, se generó un registro de eventos, los cuales fueron utilizados para conformar el grupo  $G_1$ . Para obtener los 10 registros de eventos que conforman  $G_2$  se extrajo, a cada registro de eventos perteneciente a  $G_1$ , el 5% de sus eventos. En la Tabla 3.3 se muestran las características de los registros de eventos de cada grupo.

Tabla 3.3: Características de los registros de eventos de cada grupo.

Registro de eventos	Trazas	Grupo $G_1$		Grupo $G_2$	
		Eventos	Trazas únicas	Eventos	Trazas únicas
1	1500	10492	3	9961	62
2	1500	12000	2	11399	59
3	1500	16500	12	15675	231
4	1500	23096	90	21945	425
5	5000	35192	9	33473	182
6	1500	18849	38	17466	519
7	5000	32453	3	30849	82
8	1500	9000	2	8552	31
9	5000	65000	6	61762	430
10	5000	30000	2	28563	39

### 3.3.3. Análisis de los resultados

En la Figura 3.3 se muestra, por cada modelo, el total de patrones presentes en el modelo original y los identificados por la *Minería de Variantes*, a partir de los registros de eventos de los grupos  $G_1$  y  $G_2$ .

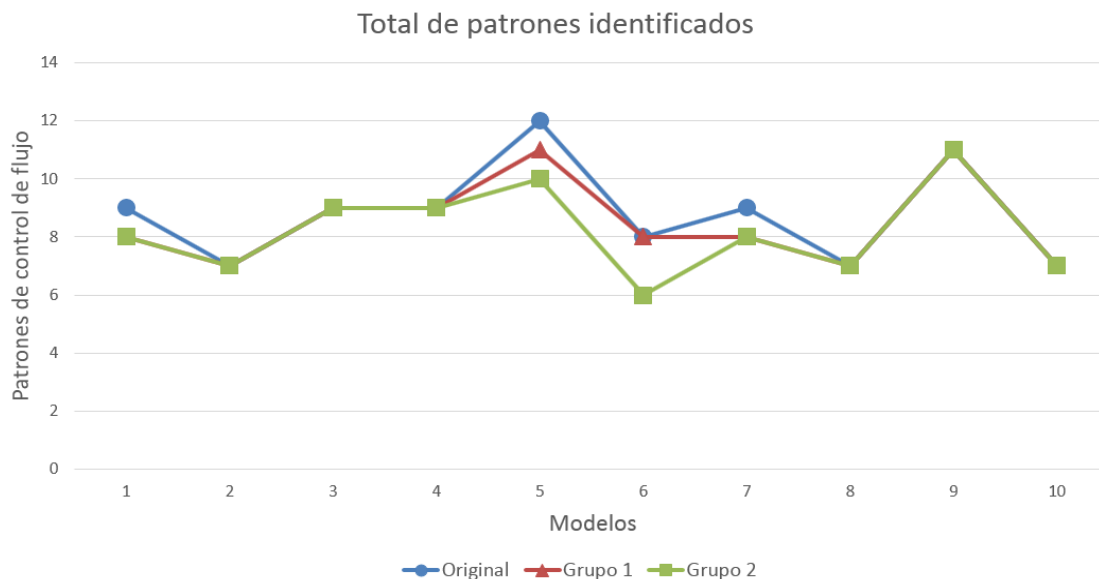


Figura 3.3: Patrones identificados en el experimento.

A partir del diseño experimental expuesto se realizaron un conjunto de pruebas. Primariamente se realizaron comparaciones por pares en un grupo, utilizando el test no paramétrico de signos con rasgos de Wilcoxon. Al analizar los datos correspondientes al primer y segundo



momento se detectó que no existen diferencias significativas (significación 0.083 para  $G_1$  y 0.063 para  $G_2$ ) entre los patrones observados en los modelos originales y los observados en el *Árbol de Variantes*.

Con el objetivo de comprobar el impacto del ruido y la ausencia de información en la técnica propuesta se compararon los valores obtenidos en las evaluaciones  $O_2$  y  $O_4$  utilizando el test de Mann-Whitney. En esta prueba estadística no se encontraron diferencias significativas entre ambos momentos (significación 0.669).

Considerando los resultados obtenidos, se puede afirmar que la técnica propuesta identificó correctamente los patrones de control de flujo a partir de registros de eventos con ruido y/o ausencia de información.

### 3.4. Comprensión del proceso en un entorno real

Para la comparación integral de las técnicas de diagnóstico, a partir de su aporte a la comprensión del proceso, se analiza un proceso que pertenece a un entorno real. Durante el análisis se muestra el impacto que tiene la presencia de cada característica en la comprensión, por cada una de las técnicas de diagnóstico utilizadas.

#### 3.4.1. Aplicación de *Minería de Variantes* a un registro de eventos del sistema *XAVIA HIS*

El Sistema de Información Hospitalaria *XAVIA HIS* es un sistema de gestión que permite la recolección, almacenamiento, procesamiento, recuperación y comunicación de información de atención al paciente para todas las actividades relacionadas con la institución de salud. Está concebido para registrar las actividades de salud orientadas a los pacientes, permitiendo además gestionar y controlar los recursos de cada una de las áreas de dichas instituciones.

La técnica *Minería de Variantes* se ha aplicado a un registro de eventos que corresponde al módulo Almacén del sistema *XAVIA HIS*. Este registro de eventos recopila 1212 ejecuciones del proceso *Solicitar Producto* entre las fechas 09/01/2012 y 27/06/2013. El proceso *Solicitar Producto* cuenta con 8 actividades diferentes con un total de 2400 ocurrencias. El análisis del modelo de referencia, presentado en la Figura 3.4, permite identificar múltiples actividades de fin del proceso y lazos en diferentes puntos del proceso. Los nodos del tipo **Decision** expresan el patrón de control de flujo *Selección exclusiva*.

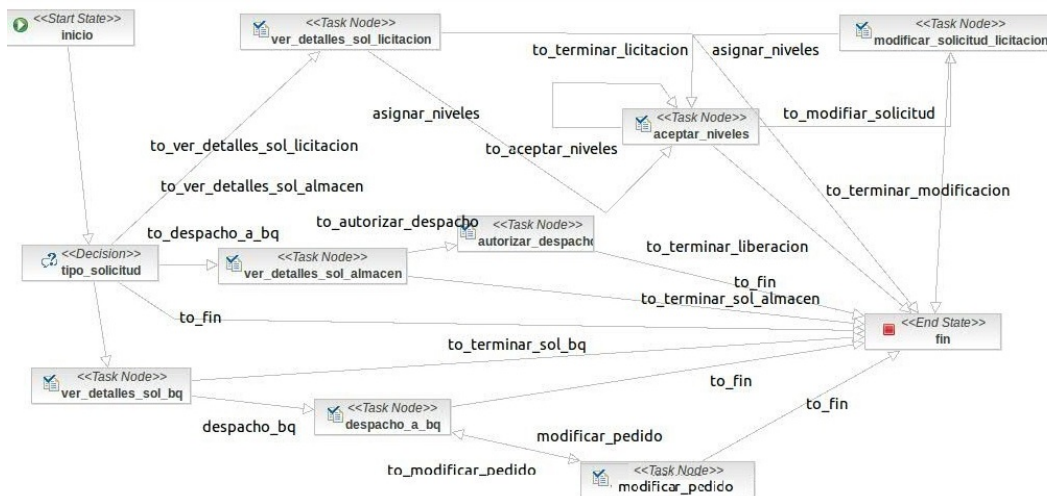


Figura 3.4: Modelo del proceso *Solicitar Producto*.

Antes de cualquier **Task Node** aparece un nodo de tipo **Decision**, denominado *tipo\_solicitud*. Esto plantea cuatro posibles alternativas. La tercera alternativa vincula el evento *tipo\_solicitud* directamente con el fin del proceso. La selección de dicha alternativa no produce evidencias en el registro de eventos. Las restantes alternativas involucradas corresponden a los subprocesos *ver\_detalles\_sol\_licitacion*, *ver\_detalles\_sol\_almacen* y *ver\_detalles\_sol\_bq*.

La aplicación de la *Minería de Variantes*, con los valores predeterminados en los parámetros de ruido y completitud, permite la obtención de dos variantes del proceso. Las variantes obtenidas corresponden a las descomposiciones mediante los patrones de control de flujo *Selección Exclusiva* y *Secuencia*, respectivamente.

### Análisis de la variante de Selección Exclusiva

En esta variante se identifica la existencia de una selección exclusiva entre tres subprocesos. La presencia de este patrón de control de flujo, en este nivel de abstracción, indica que al inicio del proceso existe un punto de decisión a partir del cual solo se ejecuta uno de los subprocesos involucrados. Esto es coherente con el modelo de referencia (Figura 3.4). En la Figura 3.5 se muestra el *Árbol de Variantes* identificado. A la derecha del modelo se puede observar la leyenda que contiene la codificación de las actividades.

En el *Perfil de Diagnóstico* asociado a esta variante se obtuvo una evaluación de 0.0 para el ruido y 1.0 para la completitud. La evaluación de 0.0 de ruido significa que para determinar esta variante no fue necesaria la consideración de comportamiento como ruido, mientras que

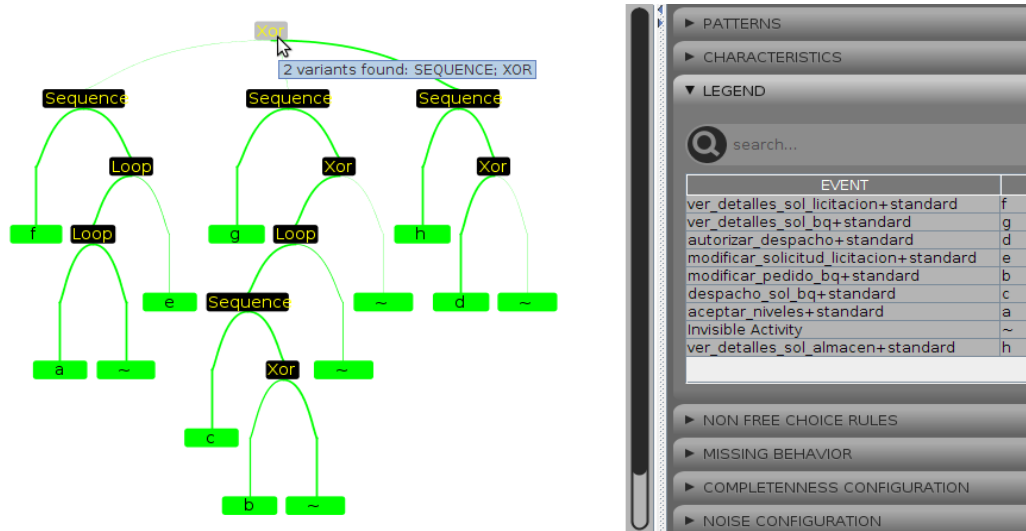


Figura 3.5: Variante del proceso mediante la descomposición por *Selección Exclusiva*.

el 1.0 de completitud indica que el comportamiento necesario para la identificación del patrón se encuentra presente en el registro de eventos. Las estimaciones de ruido y completitud obtenidas para esta descomposición corroboran que en el registro de eventos se expresa el comportamiento del patrón de control de flujo *Selección Exclusiva* en su totalidad.

El *Árbol de Variantes* proporciona una vista jerárquica del proceso, lo que posibilita el análisis de cada uno de los subprocesos que componen el proceso por separado y de acuerdo a diferentes niveles de abstracción. Se aprecia además, en el grosor de la última arista del primer nivel de abstracción, que la alternativa más ejecutada del proceso es la del subproceso `ver_detalle_sol_almacen`, conformado por las actividades `ver_detalle_sol_almacen` y `autorizar_despacho`, codificadas como `h` y `d` respectivamente. Seguidamente se realiza el análisis de los subprocesos identificados correspondientes al segundo nivel de abstracción.

### Subproceso `ver_detalle_sol_almacen`

En la Figura 3.6 se muestra la sección del modelo del proceso correspondiente al subproceso `ver_detalle_sol_almacen`. La relación de las actividades `ver_detalle_sol_almacen` y `autorizar_despacho` con el evento de fin del proceso indica que ambas actividades pueden aparecer como actividad de fin de traza. Este comportamiento se encuentra reflejado en el registro de eventos y en la descomposición obtenida para esta sección del registro de eventos. La descomposición obtenida para este subproceso, las trazas y la codificación de sus actividades se muestran en la Figura 3.7

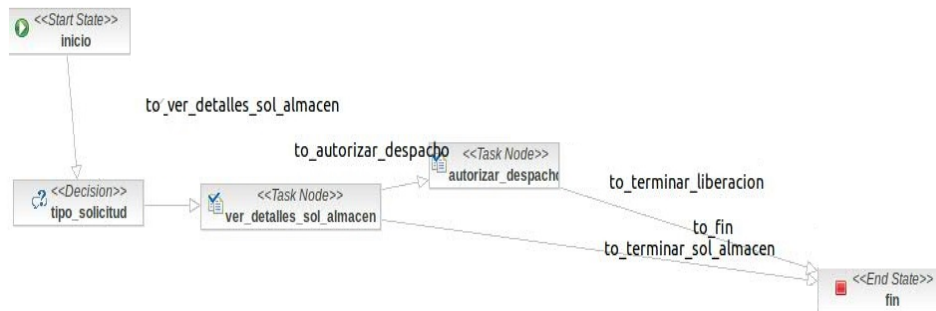


Figura 3.6: Modelo de proceso del subprocesso *ver\_detalles\_sol\_almacen*

En esta sección del registro de eventos se identifica la existencia de la secuencia entre la actividad *ver\_detalles\_sol\_almacen* y un subprocesso descompuesto mediante el patrón de control de flujo *Selección Exclusiva*. La traza que contiene únicamente a la actividad *autorizar\_despacho* representa una de las situaciones de ausencia de información reportadas en la literatura [Yzquierdo-Herrera, 2012].

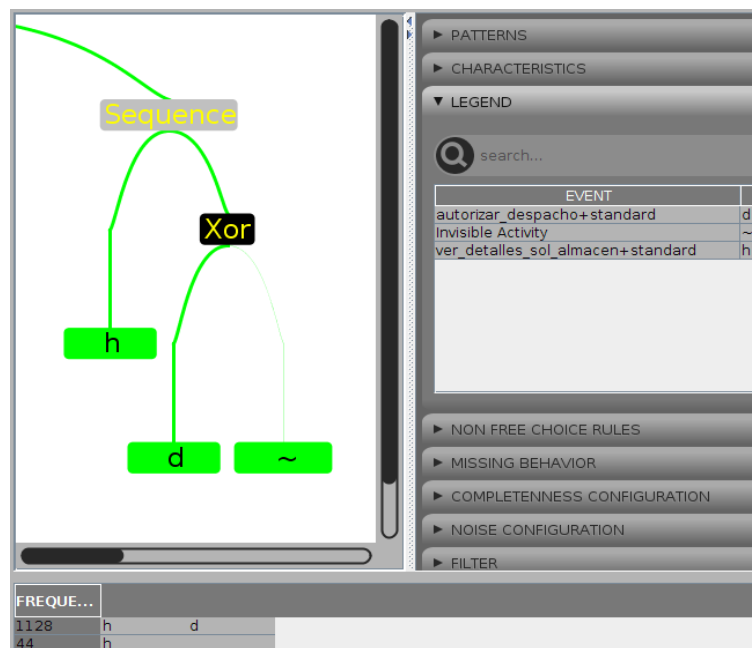


Figura 3.7: Subprocesso *ver\_detalles\_sol\_almacen*

La ausencia de representación de uno de los subprocessos que se encuentran en secuencia en alguna traza es considerada ausencia de información para el patrón de control de flujo *Secuencia*. El análisis de la sección del registro de eventos correspondiente evidencia la au-

sencia de información en la segunda traza. La identificación de la ausencia de información en esta traza provoca la inserción de una actividad invisible inmediatamente después de la actividad *ver\_detalle\_sol\_almacen*. Una tarea invisible se puede manifestar cuando se produce un salto de una o varias actividades en una situación de selección. La inserción de la actividad invisible en la traza posibilita la posterior identificación del patrón de *Selección Exclusiva*.

En el *Árbol de Variantes* se puede apreciar, en el grosor de las aristas que se refieren al subproceso cuya descomposición se realizó utilizando el patrón de *Selección Exclusiva*, que la alternativa que contiene la actividad invisible insertada es menos frecuente. Esto indica que la ejecución más común del subproceso *ver\_detalle\_sol\_almacen* es la secuencia conformada por *ver\_detalle\_sol\_almacen* y *autorizar\_despacho*. Esto es coherente con la frecuencia de las trazas únicas de la sección del registro de eventos referente a este subproceso (Figura 3.7).

### Subproceso *ver\_detalle\_sol\_licitacion*

Al analizar el modelo relativo a este subproceso se aprecia la existencia de varias actividades que finalizan el proceso y la presencia de múltiples lazos. La Figura 3.8 muestra la existencia de una relación entre las actividades *ver\_detalle\_sol\_licitacion*, *modificar\_solicitud\_licitacion* y *aceptar\_niveles* con el evento que finaliza el proceso, lo que indica que las tres actividades mencionadas pueden aparecer como actividad de fin de traza.

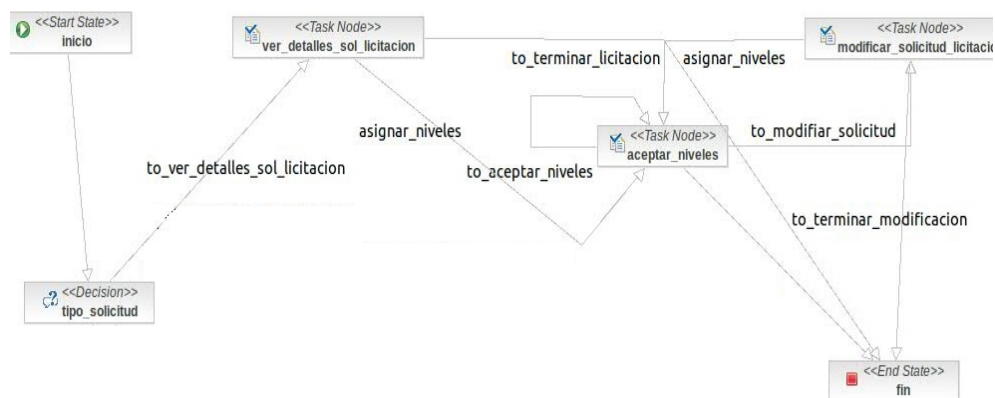


Figura 3.8: Modelo de referencia del subproceso *ver\_detalle\_sol\_licitacion*.

En la Figura 3.9 se muestra el resultado obtenido por la *Minería de Variantes* en la sección del registro de eventos asociada a este subproceso. En la sección del *Árbol de Variantes* correspondiente a este subproceso se identifica la secuencia entre la actividad *ver\_detalles\_sol\_licitacion* y un subproceso compuesto por las actividades *modificar\_solicitud\_licitacion* y *aceptar\_niveles*. La aparición de esta relación indica que después de la ocurrencia de *ver\_detalles\_sol\_licitacion* debe ejecutarse obligatoriamente alguna actividad, por lo que esta actividad no debe aparecer al final de una traza. Esto contradice lo expresado en el modelo de referencia (Figura 3.8), pero es coherente con el comportamiento reflejado en las trazas.

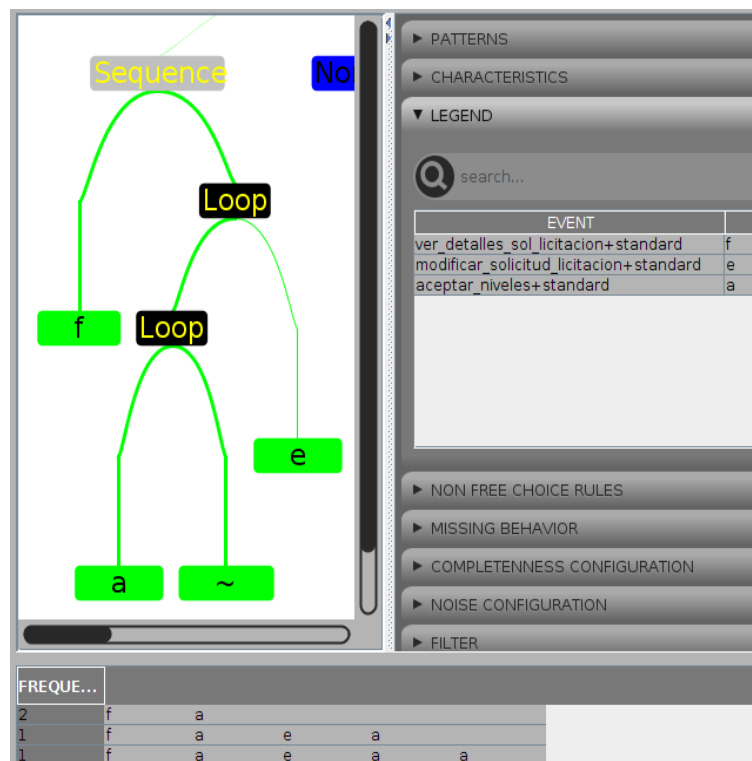


Figura 3.9: Subproceso *ver\_detalles\_sol\_licitacion*

En el subproceso restante se determinó la existencia de un lazo donde la actividad *ver\_detalles\_sol\_licitacion* constituye el subproceso **Redo**. El subproceso **Do** se descompone en un lazo entre la actividad *aceptar\_niveles* y una actividad invisible. La presencia de la actividad *modificar\_solicitud\_licitacion* en el **Redo** significa que nunca es la última actividad en ejecutarse en el *Lazo* al que pertenece y por tanto nunca debe aparecer como evento de fin de proceso.

El análisis de las trazas correspondientes a esta parte del registro de eventos muestra que solo la actividad *aceptar\_niveles* ha aparecido al final de las trazas. Por tanto, se concluye que las alternativas de terminar el proceso después de la actividad *ver\_detalles\_sol\_licitacion* o después de *modificar\_solicitud\_licitacion*, no han sido ejecutadas, al menos en el ámbito temporal que contiene el registro de eventos analizado. Además se evidencia que en dos ejecuciones del proceso no aparece la actividad *modificar\_solicitud\_licitacion*.

La descomposición obtenida para este subproceso no explica todo el comportamiento contenido en el modelo del proceso, sin embargo los patrones de control de flujo identificados son coherentes con la información existente en el registro de eventos. Se puede afirmar entonces, que en el registro de eventos no se encuentra recogido todo el comportamiento expresado en el modelo de este subproceso (Figura 3.8).

### Subproceso *ver\_detalles\_sol\_bq*

En la Figura 3.10 se muestra la sección del modelo del proceso correspondiente al subproceso *ver\_detalles\_sol\_bq*.

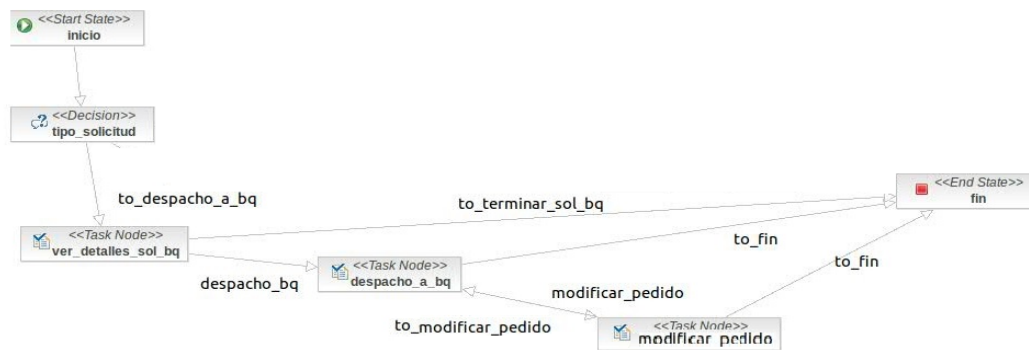


Figura 3.10: Modelo de proceso del subproceso *ver\_detalles\_sol\_bq*.

De modo similar a los subprocesos analizados anteriormente, de acuerdo al modelo de proceso, existen varias actividades que pueden aparecer como actividades de fin de proceso. En este caso son las actividades *ver\_detalles\_sol\_bq*, *despacho\_a\_bq* y *modificar\_pedido*. Por otra parte es posible identificar la existencia de un lazo que involucra las actividades *despacho\_a\_bq* y *modificar\_pedido*. Además se evidencia la existencia de una secuencia entre la actividad *ver\_detalles\_sol\_bq* y un subproceso compuesto por el resto de las actividades. El *Árbol de Variantes* obtenido se muestra en la Figura 3.11.

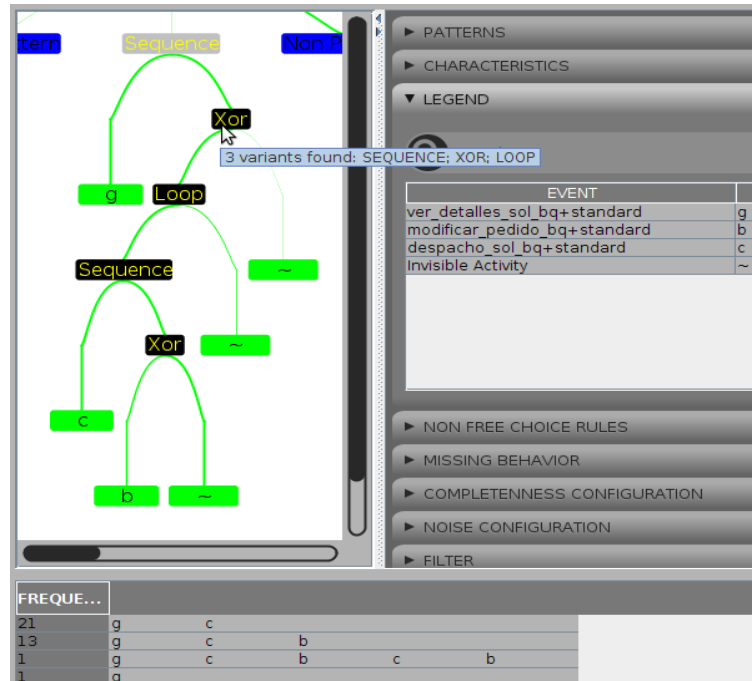


Figura 3.11: Subproceso *ver\_detalles\_sol\_bq*

El resultado de la aplicación de la técnica para este subproceso muestra que existe una traza incompleta de acuerdo a la descomposición mediante el patrón de control de flujo secuencia. Dicha traza se identifica como incompleta debido a que si dos subprocesos se encuentran en secuencia todas las trazas correspondientes deben contener una representación de ambos. En este caso en la traza que aparece solamente la actividad *ver\_detalles\_sol\_bq* no hay representación de las actividades que componen el segundo subproceso. La identificación de la situación anteriormente descrita produce la introducción de una actividad invisible en niveles inferiores de abstracción. La introducción de la misma expresa la relación directa entre la actividad *ver\_detalles\_sol\_bq* y el evento de fin de proceso. Se evidencia en el grosor de la arista que representa esta alternativa en el *Árbol de Variantes*, que esta ejecución es muy poco frecuente, lo cual coincide con las trazas. Esto sugiere una anomalía en el proceso.

En esta variante se identifica un lazo en el cual el subproceso **Redo** está compuesto por una actividad invisible o sea, un lazo del que solo existe evidencia de un subproceso, el **Do**. En el subproceso **Do** se identifica una secuencia entre la actividad *despacho\_sol\_bq* y un subproceso descompuesto por *Selección Exclusiva*, lo que indica que la actividad *despacho\_sol\_bq* siempre debe aparecer en la ejecución del subproceso. El patrón *Selección Exclusiva* identificado se evidencia entre la actividad *modificar\_pedido\_bq* y una actividad invisible, por tanto la



ejecución del lazo puede terminar tanto en la actividad *despacho\_sol\_bq* como en la actividad *modificar\_pedido\_bq*.

El análisis de la sección del registro de eventos correspondiente muestra que la descomposición obtenida representa el comportamiento presente en el registro de eventos. Además el *Árbol de Variantes* obtenido es coherente con el modelo de referencia (Figura 3.10).

### 3.4.2. Aplicación de técnicas existentes para el diagnóstico de proceso

Las características deseadas en las técnicas de diagnóstico, para la comprensión del proceso, son la identificación de patrones de control de flujo, manejo y visualización del ruido y la ausencia de información, así como la obtención de una vista global y jerárquica del proceso. A continuación, se analizan estas características en las técnicas de diagnóstico presentadas como parte del marco teórico de esta investigación. Para esto se presentan los resultados obtenidos al aplicar la *Alineación de Trazas* [Bose, R.P.J.C. and van der Aalst, W.M.P., 2012], la *Minería Difusa* [Günther and van der Aalst, W.M.P., 2007] y la *Descomposición en Bloques de Construcción* [Yzquierdo-Herrera et al., 2013a], a partir del registro de eventos del proceso *Solicitar Producto*.

#### Análisis de los resultados de la técnica *Alineación de Trazas*

En la Figura 3.12 se muestra el resultado de la aplicación de la técnica *Alineación de Trazas* al registro de eventos proveniente del sistema *XAVIA HIS*.

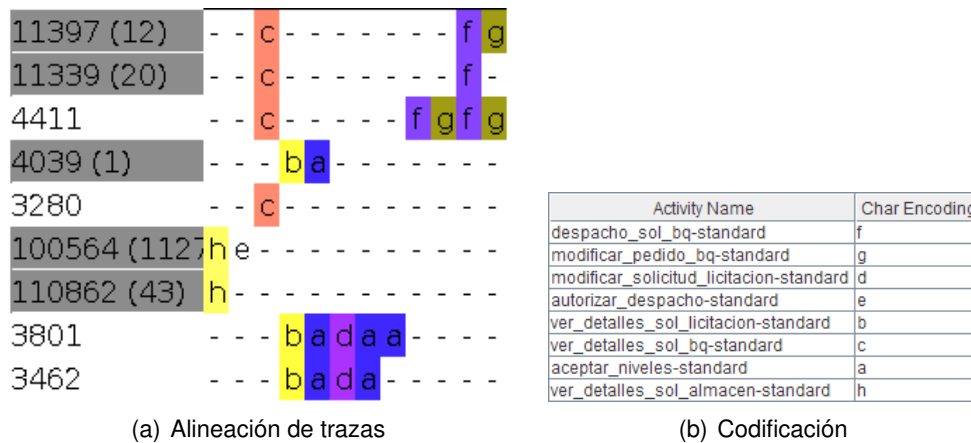


Figura 3.12: Alineación de trazas y su codificación, relativas al proceso *Solicitar Producto*.

El análisis de la alineación de trazas obtenida permite identificar que las actividades *h* y *e* nunca aparecen en las mismas trazas que las actividades *c*, *f* y *g*. Esto también sucede con el grupo conformado por las actividades *a*, *b* y *d* respecto a los grupos antes mencionados. Esta regularidad sugiere la existencia del patrón de control de flujo *Selección Exclusiva* entre tres subprocesos. Lo cual es coherente con el modelo de referencia del proceso (Figura 3.4). No obstante, identificar patrones de este tipo a partir de una alineación, puede resultar complejo, especialmente en registros de eventos que involucren más actividades. Esto es un ejemplo de que la ausencia de una vista jerárquica del proceso afecta su comprensión.

Al verificar la frecuencia de ejecución de las trazas se puede apreciar que aproximadamente el 96% de las trazas corresponden al subproceso donde se ejecutan las actividades *ver\_detalles\_sol\_almacen* y *autorizar\_despacho*. Además se puede observar que en el 3.75% de ejecuciones de este subproceso aparece únicamente la actividad *ver\_detalles\_sol\_almacen*.

Las trazas alineadas que involucran a las actividades *ver\_detalles\_sol\_bq*, *despacho\_sol\_bq* y *modificar\_pedido\_bq* indican una secuencia entre la actividad *ver\_detalles\_sol\_bq* y un subproceso conformado por las restantes. Se puede apreciar una traza donde únicamente aparece la actividad *ver\_detalles\_sol\_bq*, lo cual puede ser un comportamiento anómalo. Además, la repetición de las actividades *despacho\_sol\_bq* y *modificar\_pedido\_bq*, en una misma traza, sugiere la existencia de un lazo que involucra a ambas.

Al analizar la sección de la alineación que involucra a las actividades *ver\_detalles\_sol\_licitacion*, *aceptar\_niveles* y *modificar\_solicitud\_licitacion* se aprecia una secuencia entre *ver\_detalles\_sol\_licitacion* y las restantes. Además la aparición sucesiva de la actividad *aceptar\_niveles*, sugiere un lazo que involucra además a *modificar\_solicitud\_licitacion* y siempre culmina con *aceptar\_niveles*.

Se puede apreciar que la inspección visual de la alineación de trazas es insuficiente para extraer toda la información contenida en el registro de eventos. Esto se debe a que la identificación de comportamientos que sugieran la existencia de determinados patrones de control de flujo, se complica al no tener una vista jerárquica del proceso. Tampoco es sencillo distinguir el comportamiento que puede ser considerado como ruido, ya que aunque se ofrece la frecuencia de ejecución de las trazas, esta información no se jerarquiza. Además, cualquier inferencia se basa únicamente en lo contenido en la alineación, por lo cual no se maneja la ausencia de información. Por último, cuando el número de trazas y actividades aumenta, la

complejidad de la alineación obtenida implica una alta carga cognitiva. Todos estos elementos limitan la comprensión del proceso.

### Análisis de los resultados de la técnica *Minería Difusa*

La ejecución de la *Minería Difusa* sobre el registro de eventos en cuestión genera el mapa de proceso que se observa en la Figura 3.13

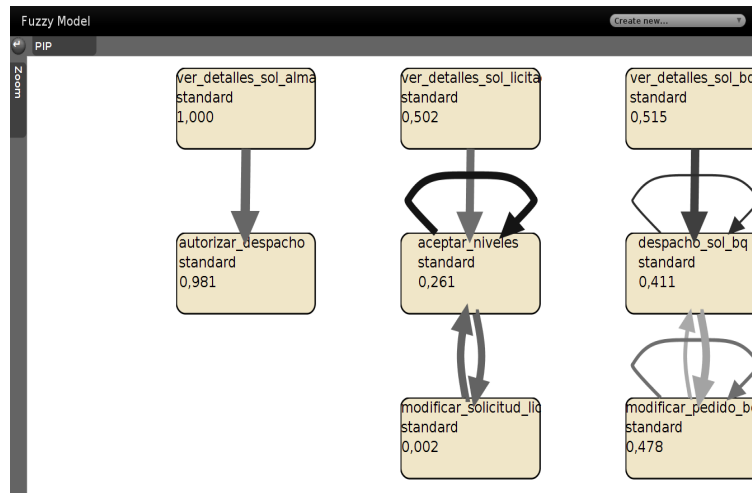


Figura 3.13: Mapa del proceso *Solicitar Producto*.

En el mapa obtenido, se aprecian tres procesos, lo cual está alineado con las tres alternativas de ejecución que existen en el proceso. Sin embargo, a diferencia del *Árbol de Variantes*, en este modelo no se muestra la relación existente entre estos procesos. Además, no se brinda información sobre la alta relevancia, desde el punto de vista de frecuencia de ejecución, que posee la alternativa que incluye las actividades *ver\_detalles\_sol\_almacen* y *autorizar\_despacho*. Esta alternativa está presente en 1172 trazas de las 1212 que contiene el registro de eventos.

La *Minería Difusa* no considera la información ausente en el registro de eventos, ni reconoce patrones de control de flujo. Esto provoca que en el subproceso *ver\_detalles\_sol\_almacen*, únicamente se refleje la secuencia entre las dos actividades que contiene y no se modele la alternativa de que únicamente se ejecute la primera actividad y el proceso termine sin ejecutar la actividad *autorizar\_despacho*. El registro de eventos contiene 44 ejecuciones del proceso que siguen la alternativa expuesta.

Respecto al subproceso *ver\_detalles\_sol\_licitacion*, las relaciones que se muestran en el mapa del proceso obtenido, sugieren correctamente la secuencia y los lazos que conforman dicho subproceso. Sin embargo, no se modela que todas las ejecuciones del proceso terminan en la actividad *aceptar\_niveles*, comportamiento presente en el registro de eventos.

La sección del mapa de proceso que se refiere al subproceso *ver\_detalles\_sol\_bq* también refleja relaciones entre las actividades que sugieren las secuencias y lazos existentes en el mismo. No obstante, al no identificar formalmente los patrones de control de flujo, no se puede distinguir entre la sección **Do** y **Redo** de los lazos. Producto de esto, no se puede inferir que el registro de eventos posee ejecuciones del proceso que terminan frecuentemente en las actividades *despacho\_sol\_bq* y *modificar\_pedido\_bq*. Tampoco se refleja la posible anomalía que representa la ejecución del proceso que contiene únicamente a la actividad *ver\_detalles\_sol\_bq*. Además se muestra una relación de repetición sobre la actividad *despacho\_sol\_bq*, que puede interpretarse erróneamente como un lazo que contenga a esta actividad en el **Do** y una actividad invisible en el **Redo**. Sin embargo, ninguna traza contiene un comportamiento de este tipo, ni el modelo de referencia contempla esta alternativa.

Los resultados obtenidos evidencian que la *Minería Difusa*, al manejar y visualizar el ruido, incluye en el modelo resultante comportamientos infrecuentes, lo cual es positivo para la comprensión del proceso. No obstante, al no identificar formalmente los patrones de control de flujo y no considerar la ausencia de información, el modelo obtenido puede generar ciertas interpretaciones incorrectas del proceso, afectando su comprensión.

### **Análisis de los resultados de la técnica *Descomposición en Bloques de Construcción***

En la Figura 3.14 se muestra el resultado de la aplicación de la técnica Descomposición en Subprocesos al registro de eventos proveniente del sistema *XAVIA HIS*.

El patrón de control de flujo identificado en el primer nivel de descomposición es el de *Selección Exclusiva*. Las alternativas involucradas corresponden a los subprocesos *ver\_detalles\_sol\_bq* (BB\_2\_2), *ver\_detalles\_sol\_licitacion* (BB\_2\_3) y *ver\_detalles\_sol\_almacen* (BB\_2\_4). Este resultado es coherente con el modelo de referencia del proceso (Figura 3.4).

El primer subproceso presenta diferencias visuales respecto a los resultados obtenidos mediante la aplicación de la técnica propuesta. Sin embargo, esto se debe fundamentalmente a que para algunos patrones de control de flujo no se manejan los mismos conceptos. En

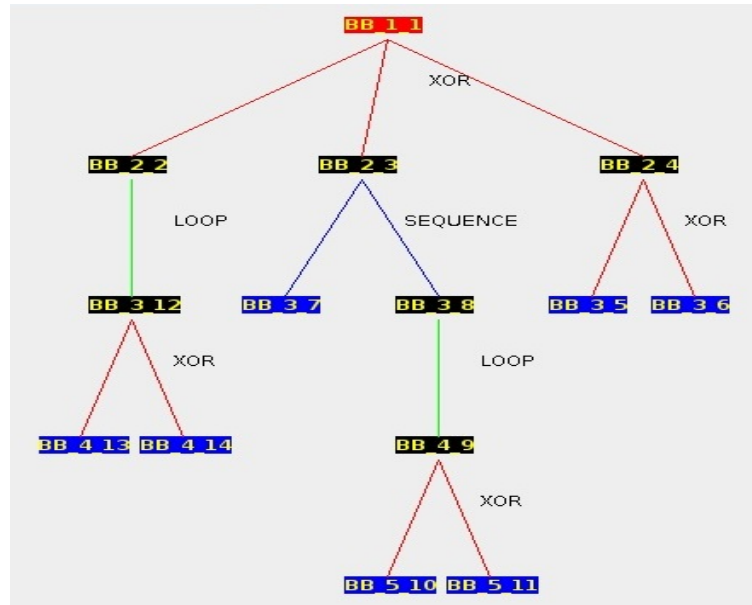


Figura 3.14: Descomposición en Bloques de Construcción.

el caso del *Lazo* la *Descomposición en Bloques de Construcción* no identifica qué parte del subproceso corresponde al **Do** y cual al **Redo**, sino que solamente identifica el subproceso que se repite. Por otra parte la *Descomposición en Bloques de Construcción* no identifica la secuencia entre la actividad *despacho\_sol\_bq* y el subproceso constituido por las actividades *modificar\_pedido\_bq* y una actividad invisible (Figura 3.10).

En el subproceso *ver\_detalle\_sol\_licitacion* existe una incongruencia en la identificación del patrón de control correspondiente al último nivel. En este caso mediante la *Minería de Variantes* se identifica un lazo y la *Descomposición en Bloques de Construcción* permite determinar la existencia del patrón de control de flujo *Selección Exclusiva*. Debido a que en ambos casos el subproceso identificado en el nivel anterior se descompuso mediante el patrón de control de flujo *Lazo* ambas variantes explican el comportamiento presente en el registro de eventos. No obstante, la descomposición propuesta por la *Minería de Variantes* presenta una mayor precisión, ya que explica el comportamiento presente en el modelo de referencia del proceso, sin admitir comportamiento adicional.

La aplicación de la técnica *Descomposición en Bloques de Construcción* evidencia el impacto positivo que generan en la comprensión del proceso la identificación de los patrones de control de flujo y una vista jerárquica del proceso. No obstante, la existencia de comportamientos infrecuentes y la representación del *Lazo* conducen, en ciertos contextos, a la identificación

de patrones erróneos. Esto afecta la comprensión en su dimensión empírica, al crear un modelo que, en ciertos elementos, se distancia de la ejecución real del proceso.

### **3.5. Conclusiones parciales**

La herramienta informática desarrollada para implementar la técnica propuesta, permitió su aplicación en registros de eventos que combinan aleatoriamente situaciones de ruido y ausencia de información. Las pruebas estadísticas realizadas mostraron que no existen diferencias significativas entre los modelos originales y los obtenidos con la técnica propuesta, lo cual resalta la efectividad de la técnica en la detección de los patrones de control de flujo.

Adicionalmente, se analizaron los resultados de la aplicación de la *Minería de Variantes* a un registro de eventos proveniente de un entorno real. En una de las variantes de descomposición se identificaron patrones de control de flujo que son coherentes con el modelo de referencia del proceso. Estos resultados fueron comparados con los obtenidos por otras técnicas para el diagnóstico de procesos. La asunción de comportamiento en el registro de eventos permitió identificar patrones de control de flujo y variantes del proceso que no fueron identificados por las demás técnicas. Tanto la evaluación en un entorno real, como el experimento realizado, permiten concluir que la técnica propuesta mejora la comprensión del proceso.

# Conclusiones

Se analizaron las técnicas de minería de proceso que han sido empleadas en el diagnóstico, identificándose los elementos que impactan en la comprensión del proceso. Se construyó una nueva técnica para el diagnóstico, la cual fue validada mediante pruebas estadísticas y la aplicación en un entorno real. De esta manera se arriba a las siguientes conclusiones:

- La identificación de patrones de control de flujo, manejo y visualización del ruido y la ausencia de información, así como la obtención de una vista global y jerárquica del proceso son características que impactan en la comprensión del proceso durante su diagnóstico.
- El algoritmo propuesto para la identificación de los patrones de control de flujo permite la descomposición en subprocesos, a partir de la combinación de los comportamientos presentes y ausentes en el registro de eventos. Esta descomposición permite obtener diferentes variantes de descomposición jerárquica del proceso.
- El perfil de diagnóstico obtenido mejora la comprensión del proceso al ubicar en contexto el comportamiento considerado como ruido y el asumido como ausente, durante la identificación de los patrones de control de flujo.
- Las pruebas estadísticas realizadas evidencian que la técnica propuesta identifica correctamente los patrones de control de flujo, en registros de eventos con ruido y/o ausencia de información.
- El complemento desarrollado para el marco de minería de proceso ProM, permite utilizar la técnica propuesta en escenarios reales, durante la etapa de diagnóstico de procesos.
- Los resultados obtenidos al aplicar la *Minería de Variantes* y otras técnicas de diagnóstico, sobre un registro de eventos proveniente de un entorno real, permiten afirmar que la solución propuesta mejora la comprensión en el diagnóstico de procesos.

# Recomendaciones

- Extender la aplicación de la técnica propuesta a otros entornos reales, donde existan registros de eventos con ruido y/o ausencia de información.
- Utilizar la información contenida en el *Perfil de Diagnóstico* obtenido para la recomendación de algoritmos de descubrimiento de procesos.
- Incorporar, al *Perfil de Diagnóstico* información de la perspectiva de tiempo, que facilite el análisis del rendimiento del proceso.
- Desarrollar un algoritmo para identificar, de las variantes propuestas por la técnica, aquellas que optimicen aspectos de interés para el análisis como aptitud, tiempo de repuesta, precisión consumo de recursos.
- Extender la técnica desarrollada para la identificación de variaciones temporales en el proceso y análisis transversal de procesos.



# Referencias bibliográficas

- [Agarwal et al., 1999] Agarwal, R., De, P., and Sinha, A. P. (1999). Comprehending object and process models: An empirical study. *Software Engineering, IEEE Transactions on*, 25(4):541–556. 00100.
- [Becker et al., 2000] Becker, J., Rosemann, M., and von Uthmann, C. (2000). Guidelines of business process modeling. In *Business Process Management*, pages 30–49. Springer.
- [Betsy Burton, 2012] Betsy Burton (2012). BPM best practices of highly adaptive organizations.
- [Bezerra and Wainer, 2012] Bezerra, F. and Wainer, J. (2012). A dynamic threshold algorithm for anomaly detection in logs of process aware systems. *Journal of Information and Data Management*, 3(3):316.
- [Bezerra and Wainer, 2013] Bezerra, F. and Wainer, J. (2013). Algorithms for anomaly detection of traces in logs of process aware information systems. *Information Systems*, 38(1):33–44.
- [Bose, 2012] Bose, J. C. (2012). *Process Mining in the Large: Preprocessing, Discovery, and Diagnostics*. PhD thesis, Eindhoven University of Technology.
- [Bose, R.P.J.C. and van der Aalst, W.M.P., 2012] Bose, R.P.J.C. and van der Aalst, W.M.P. (2012). Process diagnostics using trace alignment: Opportunities, issues, and challenges. *Information Systems*, 37(2):117–141.
- [Buijs et al., 2012] Buijs, J. C., van Dongen, B. F., and van der Aalst, W. M. (2012). On the role of fitness, precision, generalization and simplicity in process discovery. In *On the Move to Meaningful Internet Systems: OTM 2012*, pages 305–322. Springer. 00008.

- [Burattin and Sperduti, 2011] Burattin, A. and Sperduti, A. (2011). PLG: A framework for the generation of business process models and their execution logs. In *Business Process Management Workshops*, pages 214–219. Springer. 00028.
- [Canfora et al., 2005] Canfora, G., García, F., Piattini, M., Ruiz, F., and Visaggio, C. A. (2005). A family of experiments to validate metrics for software process models. *Journal of Systems and Software*, 77(2):113–129. 00070.
- [Davenport, 2013] Davenport, T. H. (2013). *Process Innovation: Reengineering Work Through Information Technology*. Harvard Business Press. 06396.
- [De Medeiros et al., 2007] De Medeiros, A., Weijters, A., and van der Aalst, W.M.P. (2007). Genetic process mining: An experimental evaluation. *Data Mining and Knowledge Discovery*, 14(2):245–304.
- [De Weerd et al., 2013] De Weerd, J., Baesens, B., and Vanthienen, J. (2013). A comprehensive benchmarking framework (CoBeFra) for conformance analysis between procedural process models and event logs in ProM. In *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2013), part of the IEEE Symposium Series in Computational Intelligence 2013*. 00000.
- [De Weerd et al., 2012] De Weerd, J., De Backer, M., Vanthienen, J., and Baesens, B. (2012). A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Information Systems*, 37:654–676.
- [Depaire et al., 2013] Depaire, B., Swinnen, J., Jans, M., and Vanhoof, K. (2013). A process deviation analysis framework. In Rosa, M. L. and Soffer, P., editors, *Business Process Management Workshops*, number 132 in Lecture Notes in Business Information Processing, pages 701–706. Springer Berlin Heidelberg.
- [Figl et al., 2013] Figl, K., Recker, J., and Mendling, J. (2013). A study on the effects of routing symbol design on process model comprehension. *Decision Support Systems*, 54(2):1104–1118. 00005.
- [Günther and van der Aalst, W.M.P., 2007] Günther, C. and van der Aalst, W.M.P. (2007). Fuzzy mining - adaptive process simplification based on multi-perspective metrics. In *Business Process Management*, volume 4714 LNCS of *5th International Conference on Business Process Management, BPM 2007*, pages 328–343, Brisbane.

## Referencias bibliográficas

- [Hahn and Kim, 1999] Hahn, J. and Kim, J. (1999). Why are some diagrams easier to work with? effects of diagrammatic representation on the cognitive intergration process of systems analysis and design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 6(3):181–213. 00068.
- [Hammer and Champy, 1993] Hammer, M. and Champy, J. (1993). Reengineering the corporation: A manifesto for business revolution. *Business horizons*, 36(5):90–91.
- [Hendricks et al., 2007] Hendricks, K. B., Singhal, V. R., and Stratman, J. K. (2007). The impact of enterprise systems on corporate performance: A study of ERP, SCM, and CRM system implementations. *Journal of Operations Management*, 25(1):65–82.
- [Jagadeesh Chandra Bose and van der Aalst, W.M.P., 2009] Jagadeesh Chandra Bose, R. and van der Aalst, W.M.P. (2009). Abstractions in process mining: A taxonomy of patterns. *Business Process Management*, pages 159–175.
- [Jagadeesh Chandra Bose and van der Aalst, W.M.P., 2010] Jagadeesh Chandra Bose, R. and van der Aalst, W.M.P. (2010). Trace alignment in process mining: opportunities for process diagnostics. *Business Process Management*, pages 227–242.
- [Jan Claes and Geert Poels, 2012] Jan Claes and Geert Poels (2012). Process mining and the ProM framework: An exploratory survey. In *LNBIP*. Springer.
- [Jans et al., 2011] Jans, M., Van Der Werf, J., Lybaert, N., and Vanhoof, K. (2011). A business process mining application for internal transaction fraud mitigation. *Expert Systems with Applications*, 38(10):13351–13359.
- [Johansson et al., 1993] Johansson, H. J., McHugh, P., Pendlebury, A. J., and Wheeler, W. A. (1993). *Business process reengineering: Breakpoint strategies for market dominance*. Wiley Chichester. 00665.
- [Kirschner, 2002] Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and instruction*, 12(1):1–10. 00408.
- [Kotovsky et al., 1985] Kotovsky, K., Hayes, J. R., and Simon, H. A. (1985). Why are some problems hard? evidence from tower of hanoi. *Cognitive psychology*, 17(2):248–294. 00601.

- [Krogstie et al., 2006] Krogstie, J., Sindre, G., and Jorgensen, H. a. (2006). Process models representing knowledge for action: a revised quality framework. *European Journal of Information Systems*, 15(1):91–102. 00199.
- [Leemans et al., 2013] Leemans, S. J., Fahland, D., and van der Aalst, W. M. (2013). Discovering block-structured process models from event logs containing infrequent behaviour. In *BPM Conference*.
- [Li et al., 2011a] Li, C., Reichert, M., and Wombacher, A. (2011a). Mining business process variants: Challenges, scenarios, algorithms. *Data and Knowledge Engineering*, 70(5):409–434.
- [Li et al., 2011b] Li, J., Bose, R. P. J. C., and Van Der Aalst, W. M. P. (2011b). Mining context-dependent and interactive business process maps using execution patterns. *BPM Worksh*, pages 109–121.
- [Lindland et al., 1994] Lindland, O. I., Sindre, G., and Solvberg, A. (1994). Understanding quality in conceptual modeling. *Software, IEEE*, 11(2):42–49. 00725.
- [López-Jiménez et al., 2014] López-Jiménez, R., Pupo-Hernández, E., Yzquierdo-Herrera, R., and Pérez-Alfonso, D. (2014). Técnicas de minería de proceso para el diagnóstico de procesos de negocio. process mining techniques for business processes diagnosis. In *I Conferencia Científica Internacional UCIENCIA*, La Habana, Cuba. Universidad de las Ciencias Informáticas.
- [Ma, 2012] Ma, L. (2012). *How to Evaluate the Performance of Process Discovery Algorithms*. Master thesis, Eindhoven University of Technology, Netherlands.
- [Mendling, 2008] Mendling, J. (2008). *Metrics for process models: empirical foundations of verification, error prediction, and guidelines for correctness*, volume 6. Springer. 00194.
- [Mendling et al., 2007] Mendling, J., Reijers, H. A., and Cardoso, J. (2007). What makes process models understandable? In *Business Process Management*, pages 48–63. Springer. 00183.
- [Mendling et al., 2010a] Mendling, J., Reijers, H. A., and Recker, J. (2010a). Activity labeling in process modeling: Empirical insights and recommendations. *Information Systems*, 35(4):467–482.

- [Mendling et al., 2010b] Mendling, J., Reijers, H. A., and van der Aalst, W. M. (2010b). Seven process modeling guidelines (7PMG). *Information and Software Technology*, 52(2):127–136.
- [Mendling et al., 2012] Mendling, J., Strembeck, M., and Recker, J. (2012). Factors of process model comprehension. findings from a series of experiments. *Decision Support Systems*, 53(1):195–206.
- [Miller, 1956] Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2):81. 19097.
- [Moher et al., 1993] Moher, T. G., Mak, D. C., Blumenthal, B., and Levanthal, L. M. (1993). Comparing the comprehensibility of textual and graphical programs. In *Empirical Studies of Programmers: Fifth Workshop*, pages 137–161. Ablex, Norwood, NJ. 00086.
- [Molka et al., 2013] Molka, T., Gilani, W., and Zeng, X.-J. (2013). Dotted chart and control-flow analysis for a loan application process. In Rosa, M. L. and Soffer, P., editors, *Business Process Management Workshops*, number 132 in Lecture Notes in Business Information Processing, pages 223–224. Springer Berlin Heidelberg.
- [Pérez-Alfonso et al., 2013] Pérez-Alfonso, D., Yzquierdo-Herrera, R., and Lazo-Cortés, M. (2013). Recommendation of process discovery algorithms: a classification problem. *RESEARCH IN COMPUTING SCIENCE*, 61:33–42.
- [Poels et al., 2011] Poels, G., Maes, A., Gailly, F., and Paemeleire, R. (2011). The pragmatic quality of resources-events-agents diagrams: an experimental evaluation. *Information Systems Journal*, 21(1):63–89. 00018.
- [Purchase, 1997] Purchase, H. (1997). Which aesthetic has the greatest effect on human understanding? In *Graph Drawing*, pages 248–261. Springer. 00416.
- [Recker and Dreiling, 2011] Recker, J. and Dreiling, A. (2011). The effects of content presentation format and user characteristics on novice developers understanding of process models. *Communications of the Association for Information Systems*, 28. 00027.
- [Recker et al., 2013] Recker, J., Mendling, J., and Hahn, C. (2013). How collaborative technology supports cognitive processes in collaborative process modeling: A capabilities-gains-outcome model. *Information Systems*, 38(8):1031–1045. 00004.

- [Recker et al., 2014] Recker, J., Reijers, H., and Wouw, S. v. d. (2014). Process model comprehension: The effects of cognitive abilities, learning style, and strategy. *Communications of the Association for Information Systems*, 34(1). 00000.
- [Reijers et al., 2011a] Reijers, H. A., Freytag, T., Mendling, J., and Eckleder, A. (2011a). Syntax highlighting in business process models. *Decision Support Systems*, 51(3):339–349. 00020.
- [Reijers and Mendling, 2011] Reijers, H. A. and Mendling, J. (2011). A study into the factors that influence the understandability of business process models. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(3):449–462. 00079.
- [Reijers et al., 2011b] Reijers, H. A., Mendling, J., and Dijkman, R. M. (2011b). Human and automatic modularizations of process models to enhance their comprehension. *Information Systems*, 36(5):881–897. 00056.
- [Rummler and Brache, 2012] Rummler, G. A. and Brache, A. P. (2012). *Improving Performance: How to Manage the White Space on the Organization Chart*. John Wiley & Sons. 01676.
- [Russell et al., 1995] Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., and Edwards, D. D. (1995). *Artificial intelligence: a modern approach*, volume 2 of *Prentice Hall Series in Artificial Intelligence*. Prentice hall Englewood Cliffs, 2 edition.
- [Sarshar and Loos, 2005] Sarshar, K. and Loos, P. (2005). Comparing the control-flow of epc and petri net from the end-user perspective. In *Business Process Management*, pages 434–439. Springer. 00061.
- [Song and Aalst, 2007] Song, M. and Aalst, W. M. P. V. d. (2007). Supporting process mining by showing events at a glance. In *Proceedings of the 17th Annual Workshop on Information Technologies and Systems (WITS)*, pages 139–145.
- [Stamper, 1996] Stamper, R. K. (1996). *Signs, Information, Norms and Systems*. De Gruyter, Berlin.
- [van der Aalst, 2012] van der Aalst, W. (2012). A decade of business process management conferences: Personal reflections on a developing discipline. In Barros, A., Gal, A., and Kindler, E., editors, *Business Process Management*, volume 7481 of *Lecture Notes in Computer Science*, pages 1–16. Springer Berlin / Heidelberg.

- [van der Aalst et al., 2012] van der Aalst, W., Adriansyah, A., De Medeiros, A., Arcieri, F., Baier, T., Blickle, T., Bose, J., Van Den Brand, P., Brandtjen, R., Buijs, J., Burattin, A., Carmona, J., Castellanos, M., Claes, J., Cook, J., Costantini, N., Curbera, F., Damiani, E., De Leoni, M., Delias, P., Van Dongen, B., Dumas, M., Dustdar, S., Fahland, D., Ferreira, D., Gaaloul, W., Van Geffen, F., Goel, S., Günther, C., Guzzo, A., Harmon, P., Ter Hofstede, A., Hoogland, J., Ingvaldsen, J., Kato, K., Kuhn, R., Kumar, A., La Rosa, M., Maggi, F., Malerba, D., Mans, R., Manuel, A., McCreesh, M., Mello, P., Mendling, J., Montali, M., Motahari-Nezhad, H., Zur Muehlen, M., Munoz-Gama, J., Pontieri, L., Ribeiro, J., Rozinat, A., Seguel Pérez, H., Seguel Pérez, R., Sepúlveda, M., Sinur, J., Soffer, P., Song, M., Sperduti, A., Stilo, G., Stoel, C., Swenson, K., Talamo, M., Tan, W., Turner, C., Vanthienen, J., Varvaressos, G., Verbeek, E., Verdonk, M., Vigo, R., Wang, J., Weber, B., Weidlich, M., Weijters, T., Wen, L., Westergaard, M., and Wynn, M. (2012). *Process mining manifesto*, volume 99 LNBIP of *9th International Conference on Business Process Management, BPM 2011P*. Clermont-Ferrand.
- [van der Aalst, W.M.P., 2011] van der Aalst, W.M.P. (2011). *Process Mining. Discovery, Conformance and Enhancement of Business Processes*. Springer, Heidelberg, Dordrecht, London et. al.
- [van der Aalst, W.M.P. et al., 2004] van der Aalst, W.M.P., Weijters, A. J. M. M., and Maruster, L. (2004). Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142.
- [vanden Broucke et al., 2013] vanden Broucke, S., Delvaux, C., Freitas, J., Rogova, T., Vanthienen, J., and Baesens, B. (2013). Uncovering the relationship between event log characteristics and process discovery techniques. *Business Process Management Workshops*, Beijing (China). 00000.
- [Verbeek et al., 2011] Verbeek, H., Buijs, J., Van Dongen, B., and Van Der Aalst, W. (2011). *XES, XESame, and ProM 6*, volume 72 LNBIP of *CAiSE Forum 2010 on Information Systems Evolution*. Hammamet.
- [Wand and Weber, 2002] Wand, Y. and Weber, R. (2002). Research commentary: information systems and conceptual modeling - a research agenda. *Information Systems Research*, 13(4):363–376. 00598.

- [Weijters et al., 2006] Weijters, A., van der Aalst, W.M.P., and de Medeiros, A. K. A. (2006). Process mining with the heuristics miner-algorithm. *Technische Universiteit Eindhoven, Tech. Rep. WP*, 166.
- [Weske and Heidelberg, 2007] Weske, M. and Heidelberg, S.-V. B. (2007). *Business Process Management. Concepts, Languages, Architectures*.
- [Yzquierdo-Herrera, 2012] Yzquierdo-Herrera, R. (2012). *Modelo para la estimación de información ausente en las trazas usadas en la minería de proceso*. PhD thesis, Universidad de las Ciencias Informáticas.
- [Yzquierdo-Herrera, 2013] Yzquierdo-Herrera, R. (2013). Minería de proceso como herramienta para la auditoría. *Ciencias de la Información*, 44(2):25–32.
- [Yzquierdo-Herrera et al., 2013a] Yzquierdo-Herrera, R., Silverio-Castro, R., and Lazo-Cortés, M. (2013a). Sub-process discovery: Opportunities for process diagnostics. In Poels, G., editor, *Enterprise Information Systems of the Future*, number 139 in Lecture Notes in Business Information Processing, pages 48–57. Springer Berlin Heidelberg.
- [Yzquierdo-Herrera et al., 2013b] Yzquierdo-Herrera, R., Silverio-Castro, R., and Lazo-Cortés, M. (2013b). Tratamiento de la ausencia de información en la minería de procesos. *Revista Facultad de Ingeniería Universidad de Antioquia*, 0(69):67–78. 00000.