

Universidad de las Ciencias Informáticas

Facultad 3



Título: Algoritmo para la detección de patrones de control de flujo en los registros de eventos reales.

Autores: Reynaldo López Jiménez
Eudel Pupo Hernández

Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas.

Tutores: Ing. Damián Pérez Alfonso

Dr. Raykenler Yzquierdo Herrera

Co-Tutor: Lic. Raynel Batista Tellez

La Habana, Cuba 2014

MINISTERIO DE EDUCACIÓN SUPERIOR
UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

Declaramos ser autores del presente trabajo de diploma y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales del mismo, con carácter exclusivo. Autorizamos a dicho centro para que haga el uso que estime pertinente con este trabajo.

Para que así conste firmamos la presente a los ____ días del mes de _____ del año ____.

Reynaldo López Jiménez

Eudel Pupo Hernández

Firma de los Autores

Damián Pérez Alfonso

Raykenler Yzquierdo Herrera

Firma de los Tutores

A Zoraida, Mario, Noelia y Lumey.

A Eubel y Carlitos.

A los que tiene un sueño y luchan cada día por hacerlo realidad.

Agradecimientos

A mis abuelos Mario y Zoraida por ser mi soporte, apoyarme y brindarme todo el cariño del mundo a lo largo de toda mi vida.

A mi mamá Noelia y a mi hermana Lumey por darme la fuerza para seguir adelante y hacer que trate cada día que se sientan orgullosas.

A mi papá Reynaldo por todo el apoyo y preocupación.

A mi tía Cora por su confianza en mí y todo el soporte y ayuda que me ha dado.

A mi vecina Yara, mi otra abuela, por tantos años llamándome ingeniero.

A toda mi familia por ser imprescindible en toda mi formación y en mi vida.

A mis hermanos de la Lenin: Mario, Adrián, Carrillo, Beny, Alexis y Race por ser la mejor segunda familia que se puede pedir.

A Laurita, mi amiga de siempre, y su familia, por todo el cariño y apoyo incondicional que me han brindado.

A Alfredo, Katerín, Yorgüü, Osiel, Rancel, Talía, Susana, Miosotis, Yaimara y Frank por tantas enseñanzas, consejos y apoyo.

A Pupo, el mejor compañero de tesis que podría tener, por enseñarme a ver la vida de otro modo y ayudarme a ser una mejor persona.

A Vilma por ser un gran empuje en mi carrera y brindarme siempre su sabiduría y confianza.

A mis tutores Damián, Rainer y Raykenler por confiar en nosotros, por ofrecernos sus consejos y amistad.

A todas las personas que, de una forma u otra, me han ayudado a que cada día trate de ser un mejor profesional y una mejor persona.

A todos, muchas gracias,

Reynaldo.

A mis padres Aracelis y Anselmo y a mis hermanos Erianni y Eubel por su apoyo incondicional y sus incontables sacrificios. A mi tía Bárbara por hacer siempre suyos mis problemas y tratarme como a su hijo. A mi familia que siempre ha estado a mi lado y con los que siempre he podido contar.

A mis amigos del grupo, a los que se fueron y a los que quedamos, a los del viejo y a los del nuevo, a todos por igual. A todos por su parte en mi formación como persona y como profesional. A todos por su apoyo en los momentos difíciles y por convertir el resto en mejores momentos. A todos por la parte que han aportado a mi personalidad tal vez sin siquiera imaginárselo.

A todos los que sería injusto no dedicarles una línea cuando me han dedicado tanto tiempo: Yorgüü, Katy, Alieski, Alfredo, Susy, Nicole, Velozo, Nay, Pastor, Leidis, Carlos, Yoendri, Osiel, Onelio y a todos los demás que no puedo incluir porque no me alcanzarían las líneas.

A mis colegas de la Plataforma Informativa SOE por guiarme en mis primeros pasos y acompañarme en el resto del viaje. A mis compañeros de Capital Humano por aceptarme como uno de ellos y dejarme pertenecer su familia.

A Damián y Raykenler que se han convertido en algo más que tutores. Por siempre tener un momento, siempre hacer el espacio y siempre estar preocupados, claro cada uno a su modo.

A Rainer por no escatimar horas y sacrificios, por siempre creer, por servir de guía y soporte en cualquier circunstancia, por ser un amigo.

A Reynaldo por entenderme y soportarme o al menos por intentarlo.

A todos los que sus ideas, recomendaciones, comentarios, críticas molestas e incluso bromas pesadas han aportado a la realización de este trabajo.

A todos. . . muchas gracias.

Eudel.

Resumen

Los Sistemas de Gestión de Procesos de Negocios permiten almacenar registros de los eventos de los procesos que gestionan. La minería de proceso desarrolla técnicas y herramientas que permiten extraer y analizar la información almacenada en los registros de eventos. 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. Existen diferentes técnicas de diagnóstico de proceso las cuales suelen verse afectadas por características presentes en los registros de eventos como el ruido y la ausencia de información. Adicionalmente brindan poca información sobre los comportamientos infrecuentes del proceso y/o la completitud del registro de eventos.

Se propone una técnica para el diagnóstico de proceso denominada Minería de Variantes que permite determinar variantes de descomposición de un proceso. Dicha técnica posibilita la identificación de los patrones de control de flujo y contextualizar su aparición en los subprocesos correspondientes. Asimismo la utilización de esta técnica brinda la oportunidad de identificar los comportamientos infrecuentes en el registro de eventos y los que han sido asumidos para la identificación de los patrones de control de flujo.

Se muestra un análisis de los resultados de la aplicación de la Minería de Variantes para el diagnóstico de proceso en registros de eventos provenientes de entornos reales. Estos resultados son comparados con los obtenidos a partir de la utilización de otras técnicas para el diagnóstico de proceso.

Palabras claves: diagnóstico de proceso, minería de proceso, minería de variantes, variantes de modelos de proceso.

Índice

Introducción.....	1
Capítulo 1: Minería de proceso y técnicas para el diagnóstico de proceso.....	7
1.1 Minería de proceso	7
1.2 Diagnóstico de proceso.....	9
1.3 Técnicas de diagnóstico de proceso	10
1.3.1 Análisis de Diagramas de Puntos	10
1.3.2 Visualización de Flujo y Alcance.....	12
1.3.3 Arreglos Tándem.....	12
1.3.4 Descubrimiento de Patrones usando Árboles de Sufijo	12
1.3.5 Alineación de Trazas	13
1.3.6 Chequeo de Conformidad	14
1.3.7 Minería Difusa.....	15
1.3.8 Minería Basada en Teoría de Regiones.....	16
1.3.9 Descomposición en Subprocesos utilizando Bloques de Construcción.....	16
1.4 Marco de trabajo para la minería de proceso ProM.....	18
1.5 Consideraciones del capítulo	19
Capítulo 2: Minería de Variantes, una técnica para el diagnóstico de proceso	21
2.1 Definiciones.....	21
2.2 Técnica Minería de Variantes.....	24
2.2.1 Parámetros de configuración.....	25
2.2.2 Perfil de diagnóstico.....	25
2.2.3 Descomposición en subprocesos	26
2.2.4 Identificación de variantes del proceso	28
2.2.5 Descubrimiento de patrones de control de flujo.....	30
2.3 Incorporación de la técnica al marco de trabajo ProM	32
2.4 Consideraciones del capítulo	38
Capítulo 3: Utilización de la técnica Minería de Variantes para el diagnóstico de proceso	39
3.1 Aplicación a un registro de eventos del Desafío de Inteligencia de Procesos de Negocios 2013	39
3.1.1 Desafío de Inteligencia de Procesos de Negocios 2013.....	39

3.1.2 Analizando el registro de eventos y el mapa de proceso	40
3.1.3 Aplicación de la técnica Minería de Variantes	41
3.2 Aplicación a un registro de eventos proveniente de un entorno real	44
3.2.1 Entendiendo el entorno	44
3.2.2 Análisis de la variante de Selección Exclusiva	45
3.2.3 Análisis de la variante de Secuencia	54
3.2.4 Análisis de los resultados de la aplicación de técnicas de diagnóstico de proceso	58
3.3 Consideraciones del capítulo	63
Conclusiones	64
Referencias bibliográficas	66

Índice de figuras

Figura 1: Ciclo de Vida de BPM (AALST W. M. P., 2011)	2
Figura 2: Representación de los tres tipos de técnicas de minería de proceso: descubrimiento, conformidad y mejoramiento (AALST W. M. P., 2011).....	8
Figura 3: Análisis de Diagrama de Puntos.....	11
Figura 4: Alineación de Trazas.	14
Figura 5: Minería Difusa.....	15
Figura 6: Técnica Descomposición en Subprocesos utilizando Bloques de Construcción.....	17
Figura 7: Variantes de modelo de proceso (LI C., 2011).	23
Figura 8: Esquema representativo del funcionamiento de la técnica Minería de Variantes.....	24
Figura 9: Relación existente entre los complementos.....	33
Figura 10: Estructura del complemento de procesamiento.	34
Figura 11: Componentes del paquete patterns	36
Figura 12: Estructura del complemento de visualización.	37
Figura 13: Mapa de proceso identificado en (SEPPE K.L.M, 2013)	40
Figura 14: Resultado de la aplicación de la técnica Minería de Variantes sobre las 5 ejecuciones más frecuentes.	41
Figura 15: Variantes identificadas en el primer subproceso.	42
Figura 16: Descomposición del segundo subproceso.....	43
Figura 17: Sección del registro de eventos correspondiente al segundo subproceso.	44
Figura 18: Modelo del proceso Solicitar Producto.....	45
Figura 19: Variante del proceso mediante la descomposición por Selección Exclusiva.....	46
Figura 20: Evaluaciones del comportamiento presente en el registro de eventos.	47
Figura 21: Modelo de proceso del subproceso ver_detalle_sol_licitacion.....	48
Figura 22: Subproceso ver_detalle_sol_licitacion.....	48
Figura 23: Modelo de proceso del subproceso ver_detalle_sol_almacen.....	50
Figura 24: Subproceso ver_detalle_sol_almacen.....	50
Figura 25: Modelo de proceso del subproceso ver_detalle_sol_bq.....	52
Figura 26: Subproceso ver_detalle_sol_bq.....	53
Figura 27: Variante del proceso mediante la descomposición por Secuencia.	55
Figura 28: Evaluación del comportamiento presente en el registro de eventos.....	55
Figura 29: Primer subproceso en secuencia.	56
Figura 30: Segundo subproceso en secuencia.	57
Figura 31: Alineación de Trazas.	59
Figura 32: Codificación empleada en la Alineación de Trazas.	59
Figura 33: Descomposición en Subprocesos.....	61

Introducción

En el desarrollo de las organizaciones la gestión de sus procesos ha tomado cada vez mayor protagonismo. El enfoque basado en procesos, base para la gestión de estos, ha permitido la identificación y gestión sistemática de los procesos asociados a una organización así como sus interacciones.

Como soporte a esta forma de organizar las actividades han surgido los Sistemas de Gestión de Procesos de Negocios (BPMS¹). Estas herramientas brindan funcionalidades para controlar la ejecución de los procesos, evaluar su funcionamiento y apoyar la toma de decisiones. Sistemas como los de Planificación de Recursos Empresariales (ERP²) o Gestión de Relación con los Clientes (CRM³) son ejemplos de BPMS. A pesar de los costos de implantación, el uso de sistemas de este tipo ha impactado positivamente en el rendimiento de las empresas (HENDRICKS K., 2007).

Gran parte de estos sistemas brindan la posibilidad de almacenar trazas de las ejecuciones de los procesos que gestionan. Estos registros de ejecuciones de procesos, denominados **registros de eventos**, recopilan información sobre los estados que transcurren durante la ejecución de los procesos. El análisis de los datos compilados en los registros de eventos presenta una oportunidad para entender el funcionamiento real de una empresa a partir de los diferentes procesos que se llevan a cabo en ella y el orden en que se realizan. Con el fin de estudiar y aprovechar esta información surge la **minería de proceso**, disciplina que desarrolla técnicas y herramientas que permiten analizar los registros de eventos, extraer información a partir de ellos y presentar de forma explícita el conocimiento que contienen (AALST W. M. P., 2011).

Más de 100 casos de éxito apuntan a la minería de proceso como una alternativa útil para la gestión de procesos y la auditoría empresarial (DEPAIRE BENOIT, 2013). La minería de proceso brinda numerosas ventajas con respecto a las técnicas

¹ Del inglés: Business Process Management Systems.

² Del inglés: Enterprise Resource Planning.

³ Del inglés: Customer Relationship Management.

tradicionales de análisis y mejora de procesos, ofreciendo resultados en menor tiempo y con mayor fiabilidad (JANS MIEKE., 2011). Las técnicas de minería de proceso han sido incorporadas en sistemas de inteligencia y gestión de negocios tales como ARIS Process Performance Manager, Enterprise Visualization Suite o Disco (AALST W. M. P., 2011).

El ciclo de vida de la Gestión de Procesos de Negocio (BPM⁴) abarca las siete fases de un proceso de negocio. Estas fases son: implementación, (re)diseño, análisis, (re)configuración, ejecución, ajuste y diagnóstico. En la Figura 1 se muestran las fases del ciclo de vida de BPM y la relación existente entre ellas.

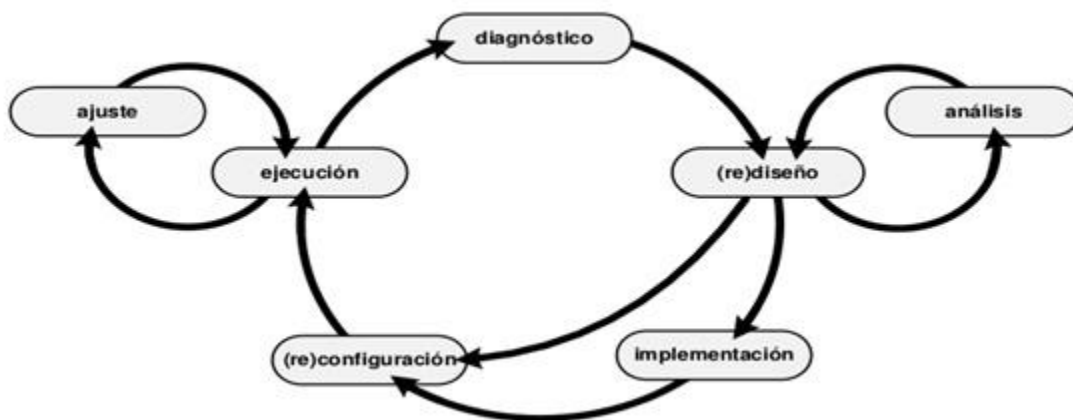


Figura 1: Ciclo de Vida de BPM (AALST W. M. P., 2011)

La fase de **diagnóstico de proceso** comprende el análisis de rendimiento, la detección de anomalías y la identificación de patrones comunes (BOSE R. P. J. C., 2012). El diagnóstico ayuda a tener una visión general del proceso, de sus aspectos más significativos y de las técnicas que pueden ser más útiles en su posterior análisis (YZQUIERDO-HERRERA R., 2012).

En la fase de diagnóstico del proceso se han empleado técnicas de **descubrimiento de proceso**, el mismo consiste en determinar modelos de procesos representativos del comportamiento expresado en un registro de eventos. Con este fin han surgido diferentes **algoritmos de descubrimiento** que permiten obtener modelos en

⁴ Del inglés: Business Process Management.

notaciones conocidas, entre las que se encuentran: Redes de Petri, Redes de Flujo de Trabajo o Redes Heurísticas (AALST W. M. P., 2004) (WEIJTERS A. J.M.M., 2006).

El descubrimiento de proceso es una tarea compleja y con disímiles retos, los cuales influyen en la calidad de los modelos descubiertos (LY L. T., 2012). El ruido, las tareas duplicadas y la ausencia de información son rasgos del registro de eventos que afectan los resultados de los algoritmos de descubrimiento (DE WEERDT J., 2012). Asimismo, la presencia en el proceso de patrones de control de flujo como el lazo y la elección no libre, representan una dificultad para algunos algoritmos (VAN DONGEN B., 2009).

En registros de eventos que presentan características como las descritas anteriormente los modelos de procesos obtenidos mediante la aplicación de algoritmos de descubrimiento contienen una elevada cantidad de nodos y relaciones, lo cual dificulta su análisis. Este tipo de modelo de proceso se conoce como **modelos poco estructurados** y son estereotipados como **modelos espaguetis** (AALST W. M. P., 2011).

Los registros de eventos provenientes de entornos reales suelen presentar ruido y ausencia de información (AALST W. M. P., 2003), causando que los algoritmos de descubrimiento presenten dificultades en la identificación de los patrones de control de flujo. La identificación de patrones de control de flujo permite determinar las actividades que se realizan sincrónicamente, los bloques de actividades que se repiten, el orden en que se ejecutan determinadas actividades y otros comportamientos relevantes en el proceso. El reconocimiento de los patrones de control de flujo posibilita mejorar la comprensión del funcionamiento general del proceso y su estructura.

El diagnóstico de proceso puede ser útil para tratar procesos poco estructurados, sin embargo, las técnicas desarrolladas en esta área presentan problemas para la identificación de los patrones de control de flujo más comunes, las desviaciones en el registro de eventos y la determinación de los niveles de jerarquización entre las actividades.

A partir de la situación problemática expuesta anteriormente se plantea como **problema a resolver**: la presencia en los registros de eventos de ruido, poca

estructuración y/o ausencia de información, dificultan la detección de patrones de control de flujo durante la fase de diagnóstico del proceso.

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

Como **objetivo general** de este trabajo se plantea: desarrollar un algoritmo que permita detectar los patrones de control de flujo en registros de eventos reales⁵ durante la fase de diagnóstico del proceso.

Como **campo de acción** se delimita el diagnóstico de proceso.

El objetivo general se puede desglosar en los siguientes **objetivos específicos**:

- Analizar las características que permiten la identificación de patrones de control de flujo en las técnicas de diagnóstico de proceso existentes.
- Diseñar un algoritmo que permita la detección de patrones de control de flujo en registros de eventos reales.
- Detectar la presencia de ruido y ausencia de información en los registros de eventos.
- Validar el algoritmo desarrollado a partir de la detección de patrones de control de flujo en registros de eventos reales.

Como **posibles resultados** a obtener mediante la realización de este trabajo se proponen:

- Descripción de un algoritmo capaz de detectar patrones de control de flujo en registros de eventos reales, considerando la presencia de ruido y ausencia de información.
- Implementación del algoritmo diseñado como complemento para el marco de trabajo para la minería de proceso ProM.

Durante el desarrollo del presente trabajo se definen como tareas específicas de la investigación las siguientes:

⁵ El término “registro de eventos reales”, en la literatura de minería de proceso, se refiere a registros de eventos que presentan características como ruido, tareas duplicadas, poca estructuración y/o ausencia de información.

- Análisis de los principales conceptos relacionados con la minería de proceso, el descubrimiento de proceso y el diagnóstico del proceso.
- Análisis del comportamiento de las principales técnicas de diagnóstico de proceso existentes.
- Caracterización de los patrones de control de flujo y su representación en un registro de eventos.
- Diseño de un algoritmo para la detección de patrones de control de flujo en registros de eventos reales.
- Implementación del algoritmo diseñado, permitiendo su integración al marco de trabajo ProM.
- Aplicación del algoritmo desarrollado en registros de eventos provenientes de entornos reales.
- Comparación de los resultados obtenidos de la aplicación del algoritmo con técnicas similares para el diagnóstico de proceso.

Para el desarrollo del presente trabajo se emplean diversos métodos científicos:

- **Analítico-sintético:** permite analizar individualmente los principales conceptos relacionados con el área de estudio a tratar, posibilitando un análisis profundo de cada uno, para luego llevar a cabo el estudio de las relaciones que se establecen entre ellos.
- **Hipotético-deductivo:** se utiliza para guiar la investigación desde el planteamiento del problema hasta la verificación de la solución a partir de las validaciones, orientando la secuencia lógica de las tareas que se realizan.
- **Histórico-lógico:** se emplea para orientar la investigación hacia los conocimientos existentes sobre las características de los patrones de control de flujo, cómo se expresan estos en los registros de eventos y cómo enfrentan el problema del manejo del ruido y la ausencia de información las técnicas de diagnóstico de proceso existentes.
- **Experimentación:** se emplea para verificar la consistencia entre los modelos descubiertos por el algoritmo y los modelos de proceso existente del proceso. Asimismo se emplea para verificar la consistencia entre los resultados obtenidos

con la aplicación del algoritmo propuesto y los resultados de la aplicación de técnicas de diagnóstico de proceso similares sobre el mismo registro de eventos.

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 de proceso y el diagnóstico de proceso. Se realiza un estudio crítico de las técnicas de diagnóstico de proceso existentes, 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: Minería de proceso y técnicas para el diagnóstico de proceso

En este capítulo se presentan los principales conceptos asociados a la minería de proceso. En el diagnóstico de proceso, una de las fases de la BPM, se han empleado técnicas de minería de proceso. La utilización de técnicas para el diagnóstico de proceso permite identificar los patrones de control de flujo, detectar las desviaciones presentes en el registro de eventos y ofrecer una vista general del proceso. Se realiza un análisis de las principales características de las técnicas de minería de proceso utilizadas para el diagnóstico, reflejando sus potencialidades y limitantes. ProM es un marco de trabajo para la minería de proceso concebido para admitir la adición de complementos y de esta manera posibilitar el desarrollo de nuevos algoritmos y técnicas.

1.1 Minería de proceso

La minería de proceso es una disciplina científica cuyo fin es descubrir, monitorear y mejorar los procesos reales, a través de la extracción de conocimiento de los registros de eventos, disponibles en los actuales sistemas de información (AALST W. M. P., 2011). Esta se enmarca dentro de la Inteligencia de Negocios (BI⁶) formando una conexión entre la Minería de Datos y el Modelado y Análisis de Procesos (AALST W. M. P., 2011).

El proceso de negocio es un concepto básico tanto en la minería de proceso como en la BI. De acuerdo a Mathias Weske, *“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 M., 2007).

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

⁶ Del Inglés: Business Intelligence

un proceso ejecutadas en un período determinado de tiempo (MA L., 2012). Las ejecuciones del proceso se denominan trazas.

Existen tres tipos fundamentales de técnicas de minería de proceso, el descubrimiento de proceso, la verificación de conformidad y el mejoramiento de modelos, como se muestra en la Figura 2.



Figura 2: Representación de los tres tipos de técnicas de minería de proceso: descubrimiento, conformidad y mejoramiento (AALST W. M. P., 2011).

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. La utilización de técnicas de verificación se emplea 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 de modo que refleje mejor lo que ocurre en la realidad de su ejecución.

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 (AALST W. M. P., 2011). La principal motivación del descubrimiento de proceso es la utilidad de los modelos obtenidos a partir de la ejecución real del

proceso. Los modelos descubiertos pueden ser representados utilizando múltiples notaciones, dentro de las que se encuentran Redes de Petri y Redes de Flujo de Trabajo (AALST W. M. P., 2011).

La función que mapea un registro de eventos hacia un modelo de proceso se denomina **algoritmo de descubrimiento** (AALST W. M. P., 2011). Existen varios algoritmos desarrollados entre los que se pueden mencionar: Alpha (AALST W. M. P., 2004), Genetic Miner (DE MEDEIROS A. K. A., 2007), Heuristic Miner (WEIJTERS A.J.M.M., 2003), Transition System Miner o Fuzzy Miner (GÜNTHER C. W., 2007).

Los registros de eventos presentan características como ruido, ausencia de información y/o presencia de patrones de control de flujo complejos. Los algoritmos de descubrimiento de procesos no manejan estas características en la misma medida, provocando, en algunos casos, la obtención de modelos de proceso de baja calidad (FUNDORA-RAMÍREZ O., 2013).

1.2 Diagnóstico de proceso

El diagnóstico de proceso permite detectar los patrones comunes presentes en un registro de eventos, así como la existencia de desviaciones en el proceso (BOSE R. P. J. C., 2012). 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 R., 2012).

El empleo de técnicas de diagnóstico de proceso permite conocer los comportamientos de mayor frecuencia de ocurrencia en el registro de eventos, lo que posibilita dirigir las técnicas de mejora hacia los elementos más críticos del proceso. Al mismo tiempo, la detección de ejecuciones anómalas existentes en el registro de eventos brinda información acerca de las posibles violaciones del modelo. La presencia de desviaciones en el modelo constituye un indicador de posibles fraudes o violaciones en las políticas y suele ser útil a las organizaciones en la realización de auditorías. El diagnóstico de proceso permite identificar el contexto en el que se ejecuta una actividad

en el registro de eventos y reconocer la existencia de dependencias entre las actividades que conforman el proceso (BOSE R. P. J. C., 2012).

Durante la fase de diagnóstico resulta importante determinar la presencia en el registro de eventos de patrones de control de flujo. En este sentido, uno de sus principales objetivos 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 R. P. J. C., 2012).

1.3 Técnicas de diagnóstico de proceso

El diagnóstico de proceso brinda información útil sobre determinados aspectos de la estructura del proceso. Con el objetivo de extraer y mostrar dicha información se han desarrollado un conjunto de técnicas útiles para el diagnóstico de proceso. 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 M., 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. El Análisis de Diagramas de Puntos 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 (SONG M., 2007). En la Figura 3 se muestra un resultado de la aplicación de esta técnica.

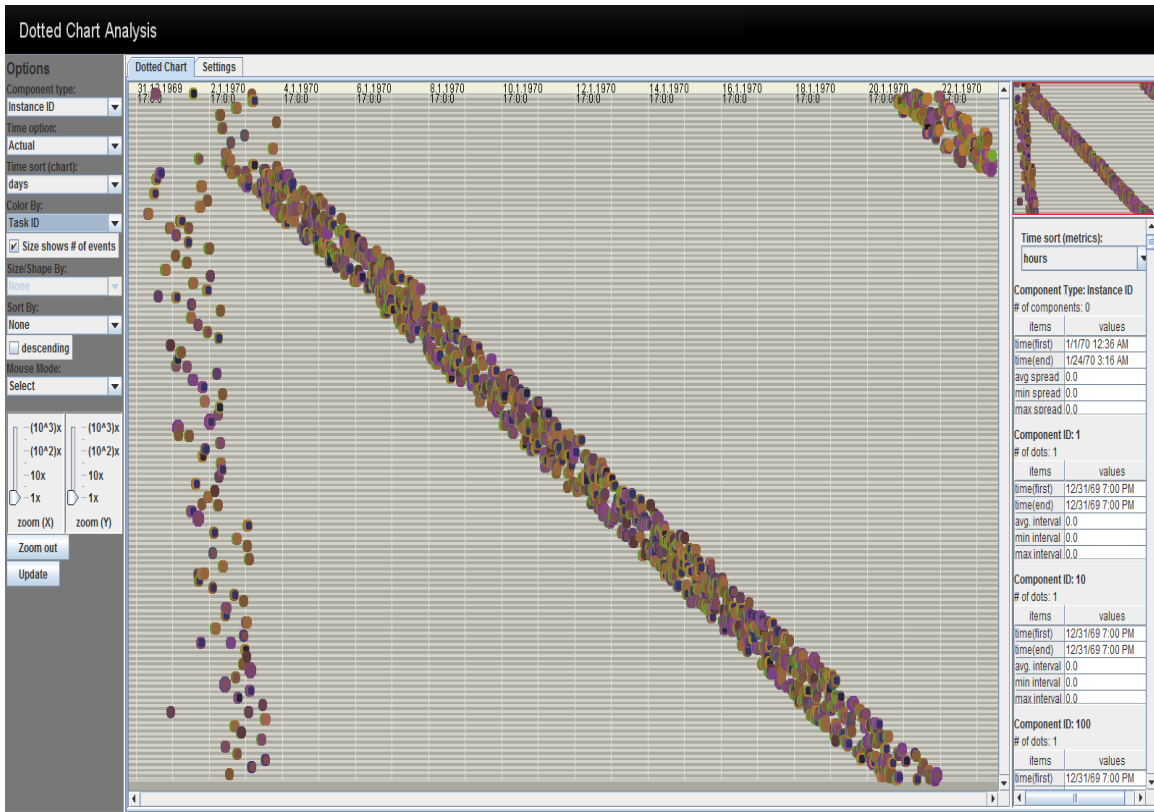


Figura 3: Análisis de Diagrama de Puntos.

El Análisis de Diagramas de Puntos presenta algunos indicadores de desempeño como 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 complejizan. Se considera que todos los eventos pertenecen al mismo nivel de abstracción en el registro de eventos, lo que dificulta la identificación de los subprocessos que componen el proceso.

1.3.2 Visualización de Flujo y Alcance

Visualización de Flujo y Alcance⁷ (GÜNTHER C .W., 2009) es una técnica de visualización de trazas cuya base es la existencia de correlaciones de clases de eventos. El empleo de esta técnica permite la detección de los patrones de los eventos co-ocurrentes debido a su cercanía. Mediante el uso de la técnica Visualización de Flujo y Alcance, cada una de las trazas que componen el registro de eventos se visualiza por separado.

Dentro de las limitaciones de esta técnica se encuentra la ausencia de una visión global del proceso, lo que dificulta enmarcar claramente el proceso en un contexto, asimismo, aunque la técnica permite la detección de patrones recurrentes no permite relacionarlos. Esta técnica determina modelos de procesos donde no se establece el nivel de abstracción de las actividades que los componen, por lo que cada actividad se considera que pertenece solamente a un determinado patrón. Este enfoque dificulta la realización de análisis de los procesos de modo jerárquico.

1.3.3 Arreglos Tándem

En (BOSE R. P. J. C., 2009) se propone el uso de Arreglos Tándem⁸ y repeticiones máximas para capturar patrones recurrentes dentro y a través de las trazas. 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 situación en (LI C., 2010) 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 existentes entre los patrones deben ser descubiertas por separado.

1.3.4 Descubrimiento de Patrones usando Árboles de Sufijo

Este tipo de técnica emplea la concatenación de las trazas existentes en el registro de eventos y la posterior aplicación de algoritmos de identificación de patrones sobre el resultado obtenido. Para la determinación de los límites de cada una de las trazas es

⁷ Del inglés: Stream Scope Visualization.

⁸ Del inglés: Tandem Arrays.

necesaria la existencia de un delimitador distinto de cualquiera de las actividades presentes en el registro de eventos.

La concatenación de las trazas puede traer consigo la obtención de una secuencia demasiado larga. En (CHEUNG C.F., 2005) se propone realizar la construcción de un Árbol de Sufijos Eficiente para el manejo de situaciones similares. Este tipo de técnica posibilita la determinación de patrones de control de flujo, sin embargo, el diagnóstico se encuentra afectado por la ausencia de variables de rendimiento del proceso y la suposición de que todas las actividades en el registro de eventos se encuentran en el mismo nivel de abstracción.

1.3.5 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 (BOSE R. P. J. C., 2012). Esta técnica se basa en la Alineación de Múltiples Secuencias (MSA⁹), 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 de un proceso en un registro de eventos.

Para realizar la Alineación de Trazas se ha empleado 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., 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.

Una de las potencialidades de la Alineación de Trazas consiste en posibilitar el descubrimiento de desviaciones en la ejecución del proceso, es decir, identificar las ocurrencias anómalas del proceso de las ocurrencias comunes. Dicha técnica permite la visualización de múltiples ejecuciones del proceso simultáneamente y facilita la

⁹ Del inglés: Multiple Sequence Alignment.

determinación visual de los patrones de control de flujo. La Figura 4 muestra un conjunto de trazas alineadas.

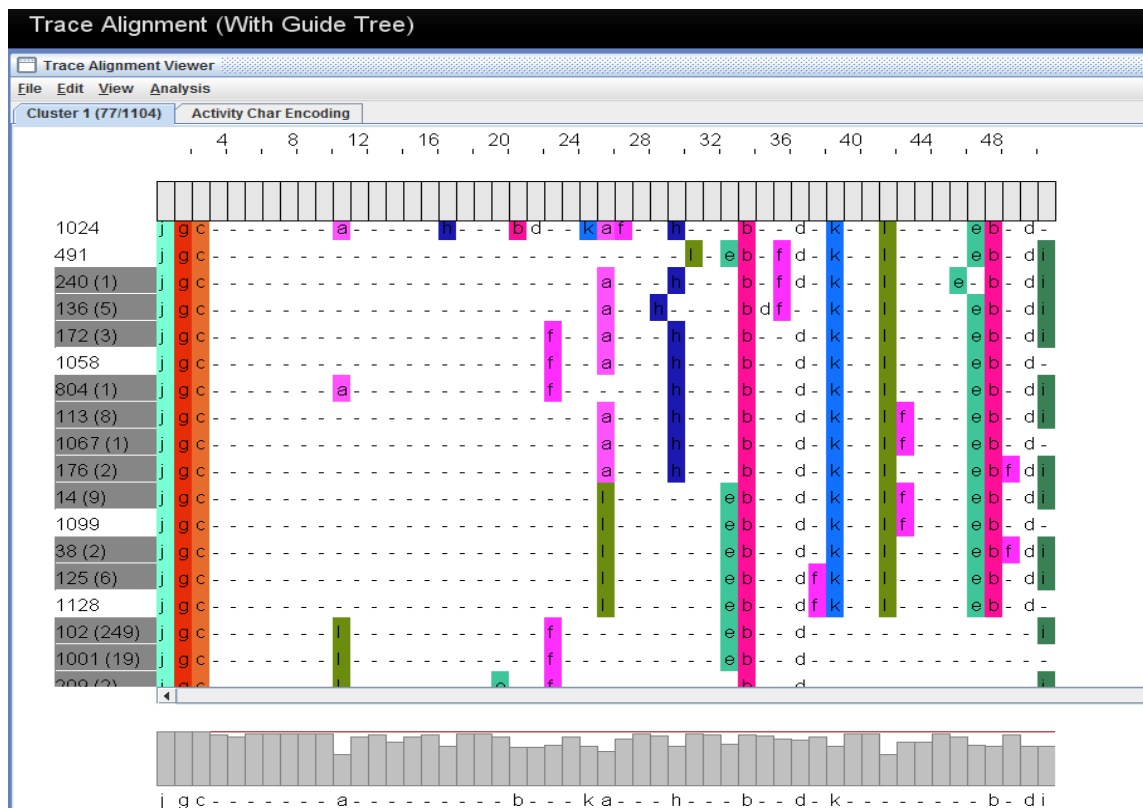


Figura 4: Alineación de Trazas.

Sin embargo, esta técnica no permite detectar los subprocessos que conforman al proceso analizado y enmarcar en estos las anomalías y patrones identificados, lo que dificulta la contextualización del aspecto detectado y el entendimiento de las causas que lo originaron (YZQUIERDO-HERRERA R., 2012). 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., 2012).

1.3.6 Chequeo de Conformidad

El Chequeo de Conformidad (AALST W. M. P., 2011) permite la identificación de las desviaciones y anomalías en el proceso ejecutado mediante la comparación del modelo descubierto con un modelo teórico existente (ROZINAT A., 2008) (ADRIANSYAH A., 2011).

Dentro de las limitaciones para la aplicación de esta técnica se encuentra la asunción de la existencia del modelo de proceso, obviando la posibilidad de que el modelo no exista o que presente problemas (BOSE R. P. J. C., 2012). Existe la alternativa de obtener el modelo de proceso a partir del registro de eventos y partiendo de dicho modelo, aplicar la técnica. Esta variante puede no arrojar resultados satisfactorios en los casos de procesos que presenten poca estructuración y/o se encuentran en un ambiente flexible. En situaciones similares la calidad del modelo obtenido afectaría el resultado de la aplicación de la técnica (BOSE R. P. J. C., 2012). Por otra parte, aunque se disponga del modelo, existe la posibilidad de que debido al alto nivel de complejidad y el tamaño del mismo sea difícil identificar y localizar problemas.

1.3.7 Minería Difusa

La Minería Difusa es una técnica de diagnóstico que muestra las actividades y sus relaciones según diferentes niveles de abstracción. La misma utiliza similitudes con la forma de representar la información en la cartografía, aplicando conceptos como: abstracción, personalización, agregación y énfasis (GÜNTHER C. W., 2007). 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 5.

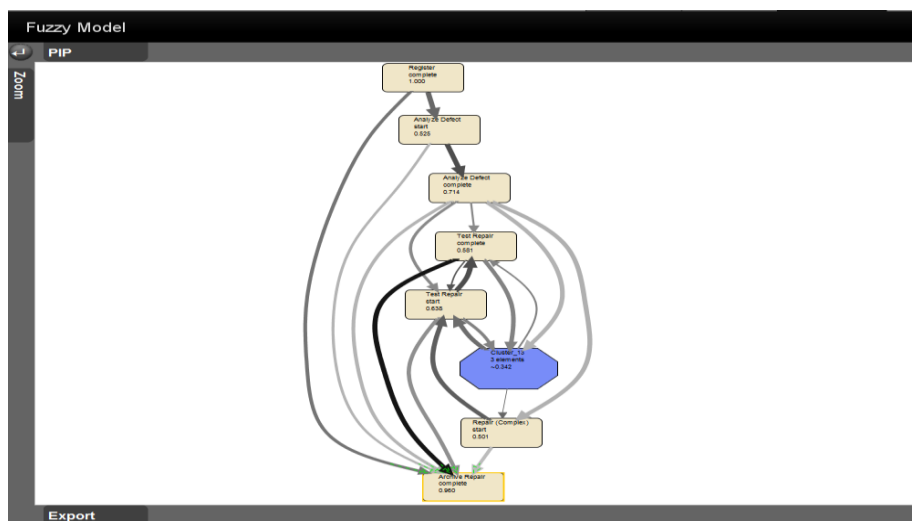


Figura 5: Minería Difusa.

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. Para facilitar la visualización de la información relevante características como color, contraste, saturación y tamaño son destacadas en determinados componentes. Por su parte la personalización permite configurar la información que se muestra de acuerdo a determinados criterios (GÜNTHER C. W., 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. Esta técnica no permite la identificación de patrones de control de flujo mostrando solo las relaciones existentes entre actividades.

1.3.8 Minería Basada en Teoría de Regiones

En (AALST W. M. P., 2009) se describe una técnica donde la relación existente entre los clústeres y las actividades es de muchos a muchos. La técnica construye un espacio de búsqueda utilizando la información presente en el registro de eventos sobre el cual aplica técnicas de teoría de regiones para construir una red de Petri.

La teoría de regiones presenta complejidad exponencial con respecto al tamaño del espacio de búsqueda, lo que provoca que la aplicación de esa técnica no sea recomendable en registros de eventos provenientes de entornos reales (DONGEN B.F.VAN, 2009). Esta técnica tiene como limitación adicional que solo se consideran Redes de Petri como modelo de proceso, y no considera otras notaciones como Redes de Flujo de Trabajo y Redes Heurísticas.

1.3.9 Descomposición en Subprocesos utilizando Bloques de Construcción

La Descomposición en Subprocesos 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 R., 2012). En la Figura 6 se muestra el resultado de la aplicación de esta técnica.

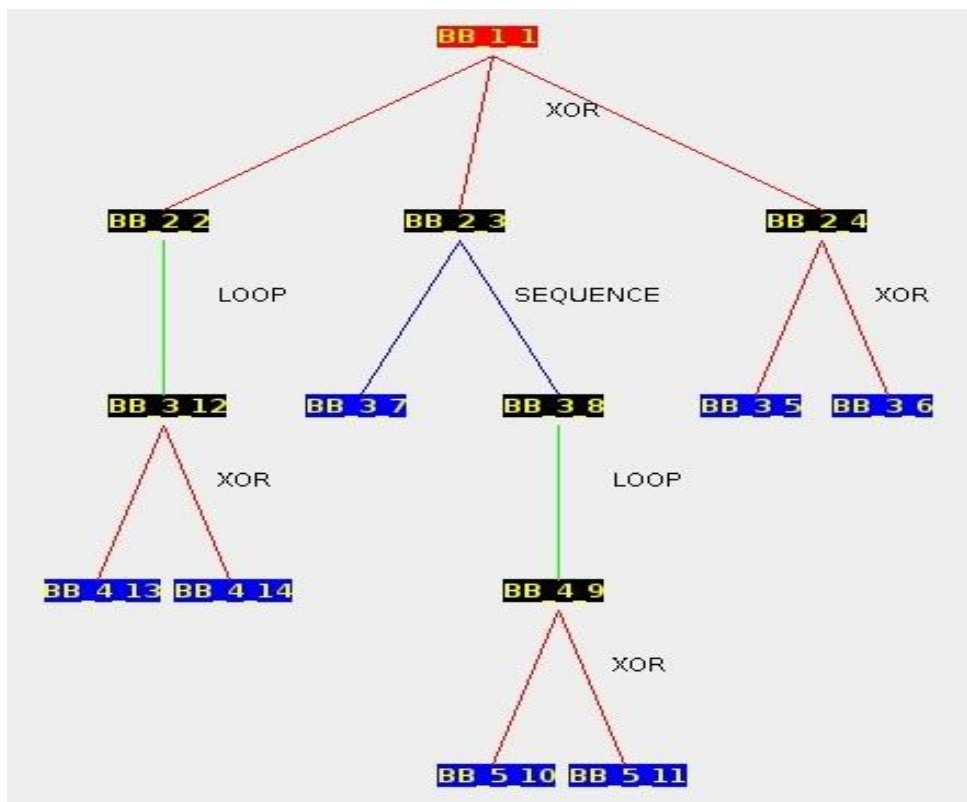


Figura 6: Técnica Descomposición en Subprocesos utilizando Bloques de Construcción.

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 (YZQUIERDO-HERRERA R., 2012). Mediante la utilización de la Descomposición en Subprocesos, 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 R., 2012).

La Descomposición en Subprocesos presenta como limitante que al utilizar como base la Alineación de Trazas, 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 Subprocesos 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.

1.4 Marco de trabajo para la minería de proceso ProM

ProM es un marco de trabajo para el desarrollo de herramientas de minería de proceso en un ambiente estandarizado (YANG H., 2012). Está desarrollado en Java y se encuentra disponible bajo licencia GPL¹⁰. ProM está concebido para admitir la adición de complementos y de esta manera posibilitar el desarrollo de nuevos algoritmos y técnicas en el campo de la minería de proceso. Los complementos 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. Este marco de trabajo cuenta con más de 550 complementos, cada uno de los cuales posibilita realizar diferentes análisis (AALST W. M. P., 2013). Las herramientas desarrolladas en ProM han sido empleadas en el análisis de procesos provenientes de diferentes dominios entre los que se encuentran gubernamental, hospitalario y sistemas ERP (AALST W. M. P., 2011).

¹⁰ Del inglés: GNU Public License.

1.5 Consideraciones del capítulo

Durante este capítulo se analizaron los conceptos relacionados con la minería de proceso y el diagnóstico de proceso. Se expusieron las características de diferentes técnicas de diagnóstico de proceso evidenciando sus ventajas y debilidades. Además se presentaron las principales características del marco de trabajo para la minería de proceso ProM.

Dentro de las principales limitaciones identificadas en las técnicas analizadas se encuentra que no son capaces de determinar los patrones de control de flujo y solo muestran una vista general del proceso. Las técnicas que presentan esta limitante son el Análisis de Diagramas de Puntos y la Visualización de Flujo y Alcance. Por otra parte, la utilización de Arreglos Tándem y el Descubrimiento de Patrones usando Árboles de Sufijos son capaces de identificar los patrones de control de flujo, pero no permiten la visualización de las relaciones o dependencias entre ellos. La Minería Difusa solo considera relaciones entre eventos y clústeres de uno a muchos y no permite la identificación de patrones de control de flujo. Por su parte, la Minería Basada en Teoría de Regiones, aunque considera relaciones de muchos a muchos entre clústeres y eventos, presenta un alto costo computacional creando cuellos de botella en su ejecución.

La Alineación de Trazas muestra una vista del registro de eventos con la cual se pueden obtener las desviaciones del proceso y distinguir entre ocurrencias comunes, pero no enmarca las anomalías y desviaciones identificadas en subprocesos, lo que dificulta su contextualización. La Descomposición en Subprocesos utilizando Bloques de Construcción no es capaz de manejar características presentes en los registros de eventos como ruido o ausencia de información. La utilización de la Alineación de Trazas como punto de partida provoca un incremento adicional en el tiempo de aplicación de la técnica.

Debido a las limitaciones de las técnicas existentes para lograr una adecuada comprensión tanto de la estructura del proceso como del comportamiento y los aspectos relevantes en el registro de eventos es necesario combinar la aplicación de

varias de las técnicas analizadas. La aplicación de múltiples técnicas para el diagnóstico del proceso representa un incremento en el costo temporal del análisis.

El análisis de las técnicas de diagnóstico permitió identificar las potencialidades presentes en algunas de ellas. De modo general las técnicas permiten obtener una vista global del proceso. La técnica Análisis de Diagramas de Puntos provee información relacionada con el rendimiento del proceso. La Minería Difusa permite el manejo del ruido presente en el registro de eventos mediante la utilización de umbrales de permisibilidad. Por otra parte, la Descomposición en Subprocesos permite la contextualización de los patrones de control de flujo identificados en los diferentes subprocesos.

Las deficiencias identificadas evidencian la necesidad de una herramienta de diagnóstico de proceso capaz de identificar y manejar las principales características del registro de eventos, determinar los patrones de control de flujo presentes y permitir la visualización de los distintos niveles de abstracción del proceso. La combinación de estas características solventaría las limitantes de las técnicas de diagnóstico de proceso disponibles actualmente.

Capítulo 2: Minería de Variantes, una técnica para el diagnóstico de proceso

En este capítulo se presenta la técnica desarrollada, denominada Minería de Variantes. La Minería de Variantes pretende construir diferentes variantes de modelos de proceso a partir de la descomposición en subprocesos considerando los patrones de control de flujo. Se utiliza un patrón de control de flujo para descomponer un subproceso descartando el comportamiento poco frecuente y/o asumiendo la presencia de comportamiento, en un ámbito definido por parámetros de configuración. La técnica propuesta ha sido desarrollada como complemento para el marco de trabajo para la minería de proceso ProM.

2.1 Definiciones

Existen un conjunto de definiciones las cuales son necesarias formalizar para posibilitar la comprensión de la técnica desarrollada.

Definición 1. Traza y registro de eventos: denotemos A_P 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 .

Definición 2. Subproceso: un subproceso es una encapsulación de las actividades del negocio la cual representa una unidad de trabajo lógica y coherente. Los subprocesos poseen 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.

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 ellos.

- 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 repiten 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**.*

Denotemos como S_p el conjunto de los subprocesos del proceso P y como $W = \{w_s, w_x, w_o, w_p, w_l\}$ el conjunto de los patrones de control de flujo. Un $s_i \in S_p$ se puede descomponer a través de diferentes $w | \forall w_j \in W$ en varios $s_k | \forall k \neq i : s_k \in S_p$, hasta el nivel de actividad. Se construye entonces un árbol en el que el mayor nivel de abstracción se encuentra en la raíz y el menor en las hojas.

Definición 3. Comportamiento del proceso: *denotemos como $l_i \subseteq L_p$ a la sección del registro de eventos relacionada al subproceso $s_i \in S_p$. El comportamiento del patrón de control de flujo w_j para el subproceso $s_i : B_{w_j s_i} | w_j \in W$, está compuesto por las posibles:*

- Sucesiones directas (SD): *se denomina sucesión directa entre la actividad **a** y la actividad **b** a la existencia en alguna traza de la secuencia **ab**.*
- Sucesiones indirectas (SI): *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: *son los eventos que aparecen de primeros y/o últimos en alguna de las trazas.*

- Eventos repetidos en las trazas: son los eventos que aparecen en más de una ocasión en una misma traza.

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

Definición 4. Variantes de proceso: las variantes de un modelo de proceso o variantes de proceso, son modelos de un proceso que describen el mismo proceso de negocio, y poseen 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.

En la Figura 7 se aprecian dos variantes (V_1 y V_2) de un proceso de negocio. En V_1 la actividad Y se muestra en secuencia con el subproceso conformado por las actividades E, A, X y F. Mientras que en V_2 la actividad Y aparece en Selección Exclusiva con el subproceso antes mencionado. Además la actividad Z aparece en V_1 pero no en V_2 .

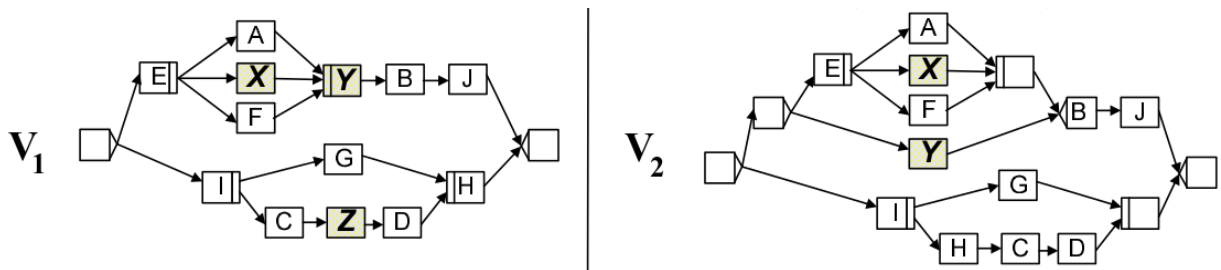


Figura 7: Variantes de modelo de proceso (LI C., 2011).

Definición 5. Árbol de variantes: un árbol de variantes es una representación de diferentes descomposiciones en subprocesos aplicadas al proceso P a partir de un registro de eventos. Este árbol está compuesto por dos tipos de nodos, los nodos subproceso y los nodos patrón.

Un nodo subproceso representa un subproceso sobre P y puede o no contener nodos patrones como hijos. 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.

2.2 Técnica Minería de Variantes

La idea general de la técnica Minería de Variantes consiste en encontrar las posibles variantes de descomposición de un subproceso bajo determinados criterios impuestos por el usuario. Estos criterios son definidos en términos de umbrales para ruido (T_n) y completitud (T_c). A partir del registro de eventos y los umbrales definidos se genera el árbol de variantes. Al mismo tiempo se obtiene un perfil de diagnóstico para el árbol descubierto. La Figura 8 muestra una representación del modelo del funcionamiento de la técnica.

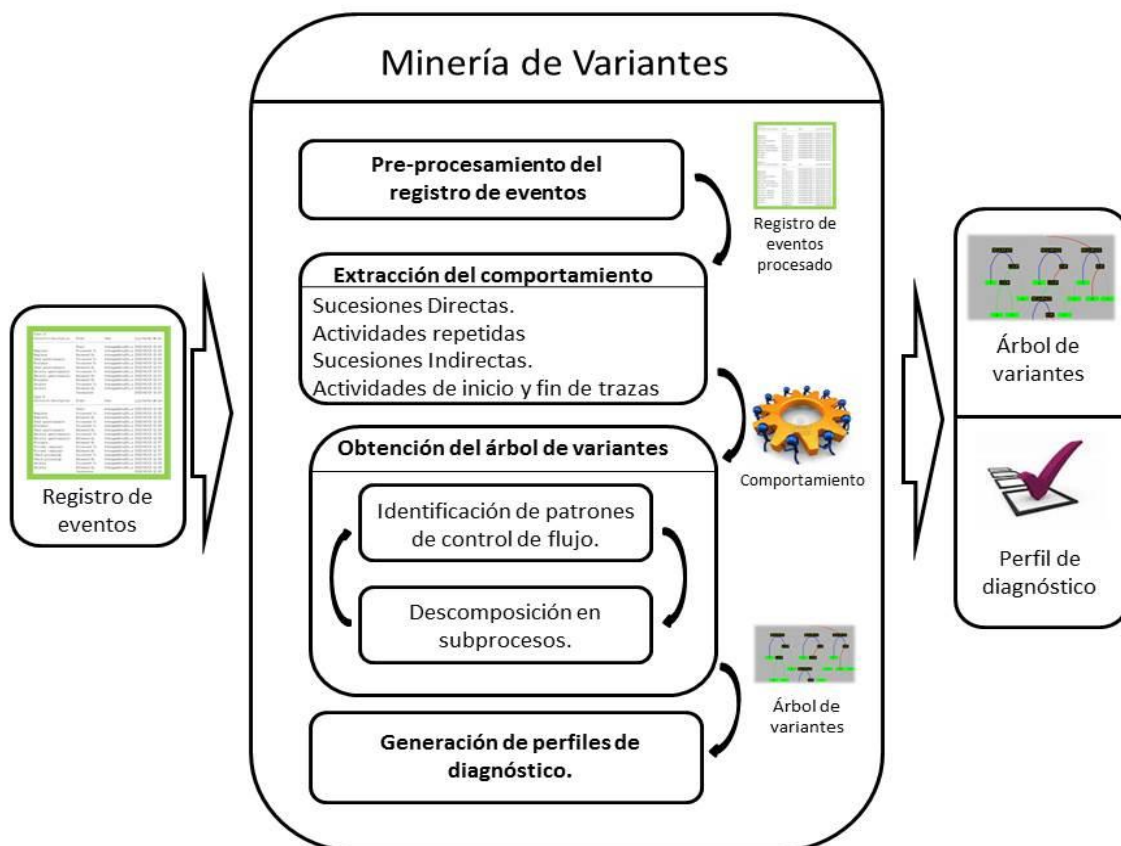


Figura 8: Esquema representativo del funcionamiento de la técnica Minería de Variantes.

Durante la etapa de pre-procesamiento se extraen y codifican las actividades presentes en el registro de eventos. Son agrupadas las trazas con igual secuencia de actividades y posteriormente ordenadas de acuerdo a su frecuencia. Además se detectan las características generales presentes en el registro de eventos, como la cantidad de trazas diferentes y la cantidad de eventos diferentes. Posteriormente se procede a la extracción del comportamiento identificativo de los patrones de control de flujo, contenido en el registro de eventos ($B_{l_i} = B_{w_s l_i} \cup B_{w_x s_i} \cup B_{w_o s_i} \cup B_{w_p s_i} \cup B_{w_l s_i} \mid w_j \in W$). Para la obtención del árbol de variantes se realizan dos procedimientos fundamentales, la identificación de los patrones de control de flujo presentes en el subproceso y luego la descomposición del subproceso según los patrones identificados. A partir de cada nodo en el árbol de variantes se construye el perfil de diagnóstico.

2.2.1 Parámetros de configuración

Los umbrales de permisibilidad 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 predeterminados están basados en el Principio de Pareto, el cual ha sido utilizado en técnicas de minería de proceso (LEEMANS S.J.J., 2013). En la Minería de Variantes se utilizan como configuración por defecto el 20% para ruido y el 80% de completitud. Por ejemplo, un umbral de ruido definido en 20% para el patrón Secuencia significa que puede ser descartado hasta el 20% del comportamiento expresado en el registro de eventos para encontrar una descomposición mediante este patrón. Por otra parte, una configuración del 80% en el umbral de completitud para Paralelismo indica que el subproceso puede ser descompuesto mediante ese patrón conteniendo solamente el 80% del comportamiento necesario para que se manifieste.

2.2.2 Perfil de diagnóstico

En el árbol de variantes se genera una información de diagnóstico ($d_{i w_j}$) asociada a cada nodo patrón ($n_{i w_j}$). El perfil de diagnóstico está compuesto por todas las informaciones de diagnóstico en el árbol de variantes. Cada $d_{i w_j}$ contiene una estimación de aptitud, una estimación de precisión, el comportamiento descartado y el comportamiento asumido. Los comportamientos identificados ($B_{w_j l_i}$) son referidos al

subproceso (s_i), cuyo padre es el nodo patrón n_{iw_j} de acuerdo al patrón w_j en el subproceso relacionado a s_i . La relación entre el comportamiento descartado y el comportamiento del patrón w_j expresado en l_i ($\beta_{w_j l_i}$) se utiliza para realizar la estimación de aptitud. De la misma manera, la estimación de precisión está basada en el comportamiento asumido y $\beta_{w_j l_i}$. Ambas estimaciones son expresadas como un valor entre 0 y 1, donde, mientras mayor el número, mejor es la evaluación.

El perfil de diagnóstico muestra además la manifestación del patrón de Selección no Libre. Las relaciones de dependencias indirectas y relaciones incorrectas son descritas utilizando reglas de negocio. La búsqueda para este patrón es realizada a partir de cada versión obtenida en el proceso de descomposición. Un ejemplo de regla de negocio es: **a** y **b** entonces **c**, lo que se puede interpretar como que si ocurren las actividades **a** y **b** entonces ocurre la actividad **c**.

2.2.3 Descomposición en subprocesos

El algoritmo de descomposición en subprocesos utiliza el registro de eventos del proceso P y los umbrales definidos para construir el árbol de variantes V_P . Para cada nuevo subproceso en $S'_p \subseteq S_p$, comenzando por P , se busca una variante de descomposición para cada w_j .

Con el objetivo de encontrar una variante de descomposición para un subproceso s_i , se extrae del registro de eventos el comportamiento del proceso β_{l_i} relacionado con el subproceso s_i . El comportamiento del proceso para el patrón w_j en s_i , que no está presente en β_{l_i} pero puede ser asumido dentro del umbral de completitud, se denota como $\beta'_{w_j s_i}$. Asimismo el comportamiento del proceso que está presente en β_{l_i} y puede ser descartado dentro del umbral de ruido es denotado como $\beta'_{l_i} \subset (\beta_{l_i} \setminus \beta_{w_j l_i})$. En el subproceso s_i es posible encontrar una variante para el patrón de control de flujo w_j si $(\beta_{w_j l_i} \setminus \beta'_{l_i}) \cup \beta'_{w_j s_i}$ es equivalente a $\beta_{w_j s_i}$.

Un perfil de diagnóstico es generado para cada variante encontrada. El nuevo subproceso generado por cada descomposición es añadido a S'_p con el objetivo de que

sea analizado posteriormente. Un subproceso s_i es analizado solo si el conjunto de actividades que lo componen, denotado por A_{s_i} , contiene más de una actividad.

A continuación se muestra el pseudocódigo del algoritmo de descomposición en subprocesos.

Algoritmo 1: Descomposición en subprocesos.

Entradas: registro de eventos (L_P), umbrales de ruido y completitud (T_P)

Salidas: árbol de variantes (V_P), perfil de diagnóstico.

1: **procedure** DECOMPOSE(L_P, T_P)

2: $W \leftarrow \{w_s, w_x, w_o, w_p, w_l\}$ «Patrones de control de flujo para la descomposición.

3: $n_{s_1} \leftarrow$ CREATESUB-PROCESS(L_P) «Crea el nodo raíz.

4: $V_P \leftarrow n_{s_1}$

5: $S'_p \leftarrow \{n_{s_1}\}$ «Crea una lista para los subprocesos nuevos.

6: **while** $n_{s_i} \leftarrow$ extraer el primer nodo subproceso existente en S'_p **do**

7: **if** s_i no es una actividad **then**

8: $\beta_{l_i} \leftarrow$ EXTRACTBEHAVIOR(n_{s_i}) «Extrae el comportamiento contenido en l_i .

9: **for all** w_j in W **do**

10: $V \leftarrow$ FINDVARIANTS(w_j, β_{l_i}, T_P)

11: **if** $|V| > 0$ **then** «Si n_{s_i} puede ser descompuesto por w_j bajo T_P .

12: **for all** nodo subproceso k en V **do**

13: adicionar el nodo k como hijo del nodo patrón $n_{i w_j}$

14: adicionar el nodo k a S'_p

15: **end for**

16: adicionar el nodo $n_{i w_j}$ como hijo del nodo subproceso n_{s_i} .

17: $d_{i w_j} \leftarrow$ CREATEDIAGNOSISINFORMATION(W_S, l_i, V)

18: adicionar $d_{i w_j}$ al perfil de diagnóstico

19: **end if**

20: **end for**

21: **end if**

22: **end while**

23: **end procedure**

2.2.4 Identificación de variantes del proceso

Una posible descomposición del subprocesos s_i utilizando el patrón de control de flujo w_j se denota como $\gamma_{w_j s_i}$. Del mismo modo se denota como A_{s_i} al conjunto de actividades presentes en s_i . Una posible descomposición $\gamma_{w_j s_i}$ contiene n conjuntos disjuntos de actividades, subconjuntos de A_{s_i} . En la sección del registro de eventos l_i , para el patrón de control de flujo w_j , se procesa cada elemento $b \in \beta_{w_j l_i}$. El procesamiento se realiza mediante una búsqueda sobre b ordenados descendientemente por su frecuencia ($\varphi(b)$).

Cada nodo en el espacio de búsqueda está definido por $Y_{w_j s_i}$, el comportamiento no procesado ($\beta'_{w_j l_i} \subseteq \beta_{w_j l_i}$), el comportamiento descartado ($n_{i w_j}$) y el comportamiento asumido ($\theta_{i w_j}$).

Existen dos situaciones en las que se genera un nuevo nodo. La primera ocurre al procesar el comportamiento $b_x \in \beta'_{w_j l_i}$, y se genera una nueva descomposición $Y'_{w_j s_i} : Y'_{w_j s_i} \neq Y_{w_j s_i}$. La segunda situación se produce si al procesar el comportamiento b_x se puede generar un $Y'_{w_j s_i}$ diferente, donde b_x puede ser descartado bajo el umbral de ruido o un $b_y \in \beta'_{w_j}$ puede ser asumido con el objetivo de mantener $Y_{w_j s_i}$.

El costo del camino desde un nodo a alguno de sus vecinos es $c = c(b_x) + c(b_y)$ donde $c(b_x) = \frac{\varphi(b_x)}{\sum_{n=1}^{|\beta_{w_j l_i}|} \varphi(b_n)}$, si $b_x \in \beta'_{w_j l_i}$ es descartado y $c(b_y) = \frac{\varphi(b_y)}{\sum_{n=1}^{|\beta_{w_j}|} \varphi(b_n)}$, si $b_y \in \beta'_{w_j}$ es asumido. El objetivo se alcanza cuando el comportamiento $\beta_{w_j l_i}$ ha sido procesado y la posible descomposición $Y_{w_j s_i}$ tiene más de un conjunto disjunto. A continuación se describe en pseudocódigo el algoritmo que realiza la búsqueda de las variantes a partir del procesamiento de los comportamientos.

Algoritmo 2: Búsqueda de variantes.

Entradas: patrón de control de flujo (w_j), comportamiento del proceso (β_{l_i}), umbrales de ruido y completitud (T_p)

Salidas: variante de descomposición en subproceso según el patrón de control de flujo w_j

```
1: procedure FINDVARIANTS( $w_j, \beta_{l_i}, T_p$ )
2:    $\beta'_{w_j l_i} \leftarrow$  SORTBYFREQUENCY( $\beta_{l_i}[w_j]$ ) «Ordena comportamiento en orden descendente
3:    $T_{W_j} \leftarrow T_p [w_j]$ . «Umbrales para el patrón  $w_j$ 
4:    $\gamma_{w_j s_i} \leftarrow$  CREATEFIRSTDECOMPOSITIONPROPOSAL( $w_j, \beta_{w_j l_i}, T_{W_j}$ )
5:   if  $\gamma_{w_j s_i} = \emptyset$  then . « $\beta_{w_j l_i}$  no es suficiente para crear una propuesta bajo  $T_{W_j}$ 
6:     return  $\emptyset$ 
7:   end if
8:    $\sigma_1 \leftarrow$  CREATESearchNode( $\beta_{w_j l_i}, \gamma_{w_j s_i}$ )
9:    $\Sigma \leftarrow \{\sigma_n\}$ . «Cola de nodos abiertos ordenada por prioridad según el costo del camino.
10:  while  $\sigma_n \leftarrow$  extraer primer nodo en  $\Sigma$  do
11:    if  $\sigma_n$  es meta &  $\beta'_{w_j l_i} = \emptyset$  then
12:       $g \leftarrow \sigma_n$ 
13:      break
14:    else if  $\sigma_n$  no está cerrado then
15:       $\Sigma_n \leftarrow$  SEARCHNEIGHBORS( $w_j, \sigma_n, \beta'_{w_j l_i}, T_{W_j}$ )
16:       $\Sigma \leftarrow \Sigma + \Sigma_n$ . «Adicionar vecinos como nodos abiertos.
17:    end if
18:  end while
19:  if  $g \neq null$  then
20:     $\gamma_{w_j s_i} \leftarrow$  descomposición potencial contenida en el nodo  $g$ 
21:    return CREATENEWSUB-PROCESSES( $w_j, \gamma_{w_j s_i}, \beta_{l_i}$ )
22:  else
23:    return  $\emptyset$ 
24:  end if
25: end procedure
```

La optimalidad y la 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 es más probable que no puedan ser descartados según los umbrales. Además posibilita 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.

El procedimiento anterior permite generalizar la búsqueda de descomposiciones para los patrones de control de flujo $w_j \in W$. El procedimiento donde se especifica el modo de obtener los nodos generados para cada w_j es el procedimiento **SEARCHNEIGHBORS**.

2.2.5 Descubrimiento de patrones de control de flujo

Cada patrón de control de flujo presenta comportamientos característicos lo que permite su identificación. Debido a estas particularidades tanto el proceso de identificación de los patrones de control de flujo como del comportamiento que los caracteriza es diferente para cada uno de ellos. A continuación se describe el proceso de identificación del comportamiento y la realización de la búsqueda para cada uno de los patrones de control de flujo.

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+1} están en secuencia si todas las SD dirigidas hacia actividades de s_{i+1} son solo de actividades de s_i o en s_{i+1} 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 SD, 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 SD 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 SD cuya actividad predecesora está en el **Redo** y sucesora aparece en el **Do**, la actividad sucesora debe terminar al menos una traza.

La primera posible descomposición 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 SD 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. 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 posible descomposición de paralelismo contiene cada una de las actividades como parte de un conjunto en paralelo con el resto. El procesamiento de las SI, en las cuales se violan las condiciones planteadas, provoca la unión de los conjuntos en los que se encuentran las actividades presentes en la SI procesada. La ausencia de representación de los conjuntos 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 SI.

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 SI 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 posible descomposición contiene a cada una de las actividades como un subproceso. El procesamiento de las SI, en las cuales se violan las condiciones planteadas, provoca la unión de los conjuntos a los que pertenecen. Es posible descartar determinadas SI 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 cada actividad b que pertenece a B existe una SI desde a hasta b y una SI desde b hasta a . Además debe existir representación de solo uno de los subprocesos en al menos una traza.

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

2.3 Incorporación de la técnica al marco de trabajo ProM

La estructura de la técnica Minería de Variantes se concibió como un complemento para el marco de trabajo ProM con el objetivo de posibilitar su incorporación al mismo. Este marco de trabajo soporta diferentes tipos de complementos. El primer tipo de complementos son los que realizan el procesamiento necesario para la obtención de

determinados objetos, a partir de objetos de entrada. El segundo tipo de complementos se utiliza para realizar la representación de los objetos, lo cual posibilita su visualización por parte de los usuarios. El último tipo de complementos se utiliza para realizar los procesos de importación y exportación de recursos en el marco de trabajo.

La utilización del objeto resultante de la aplicación de la técnica, denominado Árbol de Variantes, requiere la realización de los tres tipos de complementos. El primer complemento para realizar el procesamiento necesario para obtener el Árbol de Variantes a partir de un registro de eventos. El segundo complemento debe permitir la construcción y presentación de una vista jerárquica para un objeto de tipo Árbol de Variantes. Por último, se requieren complementos que permitan los procesos de importación y exportación de los objetos de tipo Árbol de Variantes generados. En la Figura 9 se muestra la relación entre los complementos anteriormente mencionados.

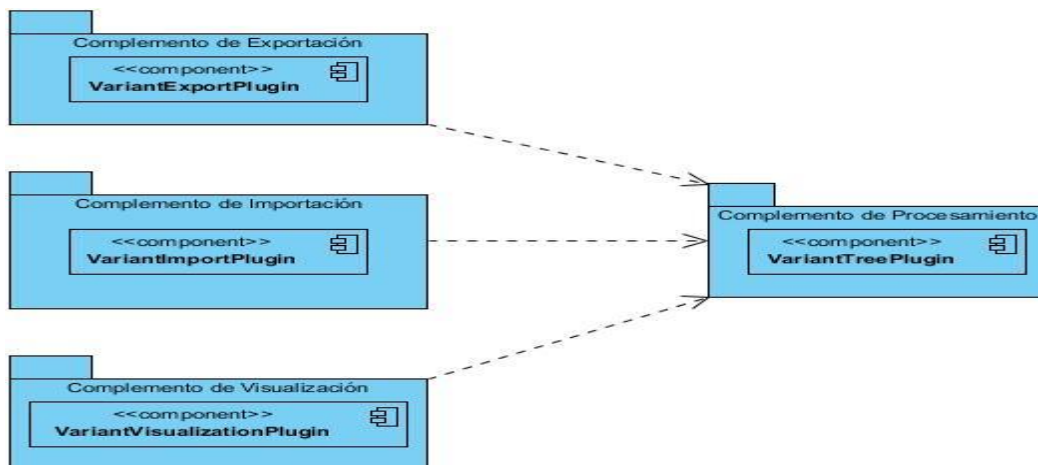


Figura 9: Relación existente entre los complementos.

El marco de trabajo ProM es capaz de identificar que el objeto de tipo Árbol de Variantes obtenido como resultado de la aplicación del complemento de procesamiento, se corresponde con el tipo de dato que recibe como entrada el complemento encargado de generar la visualización. Además identifica que dicho complemento es un complemento de visualización y automáticamente se muestra la vista generada por el mismo. En caso de existir varios complementos de visualización el ProM posibilita obtener la vista generada por cada uno de ellos y seleccionarlas alternativamente. De manera similar al importar o exportar un recurso es posible

seleccionar el complemento a emplear, en caso de existir varios para el mismo tipo de objeto. El funcionamiento del marco de trabajo ProM posibilita la integración de los complementos aunque no exista una relación directa entre los mismos. En la Figura 10 se muestra la estructura correspondiente al complemento que posibilita la obtención del Árbol de Variantes.

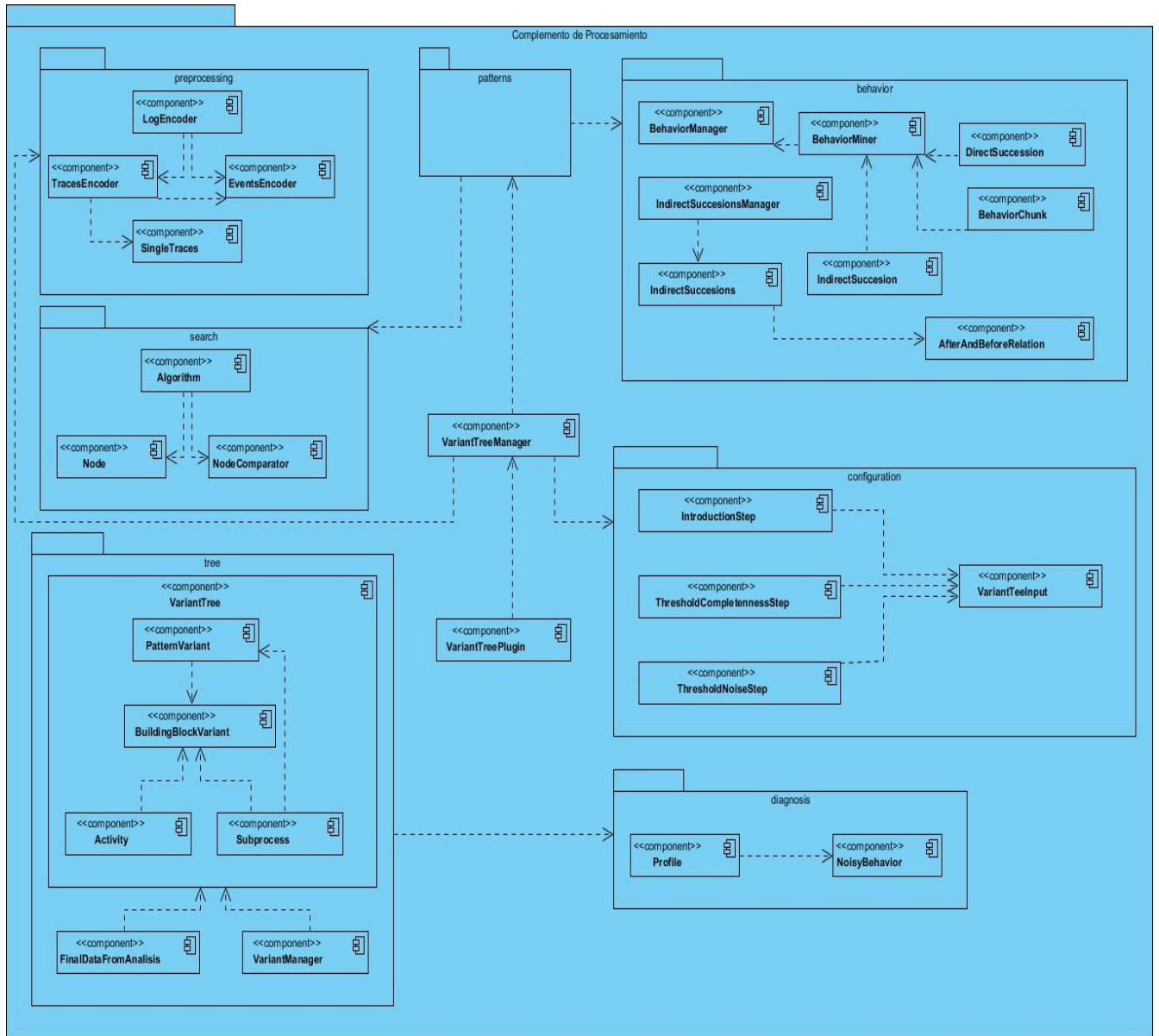


Figura 10: Estructura del complemento de procesamiento.

Este complemento está estructurado en 7 paquetes: **configuration**, **preprocessing**, **tree**, **diagnosis**, **behavior**, **search** y **patterns**. Cada uno de estos paquetes contiene los componentes necesarios para la realización de funciones específicas dentro del complemento.

En el paquete **configuration** se encuentran los componentes encargados de obtener y almacenar los parámetros de configuración introducidos por el usuario. Los datos obtenidos son empleados durante la identificación de los patrones de control de flujo. El paquete **preprocessing** contiene los componentes encargados de realizar la codificación de las actividades presentes en el registro de eventos, extraer los datos generales del registro de eventos y agrupar las trazas iguales. Los componentes pertenecientes a la estructura del Árbol de Variantes se encuentran ubicados en el paquete **tree**. En el paquete **diagnosis** se ubican los componentes asociados al perfil de diagnóstico. Los componentes presentes en el paquete **behavior** realizan la extracción del comportamiento relevante dentro del proceso. Dentro del paquete **search** se encuentran los componentes correspondientes al algoritmo de búsqueda empleado durante el descubrimiento de los patrones de control de flujo. Por último, en el paquete **patterns**, mostrado en la Figura 11, se encuentran los componentes empleados para realizar el descubrimiento de los patrones de control de flujo. Asociado al descubrimiento de cada uno de los patrones de control de flujo existe un componente especializado en la realización del procesamiento necesario.

Las funcionalidades de los complementos que permiten importar y exportar los objetos de tipo Árbol de Variantes están estrictamente dirigidas a permitir el almacenamiento o la carga de un objeto. Los mecanismos presentes en el marco de trabajo para realizar estas operaciones reducen la complejidad de la realización de dichos complementos.

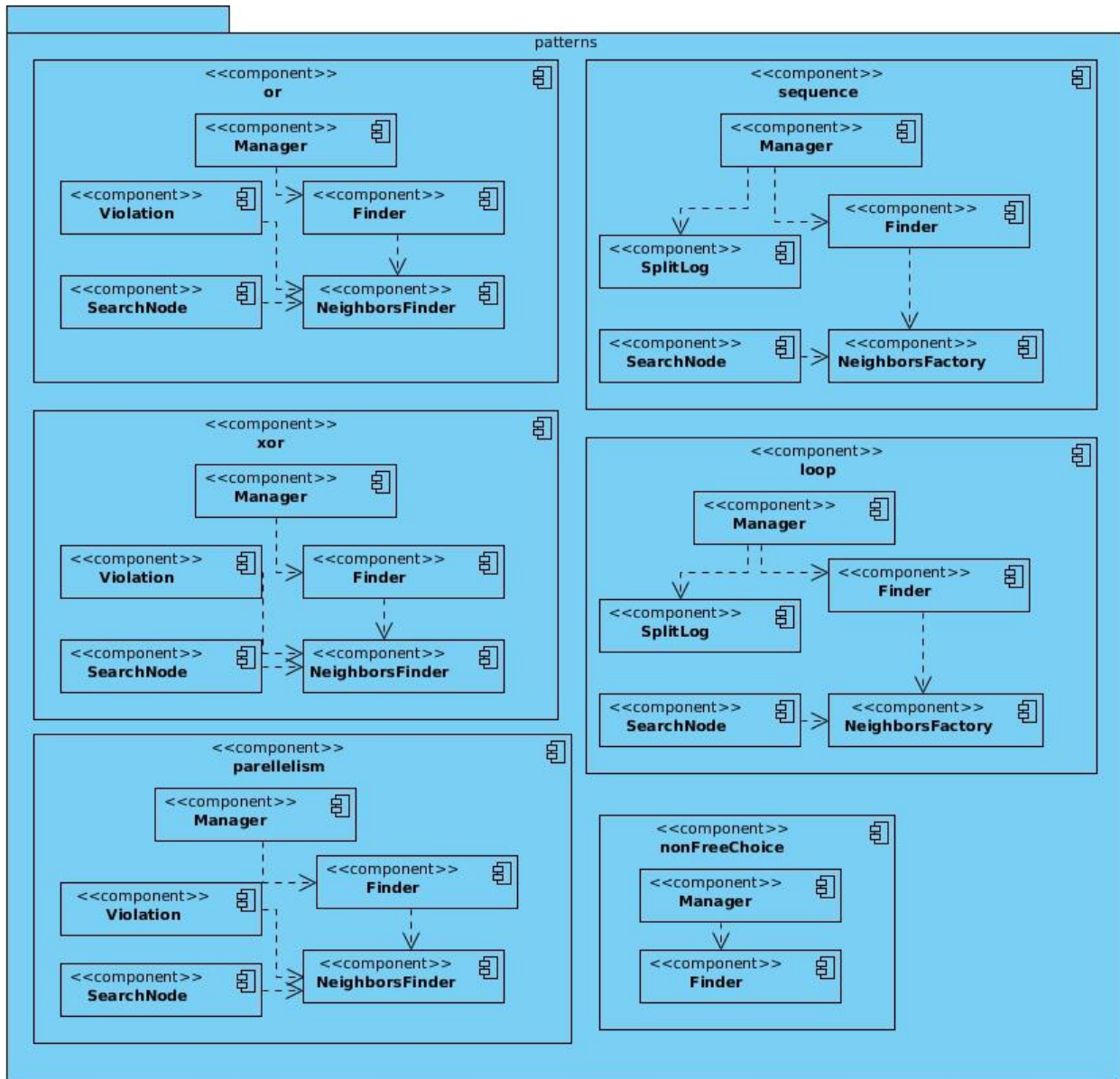


Figura 11: Componentes del paquete patterns

La Figura 12 muestra la estructura del complemento encargado de la construcción y presentación de la vista del Árbol de Variantes. Este complemento está estructurado en 4 paquetes: **toolbar**, **traces**, **tree** y **behaviorpanel**. Cada uno de estos paquetes contiene los componentes necesarios para la visualización de determinadas funcionalidades.

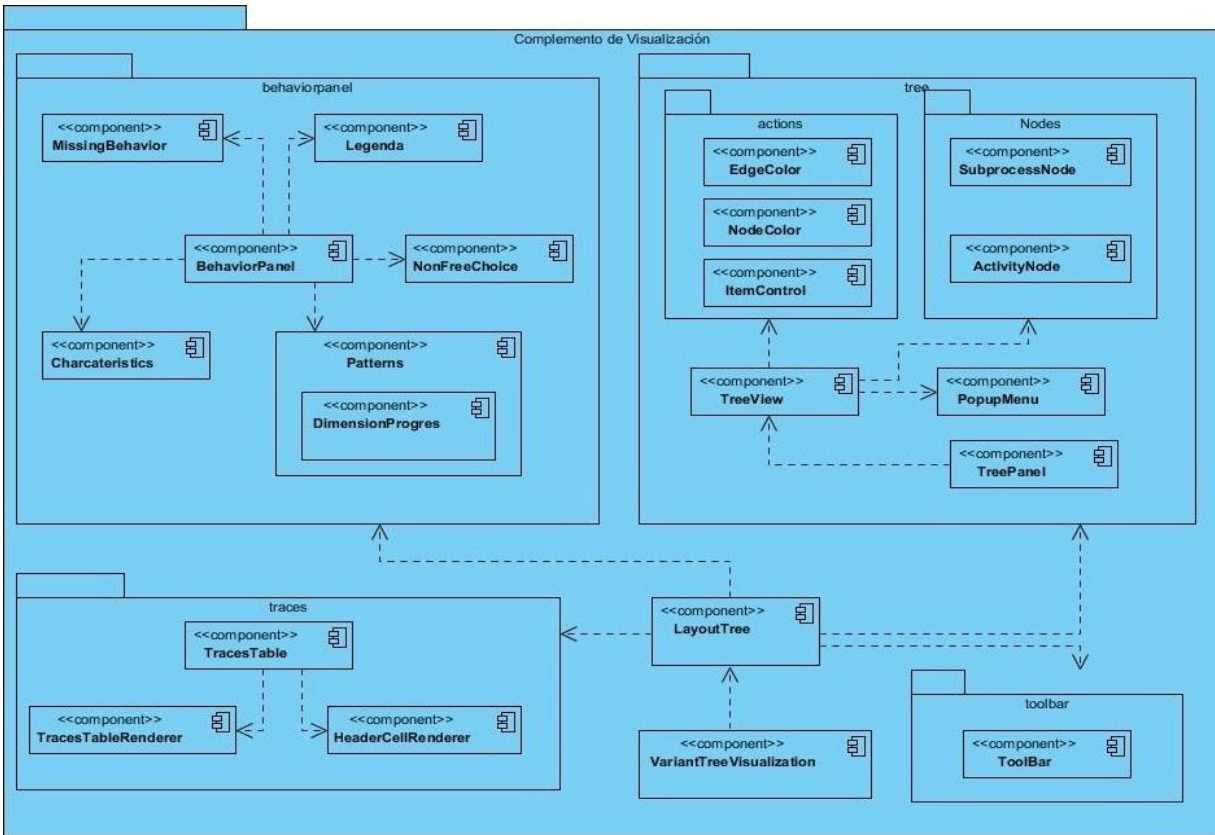


Figura 12: Estructura del complemento de visualización.

En el paquete **toolbar** se encuentran los componentes para seleccionar y posteriormente mostrar la variante que presente mejor evaluación de acuerdo a la cantidad de ruido o ausencia de información contenida en la misma. Los componentes presentes en el paquete **traces** se utilizan para mostrar la sección del registro de eventos asociada al subproceso seleccionado. Además facilitan la identificación de las trazas o el comportamiento identificado como ruido o ausencia de información en la sección del registro de eventos correspondiente al subproceso analizado. Dentro del paquete **behaviorpanel** se encuentran los componentes utilizados para la visualización de los datos del perfil de diagnóstico, la codificación de las actividades contenidas en el registro de eventos, el comportamiento ausente, las variantes identificadas así como la evaluación asociada a las mismas. Por último, en el paquete **tree** se encuentran los componentes que empleados para la construcción de la visualización jerárquica del Árbol de Variantes.

2.4 Consideraciones del capítulo

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.

El desarrollo de la técnica como un complemento del marco de trabajo ProM permite la utilización de sus resultados por parte de otros complementos del mismo. Conjuntamente con el complemento para la obtención del Árbol de Variantes fueron desarrollados complementos que permiten la visualización, exportación e importación de este tipo de objetos en el marco de trabajo ProM.

Capítulo 3: Utilización de la técnica Minería de Variantes para el diagnóstico de proceso

La técnica Minería de Variantes ha sido utilizada para el diagnóstico de proceso aplicándose a registros de eventos provenientes de entornos reales. En el presente capítulo se describen y analizan los resultados obtenidos sobre dos de estos registros de eventos. El primero de los registros analizados fue publicado como parte del Desafío de Inteligencia de Procesos de Negocio 2013 y contiene eventos generados por el sistema para la gestión de incidentes VINST el cual es utilizado por la empresa belga Volvo IT. El segundo registro de eventos analizado fue generado por el Sistema de Información Hospitalaria alas HIS desarrollado en la Universidad de las Ciencias Informáticas.

3.1 Aplicación a un registro de eventos del Desafío de Inteligencia de Procesos de Negocios 2013

3.1.1 Desafío de Inteligencia de Procesos de Negocios 2013

El Taller Internacional sobre Inteligencia de Procesos de Negocio es un evento anual organizado con la colaboración del Grupo de Trabajo para la Minería de Proceso de la IEEE¹¹. El Desafío de Inteligencia de Procesos de Negocio (BPIC¹²) se realiza como parte de este evento. Para la realización del BPIC se emplea un registro de eventos proveniente de un entorno real, sobre el cual los participantes aplican técnicas de minería de proceso. Los resultados de los análisis aplicados son presentados por medio de artículos, los cuales son evaluados por el jurado del evento.

En su edición del año 2013 se presentó un registro de eventos proveniente de la empresa belga Volvo IT (TECHNISCHE UNIVERSITEIT EINDHOVEN, 2013). El registro de eventos contiene eventos de un sistema manipulador de incidentes denominado VINST.

¹¹ Del inglés: Institute of Electrical and Electronics Engineers.

¹² Del inglés: Business Processing Intelligence Challenge.

3.1.2 Analizando el registro de eventos y el mapa de proceso

La técnica Minería de Variantes ha sido aplicada a una sección del registro de eventos similar a la utilizada en (SEPPE K.L.M, 2013). La sección del registro de eventos está compuesta por 3379 trazas y 6 eventos y agrupa las 5 ejecuciones más frecuentes del proceso.

El mapa de proceso generado a partir del registro de eventos anteriormente descrito se muestra en la Figura 13.

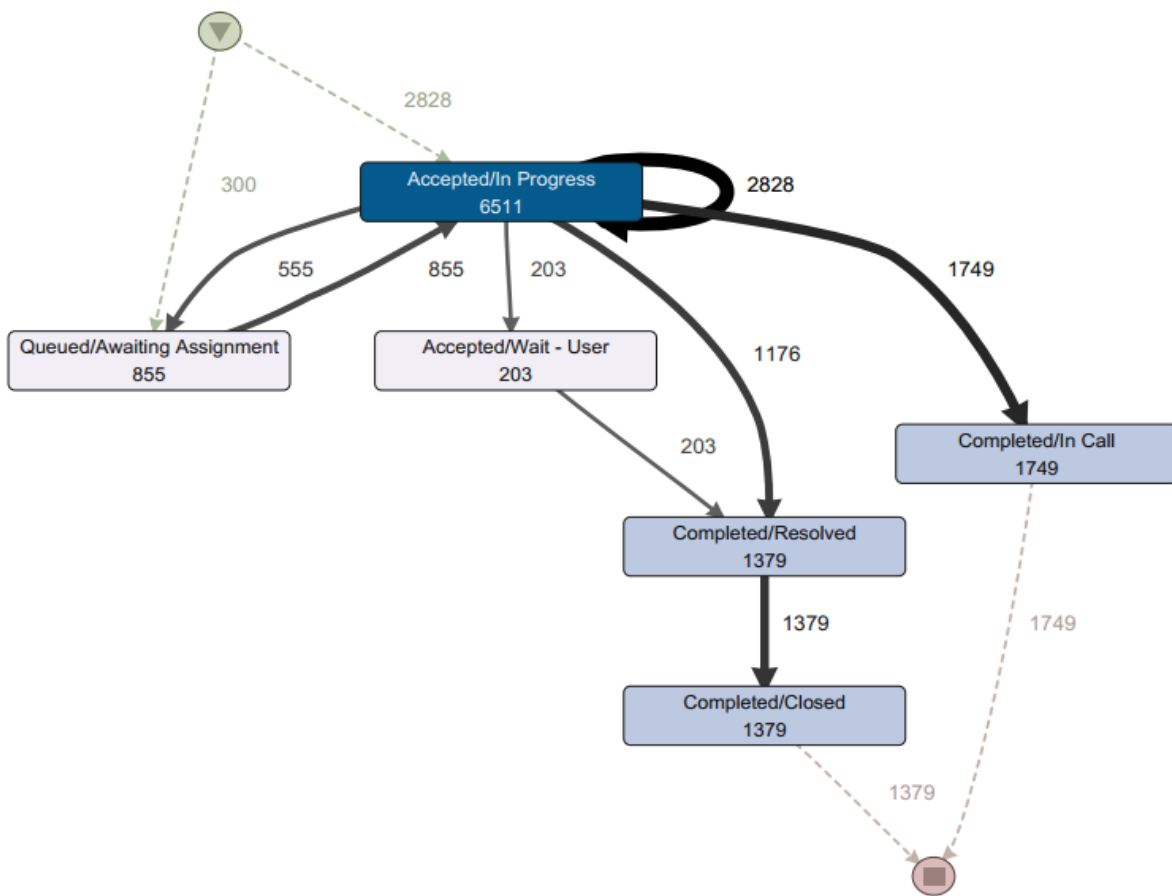


Figura 13: Mapa de proceso identificado en (SEPPE K.L.M, 2013)

3.1.3 Aplicación de la técnica Minería de Variantes

En el primer nivel de abstracción presente en el mapa de proceso se aprecia la existencia de una secuencia entre dos subprocessos. El primero está compuesto por las actividades *Accepted/In Progress* (codificada como **b**) y *Queued/Awaiting Assignment* (codificada como **a**). En el segundo subprocesso se encuentran las actividades *Accepted/Wait-User*, *Completed/Closed*, *Completed/Resolved* y *Completed//In Call* codificadas como **f**, **d**, **e** y **c** respectivamente. En el mapa de proceso se aprecia que luego de la ejecución del primer subprocesso ninguna de sus actividades puede aparecer nuevamente en la traza, evidenciando la presencia del patrón secuencia. La Figura 14 muestra el resultado de la aplicación de la técnica Minería de Variantes sobre la misma sección del registro de eventos analizada, donde se evidencia la existencia de la secuencia entre dos subprocessos.

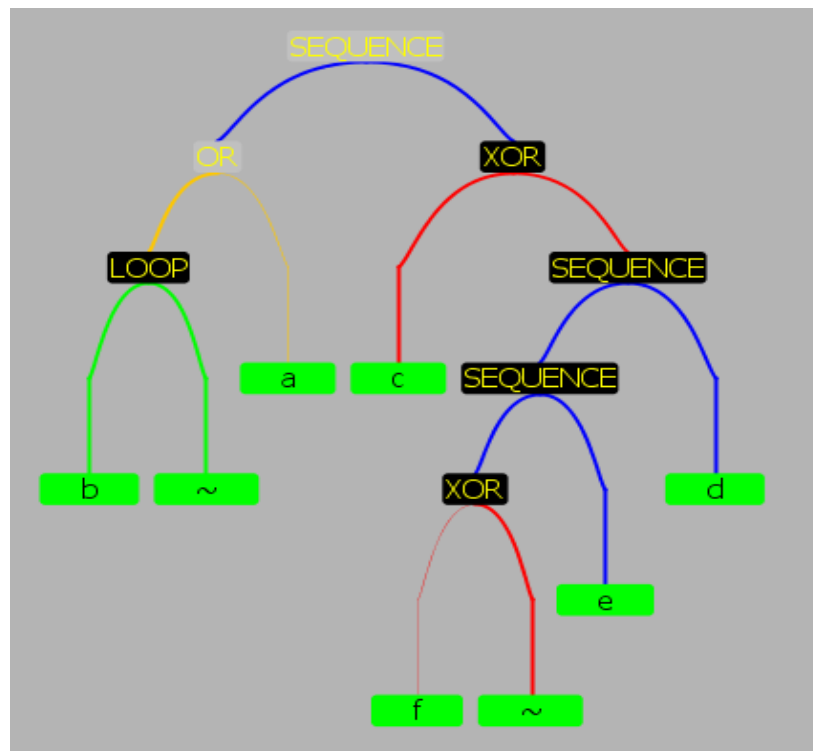


Figura 14: Resultado de la aplicación de la técnica Minería de Variantes sobre las 5 ejecuciones más frecuentes.

3.1.3.1 Análisis de los resultados para el primer subproceso identificado

En el primer subproceso identificado por la técnica Minería de Variantes, se puede apreciar la existencia de una selección no exclusiva entre las actividades *Accepted/In Progress* (**b**) y *Queued/Awaiting Assignment* (**a**). La relación existente entre las actividades indica que puede aparecer *Accepted/In Progress* repetidas veces o *Queued/Awaiting Assignment* y luego *Accepted/In Progress* o alguna repetición de las variantes anteriores. En la Figura 15 se muestran las variantes identificadas en el subproceso, sus evaluaciones de ruido y completitud, así como la sección del registro de eventos con los comportamientos relevantes destacados.



Figura 15: Variantes identificadas en el primer subproceso.

La primera variante corresponde al patrón Selección no Exclusiva. Para la identificación mediante este patrón no fue necesario la asunción de ningún comportamiento por concepto de completitud ni descartar ningún comportamiento por concepto de ruido, por lo que sus evaluaciones son 1,0 y 0,0 respectivamente. Se identifica además una variante de descomposición por el patrón de control de flujo Secuencia considerando como ruido la sucesión directa entre **b** y **a** presente en la segunda traza. La evaluación de ruido en este caso es 0,128 y la traza queda señalada en el registro de eventos. La tercera traza es la única que no presenta actividades repetidas, si se considera dicha traza como incompleta se identifica una variante de descomposición utilizando el patrón

Lazo. La actividad **b** faltaría al inicio de dicha traza al ser parte del subproceso **Do** del patrón Lazo por lo que la evaluación de completitud para esta variante es 0,91.

3.1.3.2 Análisis de los resultados para el segundo subproceso identificado

En la Figura 16 se muestra la descomposición obtenida del segundo subproceso mediante la aplicación de la técnica Minería de Variantes.

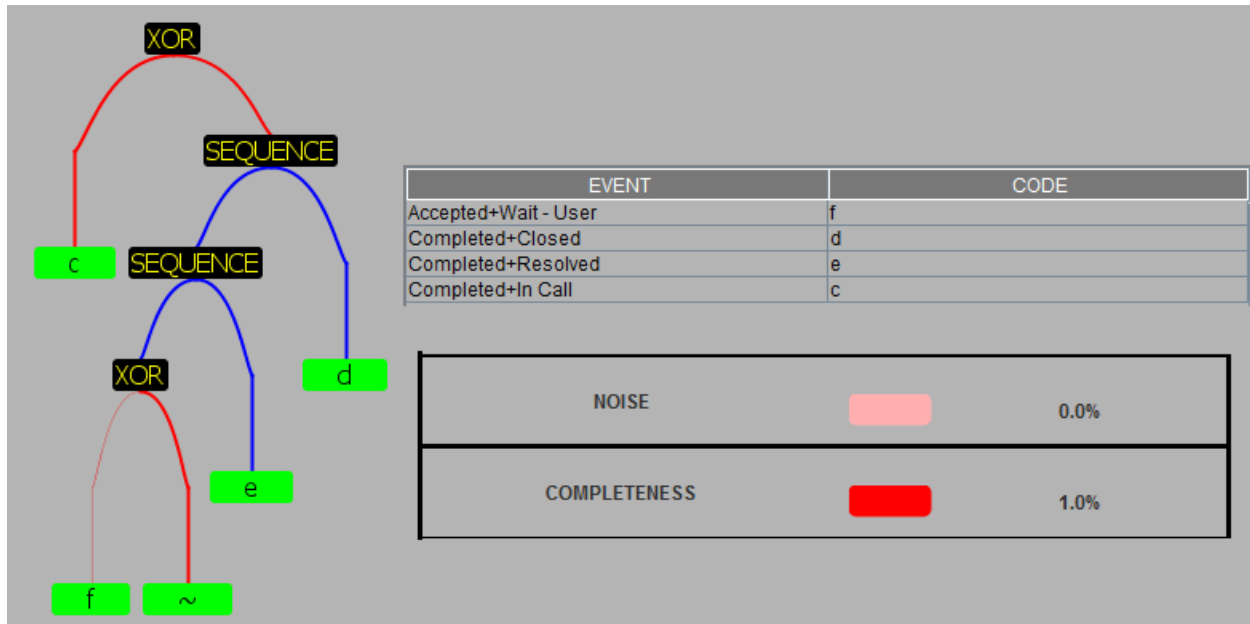


Figura 16: Descomposición del segundo subproceso.

La actividad *Completed/In Call* (**c**) aparece como parte de una selección exclusiva con un subproceso compuesto por las restantes actividades. Este comportamiento resulta consecuente con el mapa de proceso mostrado en la Figura 13 y con el comportamiento expresado en la sección del registro de eventos analizada, la cual se muestra en la Figura 17. En los niveles de abstracción inferiores se puede apreciar la identificación de la presencia de una *actividad invisible* lo cual es consecuencia de la ausencia de la actividad *Accepted/Wait-User* (**f**) en la segunda traza. Para la descomposición de los patrones de control de flujo identificados no fue necesario asumir o ignorar ningún comportamiento, por lo que las evaluaciones obtenidas son 0,0 para ruido y 1,0 para completitud.

1749	c		
1398	e	d	
232	f	e	d

Figura 17: Sección del registro de eventos correspondiente al segundo subproceso.

La descomposición obtenida mediante la técnica Minería de Variantes es consecuente con la información presente en el registro de eventos y con el mapa de proceso descrito en (SEPPE K.L.M, 2013). La posibilidad de asumir la presencia de comportamiento por concepto de completitud o ignorar comportamiento presente en el registro de eventos al considerarlo ruido permite la identificación de variantes de modelos de proceso en la descomposición en subprocesos.

3.2 Aplicación a un registro de eventos proveniente de un entorno real

3.2.1 Entendiendo el entorno

El Sistema de Información Hospitalaria alas 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. El mismo 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 (SAÑUDO ORTIZ, 2012).

La técnica Minería de Variantes se ha aplicado a un registro de eventos que corresponde al módulo Almacén del sistema alas HIS con el objetivo de comprobar su utilidad en el diagnóstico del proceso. 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, presentado en la Figura 18, permite identificar la existencia de múltiples actividades de fin del proceso y la existencia de lazos en diferentes puntos del proceso. Por otra parte es relevante la presencia de dos tipos de eventos, los eventos de tipo **Decision** y los eventos de tipo **Task Node**. La presencia de eventos del tipo

Decision sugiere la existencia de patrones de control de flujo que involucren la elección de alternativas.

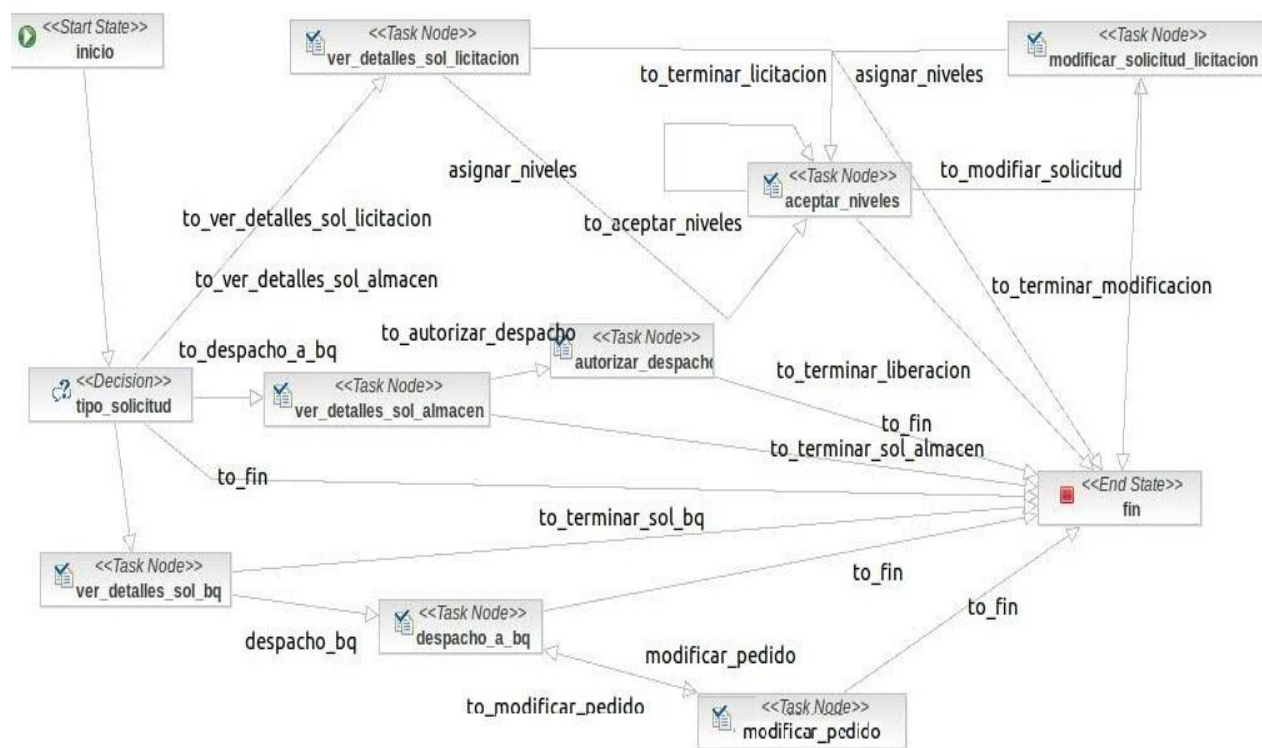


Figura 18: Modelo del proceso Solicitar Producto.

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.

3.2.2 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. Para corroborar este resultado es necesario realizar un análisis más profundo del modelo del proceso, mediante el que se evidencia que antes de la ejecución de alguno de los eventos de tipo **Task Node** existe

un evento de tipo **Decision**, denominado *tipo_solicitud*. En la Figura 19 se muestra el Árbol de Variantes identificado.

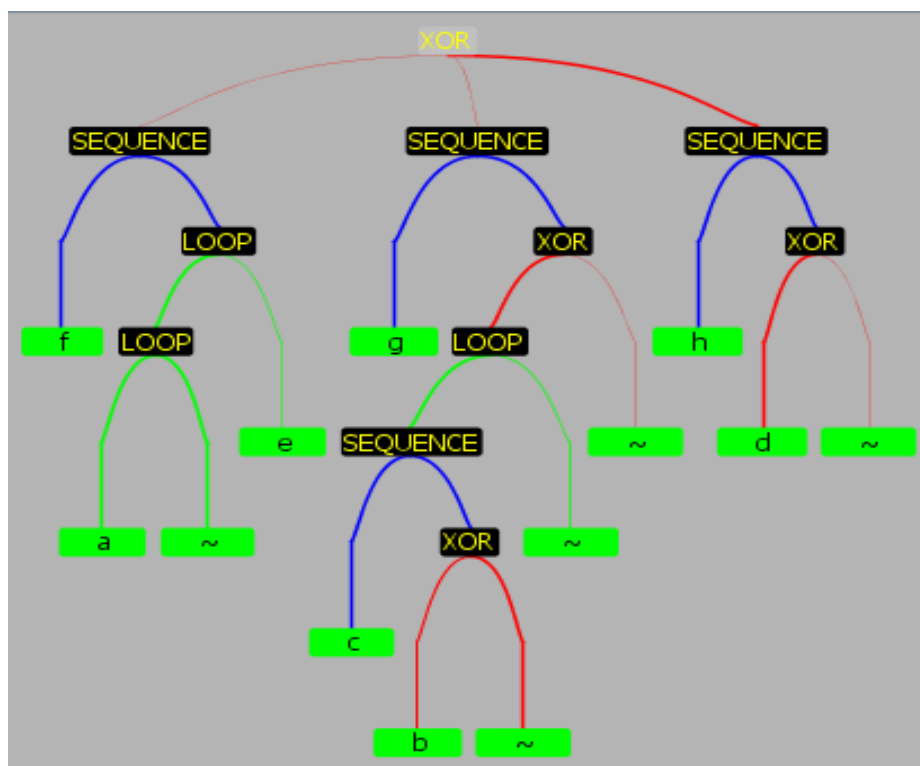


Figura 19: Variante del proceso mediante la descomposición por Selección Exclusiva.

A partir de la ejecución del evento de tipo **Decision** surgen 4 posibles alternativas. La tercera alternativa vincula el evento *tipo_solicitud* directamente con el fin del proceso. La selección de dicha alternativa no produciría evidencias en el registro de eventos por lo que el resultado obtenido en este nivel de descomposición es coherente con el modelo del proceso. Las restantes alternativas involucradas corresponden a los subprocesos **ver_detalle_sol_licitacion**, **ver_detalle_sol_almacen** y **ver_detalle_sol_bq**. Otro aspecto a tener en cuenta es el modo en que el comportamiento de esta variante se encuentra expresado en el registro de eventos, o sea, el comportamiento en el registro de eventos que ha sido necesario asumir o el que ha sido considerado como ruido para determinar esta variante. En la Figura 20 se muestran las evaluaciones de ruido y completitud identificadas.

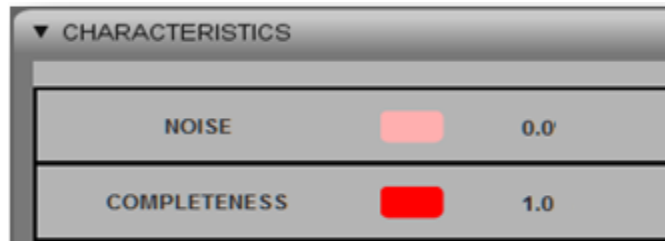


Figura 20: Evaluaciones del comportamiento presente en el registro de eventos.

La evaluación de 0,0 de ruido significa que para determinar esta variante no fue necesaria la consideración de comportamiento como ruido, mientras que 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.

Una de las potencialidades de la técnica Minería de Variantes es que 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. Seguidamente se realiza el análisis de los subprocesos identificados correspondientes al segundo nivel de abstracción.

3.2.2.1 Análisis del subproceso ver_detalles_sol_licitacion

Dentro de las características generales del subproceso que se manifiestan en el análisis del modelo se encuentra la existencia de varias actividades que finalizan el proceso y la presencia de múltiples lazos. La Figura 21 muestra la existencia de una relación entre las actividades *ver_detalles_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.

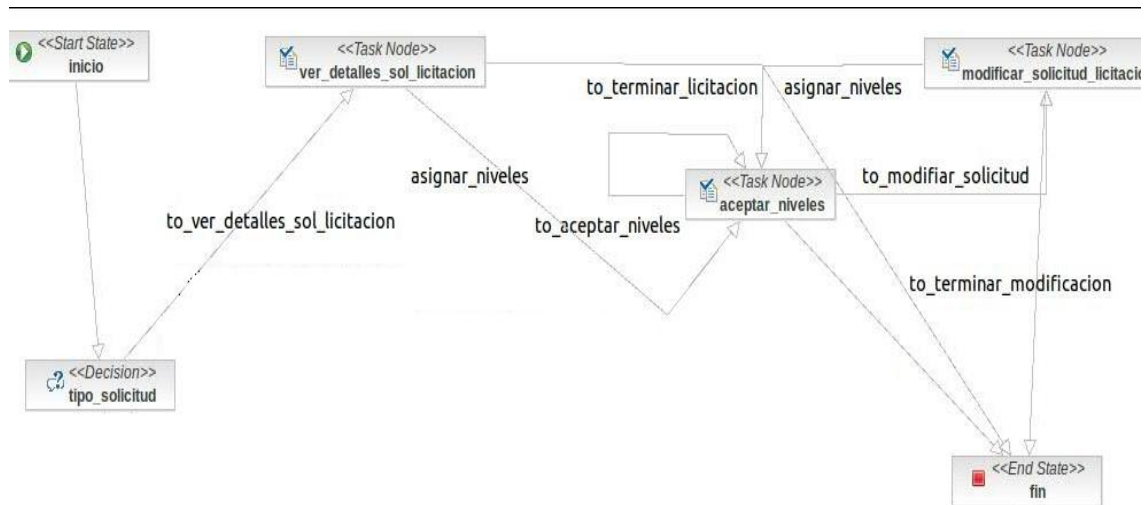


Figura 21: Modelo de proceso del subprocesso *ver_detalles_sol_licitacion*.

En la Figura 22 se muestra el resultado obtenido por la aplicación de la Minería de Variantes a la sección del registro de eventos asociada a este subprocesso, las trazas y la codificación de las actividades presentes en dicha sección.

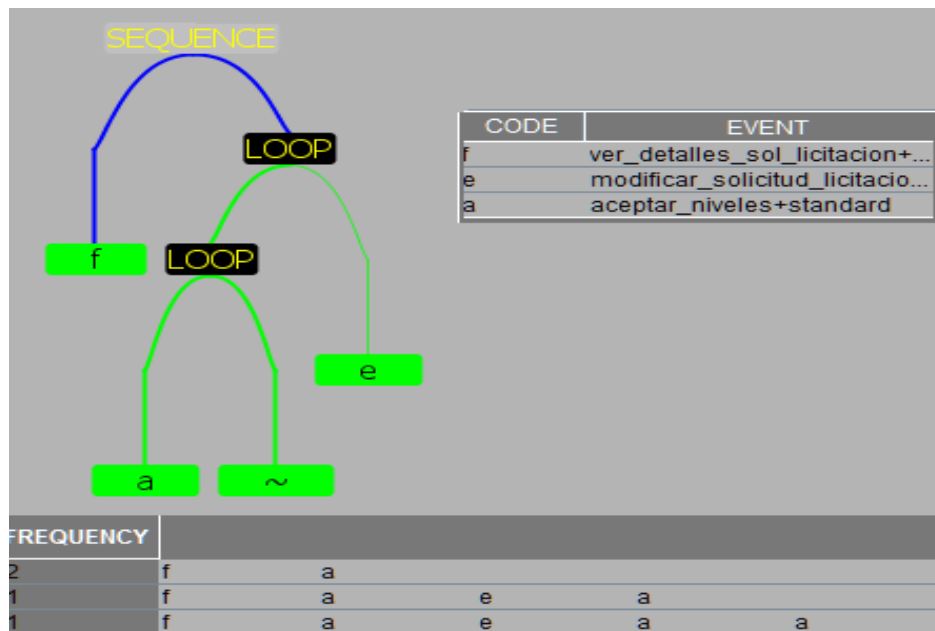


Figura 22: Subprocesso *ver_detalles_sol_licitacion*.

En la sección del Árbol de Variantes correspondiente a este subprocesso se identifica la secuencia entre la actividad *ver_detalles_sol_licitacion* y un subprocesso compuesto por las actividades *modificar_solicitud_licitacion* y *aceptar_niveles*. La aparición de esta

relación indica que después de la ocurrencia de *ver_detalle_sol_licitacion* debe existir alguna actividad, por lo que esta actividad nunca debe aparecer como evento de fin de traza.

En el subproceso restante se determinó la existencia de un lazo donde la actividad *ver_detalle_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 como el evento de fin de traza. Además se evidencia que en dos trazas 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 la totalidad del comportamiento expresado en el modelo del proceso mostrado en la Figura 18, para el subproceso **ver_detalle_sol_licitacion**, no se encuentra recogido en el registro de eventos.

3.2.2.2 Análisis del subproceso *ver_detalle_sol_almacen*

En la Figura 23 se muestra la sección del modelo del proceso correspondiente al subproceso *ver_detalle_sol_almacen*.

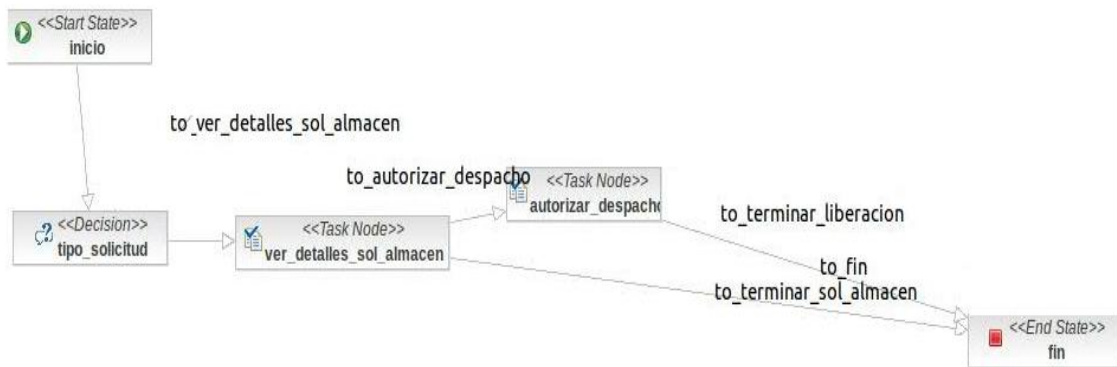


Figura 23: Modelo de proceso del subproceso ver_detalles_sol_almacen.

La relación de las actividades *ver_detalles_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 24.

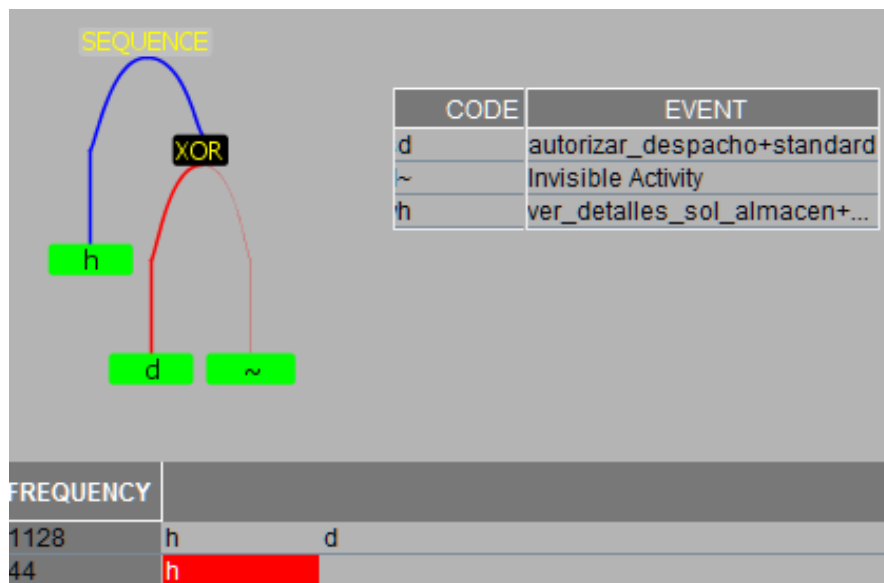


Figura 24: Subproceso ver_detalles_sol_almacen.

En esta sección del registro de eventos se identifica la existencia de la secuencia entre la actividad *ver_detalle_sol_almacen* y un subproceso descompuesto mediante el patrón de control de flujo Selección Exclusiva. Es necesario destacar la señalización de la traza que contiene solamente a la actividad *autorizar_despacho* como una traza que presenta ausencia de información según la descomposición atendiendo al patrón de control de flujo secuencia.

La ausencia de representación de uno de los subprocesos 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 ausencia de información en la segunda traza. La identificación de la ausencia de información en esa traza provoca la inserción en la misma 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 (YZQUIERDO-HERRERA R., 2012). La inserción de la actividad invisible en la traza posibilita la posterior identificación del patrón de Selección Exclusiva.

3.2.2.3 Análisis del subproceso ver_detalle_sol_bq

En la Figura 25 se muestra la parte del modelo del proceso correspondiente al subproceso *ver_detalle_sol_bq*.

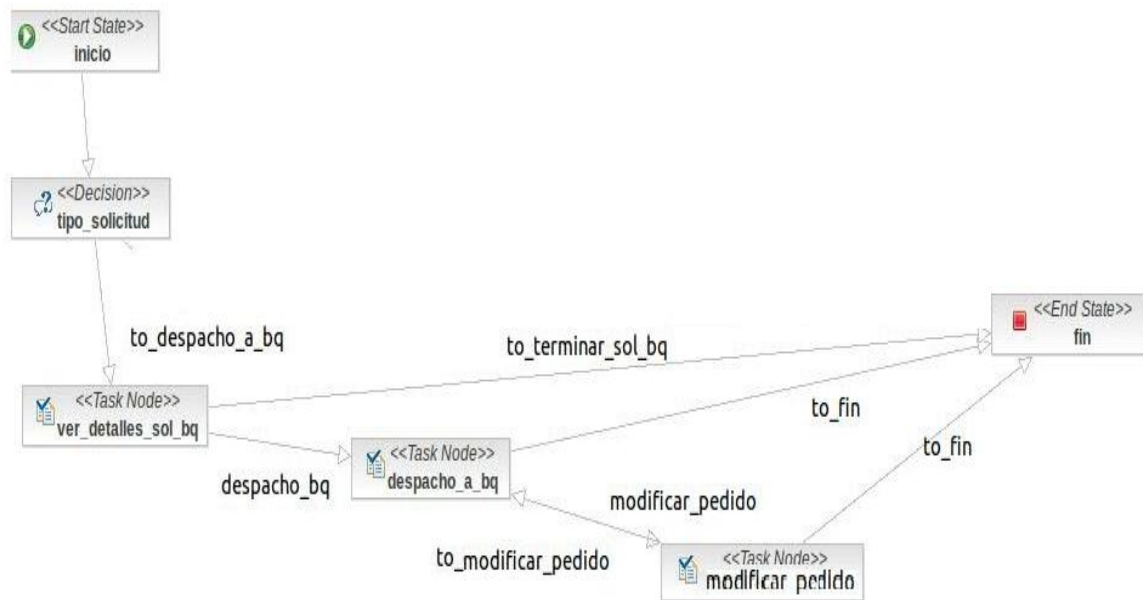


Figura 25: Modelo de proceso del subproceso ver_detalle_sol_bq.

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_detalle_sol_bq*, *despacho_a_bq* y *modificar_pedido*. Por otra parte es posible identificar la existencia de un lazo entre las actividades *despacho_a_bq* y *modificar_pedido*. Además se evidencia la existencia de una secuencia entre la actividad *ver_detalle_sol_bq* y un subproceso compuesto por el resto de las actividades. La descomposición obtenida para este subproceso, las trazas y la codificación de sus actividades se muestran en la Figura 26.

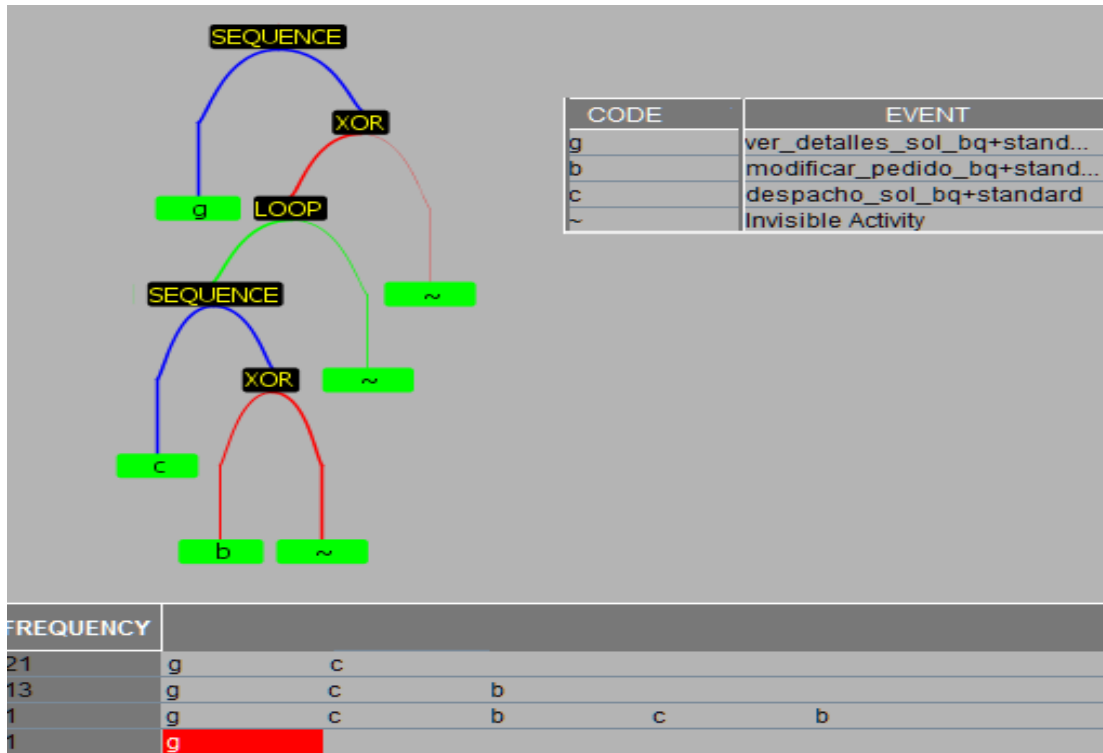


Figura 26: Subproceso ver_detalles_sol_bq.

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.

En esta variante no se identifica un lazo entre las actividades *despacho_sol_bq* y *modificar_pedido_bq*, sin embargo se identifica un lazo en el que 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 pueden aparecer como actividades de fin de proceso tanto la actividad *despacho_sol_bq* como la actividad *modificar_pedido_bq*. Además al aparecer como parte del **Do** del lazo puede repetirse el subproceso completo lo que permite la ocurrencia de repeticiones de las dos actividades. 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 aunque existen diferencias en la representación la descomposición en subprocesos obtenida es coherente con el modelo del proceso.

3.2.3 Análisis de la variante de Secuencia

El análisis de esta variante evidencia la existencia de dos subprocesos en secuencia. El primero de los subprocesos se encuentra conformado por las actividades *ver_detalles_sol_bq*, *ver_detalles_sol_licitacion* y *ver_detalles_sol_almacen*¹³. Mientras que las actividades *aceptar_niveles*, *modificar_solicitud_licitacion*, *despacho_sol_bq*, *autorizar_despacho* y *modificar_pedido_bq* conforman el segundo subproceso. La descomposición resultante se muestra en la Figura 27.

¹³ La codificación empleada para las actividades es igual a la empleada durante el análisis de la variante descompuesta por el patrón de control de flujo Secuencia.

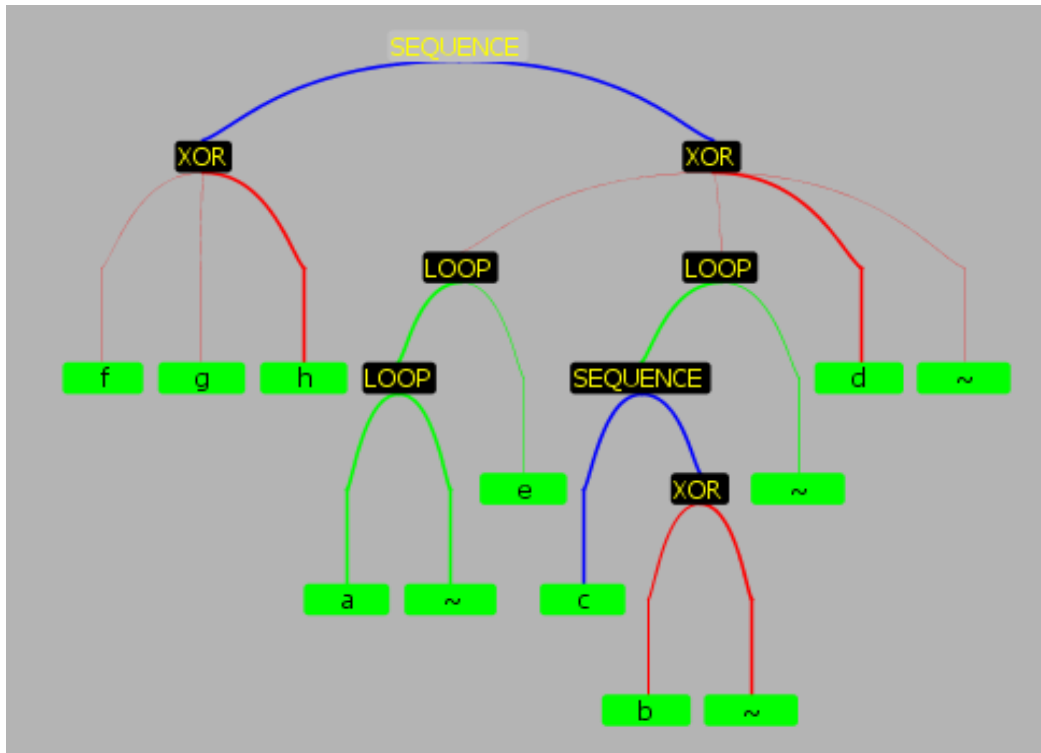


Figura 27: Variante del proceso mediante la descomposición por Secuencia.

Las evaluaciones obtenidas de ruido y completitud para esta variante muestran que para su identificación se asumió 0,0372 de ausencia de información en el registro de eventos. La existencia de una secuencia entre dichos subprocesos implica que en todas las trazas se encuentre una representación de ambos. Sin embargo, existen dos trazas en las que solamente aparece una actividad, en la primera traza la actividad *ver_detalle_sol_almacen* y en la segunda la actividad *ver_detalle_sol_licitacion*. La asunción de estas trazas como incompletas provoca la afectación de la evaluación de completitud para este patrón de control de flujo. En la Figura 28 se muestran los valores de ruido y completitud para esta variante.

▼ CHARACTERISTICS		
NOISE	<input type="checkbox"/>	0.0
COMPLETENESS	<input checked="" type="checkbox"/>	0.9628

Figura 28: Evaluación del comportamiento presente en el registro de eventos.

La identificación de trazas incompletas provoca la introducción de una *actividad invisible* en las trazas mencionadas representando la información ausente. En la Figura 29 se muestra el primer subproceso que conforma la secuencia obtenida y la sección del registro de eventos correspondiente.

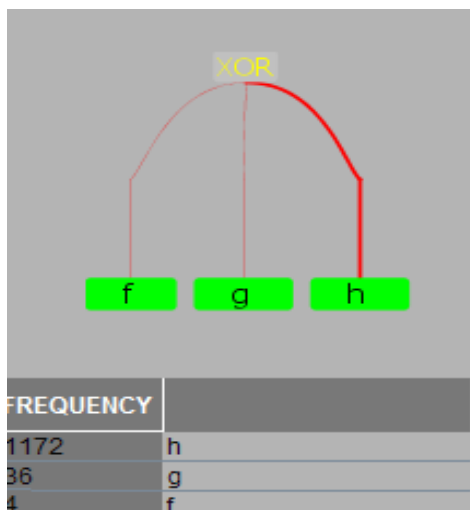


Figura 29: Primer subproceso en secuencia.

En este subproceso es posible identificar el patrón de control de flujo Selección Exclusiva entre las actividades que lo componen. En las trazas se evidencia la aparición solamente de una de las actividades que componen el subproceso, comportamiento característico del patrón identificado. Las evaluaciones de ruido y completitud obtenidas se encuentran en su valor máximo, o sea no se asumió o descartó comportamiento para la identificación del patrón en cuestión. La identificación de este patrón de control de flujo en este subproceso indica que todas las trazas inician con algunas de las actividades que lo conforman.

En la Figura 18 se aprecia que antes de la ejecución de algunos de los eventos de tipo **Task Node** se ejecuta un evento de tipo **Decision**. Las actividades involucradas en la decisión son *ver_detalle_sol_bq*, *ver_detalle_sol_licitacion* y *ver_detalle_sol_almacen*, de modo que el comportamiento identificado es coherente con el comportamiento expresado en el modelo de proceso. Por otra parte, la Figura 30 muestra la descomposición obtenida para el segundo subproceso y la sección del registro de eventos asociada.

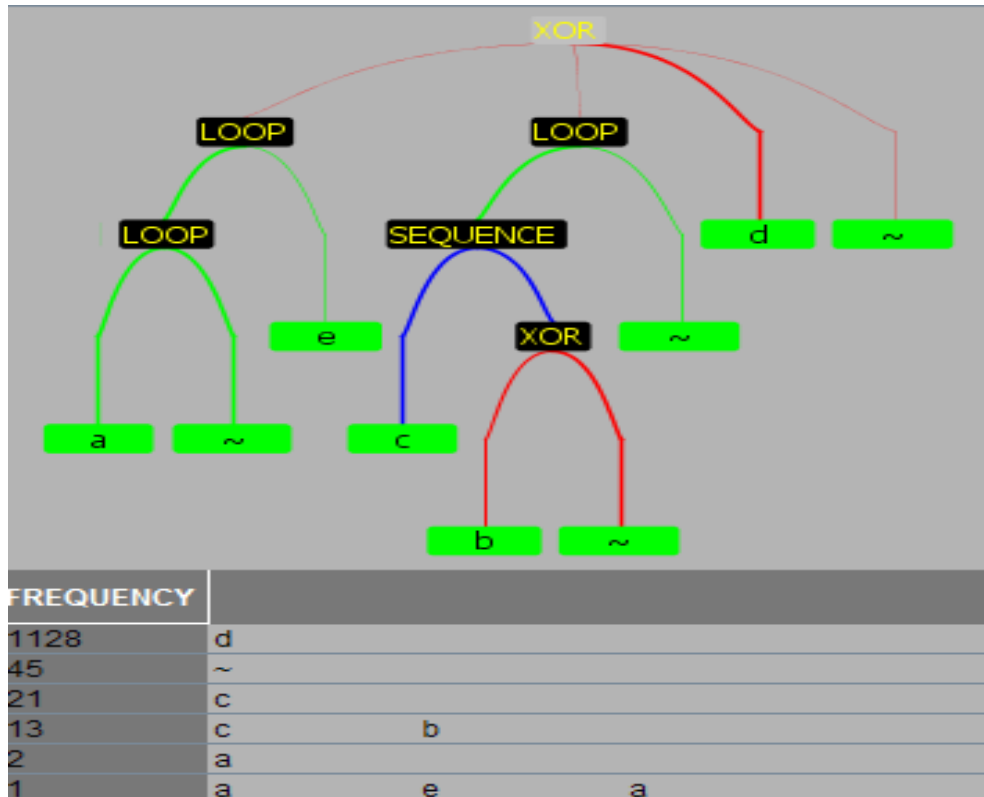


Figura 30: Segundo subproceso en secuencia.

El análisis de las trazas correspondientes al subproceso evidencia la presencia del comportamiento característico del patrón de control de flujo Selección Exclusiva. Las actividades que componen los subprocesos que han sido identificados como alternativas de la Selección Exclusiva nunca aparecen en la misma traza. La identificación de este patrón de control de flujo en este subproceso indica que todas las trazas terminan en alguno de los subprocesos involucrados en la selección. Este comportamiento es coherente con el modelo del proceso mostrado en la Figura 18, pues todas las ejecuciones del proceso deben concluir en algunas de las actividades presentes en este subproceso.

Esta variante del proceso garantiza que todas las trazas inicien y culminen con las actividades indicadas en el modelo de proceso. No obstante, la identificación de una secuencia entre dos subprocesos que a su vez se descomponen mediante Selección Exclusiva, permite la obtención de trazas que violan el comportamiento expresado en el modelo del proceso. Por ejemplo, atendiendo a la descomposición obtenida se pueden

obtener trazas en las que aparezcan las actividades *ver_detalle_sol_bq* y *aceptar_niveles*. Sin embargo, de acuerdo al modelo del proceso estas actividades nunca deberían formar parte de la misma ejecución del proceso.

La descomposición mediante el patrón de control de flujo Secuencia se realizó mediante la asunción de 0,0372 de ausencia de información. Los resultados obtenidos mediante esta variante son coherentes con el modelo de proceso y con el comportamiento presente en el registro de eventos. Sin embargo, esta variante de descomposición permite la obtención de trazas que no son admisibles por el modelo de proceso o sea, no presenta un nivel de aptitud adecuado.

3.2.4 Análisis de los resultados de la aplicación de técnicas de diagnóstico de proceso.

Los principales aspectos que se evidencian a partir de los resultados de la técnica propuesta son los patrones de control de flujo presentes en el registro de eventos y las relaciones existentes entre las actividades en cada uno de los subprocesos. Con el objetivo de comprobar la veracidad de los resultados de la aplicación de la Minería de Variantes es necesario analizar los resultados obtenidos mediante la aplicación de otras técnicas de diagnóstico de proceso al mismo registro de eventos. Las técnicas existentes que identifican con mayor claridad los aspectos determinados por la Minería de Variantes son la Alineación de Trazas y la Descomposición en Subprocesos utilizando Bloques de Construcción.

3.2.4.1 Análisis de los resultados de la técnica Alineación de Trazas

En la Figura 31 se muestra el resultado de la aplicación de la técnica Alineación de Trazas al registro de eventos proveniente del sistema alas HIS.

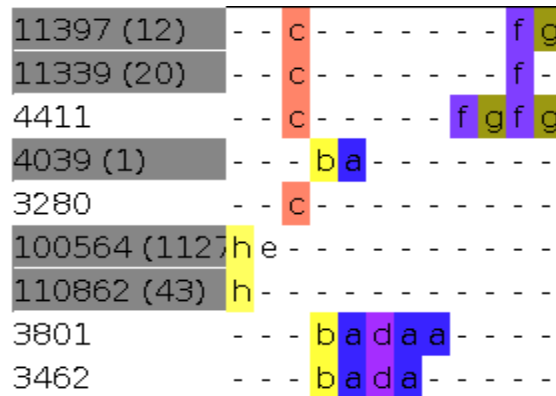


Figura 31: Alineación de Trazas.

El análisis del resultado obtenido de la aplicación de la Alineación de Trazas muestra, de modo similar a lo obtenido mediante la Minería de Variantes, la existencia del patrón de control de flujo Selección Exclusiva entre tres subprocesos. La codificación de las actividades empleada para la aplicación de la técnica se muestra en la Figura 32.

Activity Name	Char Encoding
despacho_sol_bq-standard	f
modificar_pedido_bq-standard	g
modificar_solicitud_licitacion-standard	d
autorizar_despacho-standard	e
ver_detalle_sol_licitacion-standard	b
ver_detalle_sol_bq-standard	c
aceptar_niveles-standard	a
ver_detalle_sol_almacen-standard	h

Figura 32: Codificación empleada en la Alineación de Trazas.

El primero de los subprocesos está compuesto por las actividades *ver_detalle_sol_bq*, *despacho_sol_bq* y *modificar_pedido_bq*. En este subproceso se evidencia la existencia de una secuencia entre la actividad *ver_detalle_sol_bq* y un subproceso conformado por las restantes. La única traza que no evidencia este comportamiento en el subproceso es la traza donde solo aparece la actividad *ver_detalle_sol_bq* por lo que dicha traza se podría considerar comportamiento anómalo. Además en este subproceso es posible identificar un lazo entre las actividades *despacho_sol_bq* y *modificar_pedido_bq*.

El análisis del segundo subproceso muestra que aproximadamente en el 96% de las trazas correspondientes a este subproceso se ejecutan las actividades *ver_detalles_sol_almacen* y *autorizar_despacho*. Mientras que en el 4% de las restantes ejecuciones solo aparece la actividad *ver_detalles_sol_almacen*.

El tercer subproceso está compuesto por las actividades *ver_detalles_sol_licitacion*, *aceptar_niveles* y *modificar_solicitud_licitacion*. El análisis de esta sección de la alineación muestra la existencia de una secuencia entre la actividad *ver_detalles_sol_licitacion* y el subproceso formado por las restantes. Además es posible identificar un lazo entre las actividades *aceptar_niveles* y *modificar_solicitud_licitacion* que siempre culmina con la actividad *aceptar_niveles*.

De modo general los resultados obtenidos mediante la aplicación de la técnica Alineación de Trazas en cuanto a la identificación de los patrones de control de flujo presentes reflejan el mismo resultado que los obtenidos mediante la Minería de Variantes.

3.2.4.1 Análisis de los resultados de la técnica Descomposición en Subprocesos utilizando Bloques de Construcción

En la Figura 33 se muestra el resultado de la aplicación de la técnica Descomposición en Subprocesos al registro de eventos proveniente del sistema alas HIS.

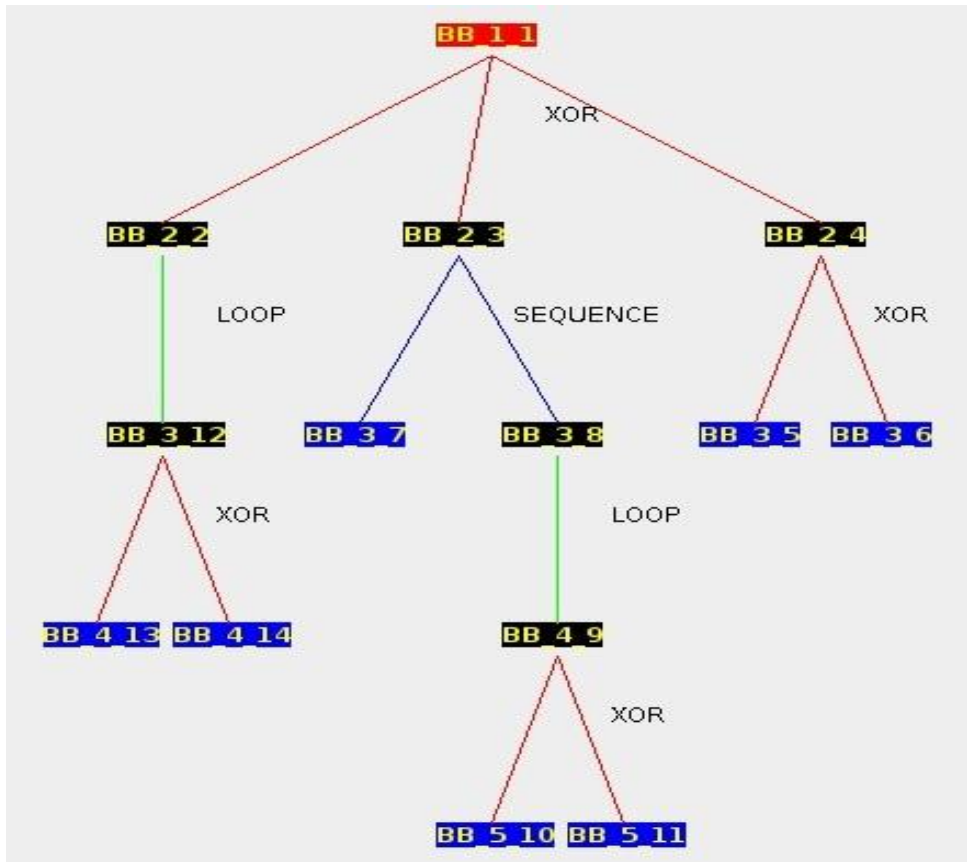


Figura 33: Descomposición en Subprocesos.

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_detalle_sol_bq**, **ver_detalle_sol_licitacion** y **ver_detalle_sol_almacen**. Este resultado es coherente con los obtenidos en la Alineación de trazas y en la Minería de Variantes.

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 el caso del Lazo la Descomposición en Subprocesos 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 subprocesos no identifica la secuencia entre la actividad *despacho_sol_bq* y el subproceso constituido por las actividades *modificar_pedido_bq* y una actividad

invisible debido a que este patrón se identifica a partir de la asunción de ausencia de información.

En el subproceso **ver_detalle_sol_licitacion** existe una diferencia con respecto a la identificación del patrón de control correspondiente al último nivel. En este caso mediante la Minería de Variantes se identifica un lazo y la técnica Descomposición en Subprocesos 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 ambos patrones de control de flujo 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 procesos sin admitir comportamiento adicional.

De manera similar a los anteriores, el subproceso **ver_detalle_sol_almacen** presenta diferencias con el resultado obtenido mediante la aplicación de la técnica propuesta. En el resultado de la aplicación de la técnica Descomposición en subprocesos no se identifica la existencia de una secuencia entre la actividad *ver_detalle_sol_almacen* y un subproceso descompuesto mediante el patrón de control de flujo Selección Exclusiva. El análisis de este resultado en la técnica Minería de Variantes mostró que la identificación de este patrón se realizó mediante la asunción de 3,76% de ausencia de información. La técnica Descomposición en Subprocesos no realiza un adecuado manejo de la ausencia de información por lo que no identifica la existencia de este patrón de control de flujo.

De modo general los resultados obtenidos mediante la aplicación de la técnica Descomposición en subprocesos son coherentes con los identificados mediante la Minería de Variantes. Sin embargo, los patrones descubiertos mediante la asunción de ausencia de información no son identificados por la técnica Descomposición en Subprocesos.

3.3 Consideraciones del capítulo

Durante este capítulo se analizaron los resultados de la aplicación de la técnica Minería de Variantes a dos registros de eventos provenientes de entornos reales. Los sistemas que registraron los eventos fueron el sistema de gestión de incidentes VINST y el Sistema de Gestión Hospitalaria alas HIS.

En el primero de ellos se tomó como referencia el mapa de proceso descrito en (SEPPE K.L.M, 2013). Mediante la aplicación de la Minería de Variantes se obtiene un modelo consecuente con el mapa de proceso. Se identifican los patrones de control de flujo presentes en el registro de eventos y estos se contextualizan a partir de la descomposición en subprocesos. Los resultados obtenidos evidencian que al asumir o descartar determinado comportamiento se pueden identificar variantes alternativas del modelo de proceso.

El análisis del segundo registro de eventos tuvo como base el modelo de proceso correspondiente. Se aplicaron las técnicas Alineación de Trazas y Descomposición en Subprocesos y se compararon sus resultados con los obtenidos por la Minería de Variantes. En una de las variantes de descomposición se identificaron patrones de control de flujo similares a los obtenidos utilizando las otras técnicas de diagnóstico. La asunción de comportamiento en el registro de eventos permitió identificar patrones de control de flujo y variantes del modelo de proceso que no fueron identificados por las demás técnicas. Los patrones de control de flujo identificados mediante la Minería de Variantes son consecuentes con el modelo de proceso.

Los resultados de la aplicación de la Minería de Variantes son similares a los obtenidos por otras técnicas de diagnóstico. La capacidad de identificar la presencia de ruido y detectar la ausencia de información permite realizar el diagnóstico sobre registros de eventos provenientes de entornos reales.

Conclusiones

En el presente trabajo se realizó un análisis de las técnicas existentes en la minería de proceso que han sido empleadas en el diagnóstico de proceso. La creación de una nueva técnica basada en el enfoque de la identificación de variantes de modelos de proceso permitió incluir las potencialidades y solventar las principales limitaciones de las técnicas existentes. El desarrollo de la técnica Minería de Variantes a partir del algoritmo propuesto permitió el cumplimiento del objetivo general planteado. La investigación realizada permitió arribar a las siguientes conclusiones:

- Los resultados de la aplicación de las técnicas existentes para el diagnóstico de proceso pueden ser complementados a partir de la identificación y el manejo de características inherentes a registros de eventos reales como ruido y/o ausencia de información.
- Al obviar o asumir la presencia de determinado comportamiento en el registro de eventos se pueden obtener, para diferentes patrones de control de flujo, diferentes variantes de descomposición jerárquica del proceso.
- El algoritmo de descomposición en subprocesos permite la identificación de los patrones de control de flujo presentes en un proceso a partir del procesamiento de diferentes combinaciones de los comportamientos presentes en el registro de eventos.
- La utilización por parte de la técnica Minería de Variantes del algoritmo de descomposición en subprocesos permite obtener el comportamiento considerado como ruido y las trazas consideradas como incompletas durante la identificación de los patrones.
- Los resultados obtenidos mediante la aplicación de la técnica desarrollada sobre registros de eventos provenientes de entornos reales y su comparación con los modelos de proceso existentes evidencian la capacidad de dicha técnica para identificar los patrones de control de flujo presentes en el proceso.
- La comparación de los resultados de la aplicación de la Minería de Variantes con otras técnicas de diagnóstico de proceso muestran su capacidad para identificar

variantes de modelos de proceso considerando la presencia de ruido y/o ausencia de información en registros de eventos.

Referencias bibliográficas

- AALST W. M. P. 2011.** *Process Mining. Discovery, Conformance and Enhancement of Business Processes.* London : s.n., 2011, Springer. 978-3-642-19344-6.
- AALST W. M. P. 2013.** *Mine your own business: Using process mining to turn big data into real value.* Eindhoven : s.n., 2013.
- AALST W. M. P., Rubin, V., Verbeek, H.M.W., van Dongen, B.F., Kindler, E., Gunther, C.W. 2009.** *Process Mining: a two-step approach to balance between underfitting and overfitting. Software and Systems Modeling.* 2009.
- AALST W. M. P., VAN DONGEN., B.F., J. HERBST, L., MARUSTER. 2003.** *Workflow Mining: A Survey of Issues and Approaches.* 2003.
- AALST W. M. P., WEIJTERS., A J M M y MARUSTER., L. 2004.** *Workflow Mining: Discovering process models from events logs.* 2004. págs. 1128-1142. Vol. 16.
- ADRIANSYAH A., DONGEN B.F., AALST W.M.P.,. 2011.** *Towards Robust Conformance Checking.* 2011.
- BOSE R. P. J. C. 2012.** *Process Mining in the Large: Preprocessing, Discovery, and Diagnostics.* 2012.
- BOSE R. P. J. C., AALST W. M. P.,. 2009.** *Abstractions in Process Mining: A taxonomy of patterns.* Berlín : Springer, 2009. págs. 159-175.
- BOSE R. P. J. C., AALST W. M. P.,. 2012.** *Process diagnostics using trace alignment: Opportunities, issues, and challenges.* 2012. págs. 117-141. Vol. 37.
- BOSE R. P. J. C., AALST W. M. P.,. 2012.** *Trace Alignment in Process Mining: Opportunities for Process Diagnostic.* 2012.
- CHEUNG C.F., YU J.X., LU H.,. 2005.** *Constructing Suffix Tree for Gigabyte Sequences with Megabyte Memory.* 2005.
- DE MEDEIROS A. K. A., WEITERS A.J.M.M., AALST. W.M.P.,. 2007.** *Genetic Miner process mining: An experimental evaluation.* 2007. págs. 245-304. Vol. 14.
- DE WEERDT J., DE BACKER M., VANTHIENEN J., BAESENS B.,. 2012.** *A multi-dimensional quality assesment of state-of-the-art process discovery algorithms using real-life event logs.* 2012. págs. 654-676. Vol. 37.
- DEPAIRE BENOIT, SWINNEN JO, JANS MIEKE y VANHOOF KOEN. 2013.** *A Process Deviation Analysis Framework.* Berlín : s.n., 2013. págs. 701-706. 978-3-642-36284-2.
- DONGEN B.F.VAN, ADRIANSIYAH A.,. 2009.** *Process Mining: Fuzzy Clustering and Performance Visualization.* 2009.

- GÜNTHER C.W. 2009.** *Process Mining in Flexible Enviroments*. University of Technology. Eindhoven : s.n., 2009.
- GÜNTHER C. W., Aalst W. M.P., 2007.** *Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics*. Brisbane, Australia : Springer, 2007. págs. 328-343.
- HENDRICKS K., SINGHAL. V. R., STRATMAN J. K., 2007.** *The impact of enterprise systems on corporate performance: A study of ERP, SCM and CRM system implementations*. 2007. págs. 65-82. Vol. 25.
- JANS MIEKE., VAN DER WERF, JAN MARTIJN, LYBAERT, NADINE., VANHOOF, KOEN. 2011.** *A business process mining application for internal transaction fraud mitigation*. 2011. págs. 13351-13359. Vol. 38.
- LEEMANS S.J.J., FAHLAND D., AALST W. M. P., 2013.** *Discovering Block-Structured Process Models From Event Logs Containing Infrequent Behaviour*. 2013.
- LI C., BOSE R.P.J.C., 2010.** *Mining Context-Dependent and Interactive Business Process Maps using Execution Patterns*. s.l. : Springer, 2010. págs. 109-121.
- LI C., REICHERT M., A WOMBACHER., 2011.** *Mining business process variants: Challenges, scenarios, algorithms*. 2011. págs. 409-434. Vol. 70.
- LY L. T., INDIONO CONRAD., MANGLE, JURGEN., RINDERLE-MA STEFANIE. 2012.** *Data Transformation and Semantic Log Purging for Process Mining*. 2012.
- MA L. 2012.** *How to evaluate the Performance of Process Discovery Algorithms*. Eindhoven University of Technology. 2012. Master Thesis.
- FUNDORA-RAMÍREZ. O. 2013.** *Impacto de las características de un registro de evento en algoritmos de descubrimiento de la minería de proceso*. 2013.
- ROZINAT A., AALST W.M.P., 2008.** *Conformance checking of processes based on monitoring real behavior*. 2008. págs. 64-65. 0306-4379.
- SAÑUDO ORTIZ, LIUVER ROMEL., BATISTA GARCÍA, LILIBETH. 2012.** Ventajas del HIS sobre otras aplicaciones bajo el modelo SaaS. [En línea] 9 de marzo de 2012. [Citado el: 28 de abril de 2014.] http://repositorio_institucional.uci.cu/jspui/handle/ident/3900. 978-959-286-019-3.
- SEPPE K.L.M, VANDEN BROUCKE, JAN VANTHIENEN, BART BAESENS. 2013.** *Third International Business Process Intelligence Challenge (BPIC'13): Volvo IT Belgium VINST*. 2013.
- SONG M., AALST W.M.P., 2007.** *Supporting process mining by showing events at a glance*. 2007. págs. 139-145.
- TECHNISCHE UNIVERSITEIT EINDHOVEN. 2013.** 9th International Workshop on Business Process Intelligence 2013. [En línea] 2013. [Citado el: 15 de 04 de 2014.] <http://www.win.tue.nl/bpi/2013/challenge>.

VAN DONGEN B., ALVES DE MEDEIROS A., WEN L., 2009. *Process Mining: Overview and Outlook of Petri Net Discovery Algorithms*. 2009. págs. 225-242.

WEIJTERS A.J.M.M., AALST W. M. P., DE MEDEIROS A. K. A., 2006. *Process mining with the heuristics miner-algorithm*. 2006. Vol. 166.

WEIJTERS A.J.M.M., AALST. W.M.P., 2003. *Rediscovering Workflow Models from Event-Based Data using Little Thumb*. 2003. págs. 151-162. Vol. 10.

WESKE M. 2007. *Business Process Management. Concepts, Languages, Architectures*. 2007. 978-3-540-73521-2.

YANG H., DONGEN B.F., HOFSTEDE A.H.M. TER., WYNN M. T., WANG JIANMIN. 2012. *Estimating completeness of event logs*. 2012.

YZQUIERDO-HERRERA R. 2012. *Modelo para la estimación de información ausente en las trazas usadas en la minería de proceso*. Universidad de las Ciencias Informáticas. 2012. Phd Thesis.

YZQUIERDO-HERRERA R., SILVERIO CASTRO R., LAZO CORTÉS M. TORRES GRAÑA A., 2012. *Diagnóstico de proceso basado en el descubrimiento de subprocesos*. 2012. Vol. 33.

YZQUIERDO-HERRERA R., Silverio-Castro R., Lazo-Cortés M., 2012. *Sub-process discovery: Opportunities for Process Diagnostics*. GHENT : Springer, 2012.