

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

Facultad 3



*Algoritmo para el diagnóstico de procesos en la
minería de proceso.*

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

Autor: Maidelis Hechavarría Borrero

Lilian Hernández Suárez

Tutores: Ms.C. Raykenler Izquierdo Herrera

Ing. Ariadna Rendón

Ing. Lisett Díaz Mesa

Ciudad de la Habana, junio del 2012

"Año 54 de la Revolución"

Declaración de autoría

Declaramos ser autores del presente trabajo de diploma y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, 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 _____.

Maidelis Hechavarría Borrero

Lilian Hernández Suárez

Firma de los Autores

Ms.C. Raykenler Izquierdo Herrera

Lisett Díaz Mesa

Firma de los Tutores



Dedicatorias

Maidelis:

A mis padres. En especial:

A la persona que sus brazos siempre se abren cuando necesito un abrazo, que su corazón sabe comprender cuando necesito una amiga, que sus ojos sensibles se endurecen cuando necesito una lección, la que su fuerza y su amor me han dirigido por la vida, a la persona que le debo todo lo que soy, a ti mi mami te dedico este trabajo.

Agradecimientos

Maidelis:

Quiero agradecerles a todas aquellas personas que de una forma u otra me apoyaran durante el transcurso de mi carrera.

Quiero agradecerle a mi madre [Sixta] por su apoyo, amor y comprensión. Por siempre confiar en mí en los momentos más difíciles, por ser la principal promotora de este sueño, la mujer más fuerte que conozco, eres mi más grande orgullo, yo te quiero tal y como eres. Si alguna vez no te he dicho cuanto te amo, cuanto te quiero y cuanto te necesito en esta vida, quiero que lo sepas en este momento, y este es mi regalo para ti. Todo lo que he sido y en lo que hoy me convierto te lo debo a ti. Te quiero con la vida. Gracias por existir.

A mi tío, Aramis, no necesito sus apellidos porque tengo su amor, que me es suficiente. Gracias por estar siempre para mí y mi hermana, sin importar ninguna circunstancia, gracias por tu apoyo, tus consejos sobre la vida, los estudios e incluso los hombres. Sin tu apoyo este trabajo no hubiera salido, aquí está tu mano; espero no haberte defraudado, porque lo he hecho para que te sientas orgullosa de tu sobrina

A mis hermanas [Lisandra, Lilian, Lismarita, Arisleidis y Adisleidis], por sentirse orgullosos de mí y creer en mí como siempre lo han hecho.

A mi niña [Lismarita], que no hemos tenido la oportunidad de crecer juntas, pero si sabiendo cuanto nos queremos, cuan orgullosa se siente una de la otra.

Gracias a jandra como yo la llamo cariñosamente, por su apoyo durante los tantos años que hemos estudiados juntas, y que sin ella tampoco estaría donde estoy.

A Adis por preocuparte por mis estudios, durante este tiempo. A Lili que sé que aunque no estará conmigo el día más importante de mi vida, ellas y las demás pensarán en mí, y pedirán para que todo salga bien, y dirán con orgullo el nombre de su hermana, porque ustedes son las mejores hermanas; los quiero mucho.

A mi papá [Ernesto] por ser unas de las grandes cosas que tengo en la vida, por enseñarme que nunca se puede uno conformar, siempre hay que superarse, para el también va dedicado este sueño. Gracias por educarme y darme lo mejor, sé que no podrá estar conmigo hoy pero estará rezando para que todo me salga bien. Quiero que sepas que te quiero con la vida, y sería capaz de hacer cualquier cosa por ti
A mis tutores, por la confianza depositada en mí, en especial a [Raykenler Izquierdo] por brindarnos su experiencia, ayuda y sus consejos siempre que los necesitamos.

A todos mis amigos y compañeros del politécnico y de la UCI con los cuales he compartido y vivido momentos inolvidables: A Areanne, Dailiana, Sandra, Mileidis, Lucrecia, Yaimi, Wilfredo, Nelson, Adrian, Yanier, Cesar, Lianet, Yaili, Yaillet, Vladimir, Lilian.

Agradecimientos

A Jaksson por los buenos momentos.

A Delio por los consejos.

A mis mejores amigas, Yulia, Marianna y Yadira, por brindarme su amistad, demostrarme que los amigos están siempre, en las buenas y malas, y ellas siempre has estado, me han secado lágrimas, me has hecho llorar, me has dado cariño y me han enseñado que en la vida todo se puede. Gracias por creer en mí en todos estos años y nunca, nunca defraudarme ni dejarme sola un solo momento.

A toda mi familia Gracias a todos por ayudarme en todo desde que entré a la Universidad, por tratar de darme lo mejor y creer en mí, gracias a todos, mis tíos, primos.

A mi ex novio, Annier, por haberme hecho reír en los momentos difíciles, por brindarme su amor y confianza. Gracias por acompañarme y darme fuerzas en todo el tiempo que estuvimos juntos, sé cuánto te agobié con mis lágrimas y mi pesimismo, pero nada, ya salió. Eres una de las mejores cosas que me ha sucedido en la vida, gracias por levantarme los ánimos, por estar cerquita de mí ahora que no somos una pareja, gracias por hacerme sentir importante para ti.

*A una persona muy especial que nos ayudo en el desarrollo de la tesis Maikel Pino.
Muchas gracias.*

Y aunque sea ya la mencioné anteriormente, quiero agradecerle nuevamente a Lilian mi compañera de tesis. Por el apoyo durante la realización de este trabajo, por los momentos que estuvimos hasta tarde sin dormir y por la paciencia.

Gracias a todas las personas que me han ayudado a construir este sueño

Síntesis

La mayoría de las empresas utilizan sistemas de información para gestionar la ejecución de sus procesos de negocio. Estos registran en forma de trazas las acciones que se van realizando cuando se ejecutan instancias o casos particulares del proceso de negocio. Al descubrimiento, monitoreo y mejora del proceso a partir de la información contenida en las trazas se le denomina Minería de Procesos. Actualmente la Minería de Procesos ha evolucionado y se ha ampliado el área de investigación, obteniendo como resultado el desarrollo de varios algoritmos que permiten obtener información a partir de un registro de eventos. Las técnicas desarrolladas en este ámbito, no son capaces de facilitar la posterior selección de las técnicas de descubrimiento, a partir del diagnóstico realizado. En el presente trabajo, se realiza un estudio de las principales investigaciones que abordan los temas de diagnóstico y descubrimiento de los procesos. En el mismo se lleva a cabo el desarrollo de un algoritmo que facilite la selección de las técnicas de descubrimiento. Finalmente, para validar la propuesta se implementa el algoritmo en la herramienta ProD.

PALABRAS CLAVE: *diagnóstico, descubrimiento, eventos, Minería de Procesos, registro de eventos, trazas.*



“El secreto del éxito es la constancia en el propósito”

Benjamín Disraeli

Índice de contenidos

Introducción.....	11
Capítulo 1: Fundamentación teórica.....	18
Introducción al capítulo.....	18
1.1. Proceso.....	18
1.2. Proceso de negocio.....	19
1.3. Modelado de procesos de negocio.....	20
1.4. Minería de Procesos.....	22
1.5. Diagnóstico de procesos.....	23
1.6. Trabajos relacionados sobre diagnóstico de procesos.....	26
1.7. Descubrimiento.....	31
Conclusiones del capítulo.....	37
Capítulo 2: Propuesta de solución.....	39
Introducción.....	39
2.1 Aspectos que caracterizan un proceso de negocio.....	39
2.2. Evaluación de las técnicas de descubrimiento de acuerdo a los aspectos que caracterizan un proceso.....	41
2.3. Evaluación de las técnicas de diagnóstico de acuerdo a los aspectos que caracterizan un proceso.....	43
2.4. Detención de las características del proceso analizado.....	44
2.5. Sugerencia de algoritmos de descubrimiento.....	46
2.5.1. Determinar patrones en un árbol de bloques de construcción.....	46
2.5.2. Detectar posibles casos afectados por el ruido.....	47
2.5.3. Determinar características del proceso.....	48
2.5.4. Sugerir técnica de descubrimiento.....	49
Conclusiones del capítulo.....	51
Capítulo 3: Validación de la solución.....	53
Introducción.....	53
4.1. Implementación del algoritmo propuesto.....	53
3.1.1. Herramienta y tecnología utilizada.....	53

Índice de contenidos

3.1.2. Aplicación ProD.	54
3.2. Resultados experimentales.	58
3.2.1. Métrica.	58
3.2.2. Características de los procesos analizados.	58
3.2.3. Análisis de los resultados.	59
Conclusiones del capítulo.	65
Conclusiones generales.	67
Recomendaciones.	68

Índice de tablas e imágenes

Tabla 1: Caracterización de las técnicas de descubrimiento.	41
Tabla 2: Evaluación de las técnicas de diagnóstico.	43
Tabla 3: Descripción de los registros de eventos.	59
Tabla 4: Sugerencias para el Proceso 1.....	61
Tabla 5: Fitness para los modelos obtenidos del Proceso 1.....	62
Tabla 6: Sugerencias para el Proceso 2.....	63
Tabla 7: Fitness de los modelos obtenidos del Proceso 2.	64
Tabla 8: Sugerencia del Proceso 3.....	65
Tabla 9: Fitness de los modelos obtenidos del Proceso 3.	65
Tabla 10: Caracterización de las técnicas de descubrimiento a partir del constructor Tareas invisibles.	LXXIII

Figura 1: Representación de los tres grupos de técnicas de la Minería de Procesos.	12
Figura 2: Árbol de bloques de construcción.	31
Figura 3: Árbol de construcción con los aspectos que caracterizan al proceso.	45
Figura 4: Árbol de construcción con los aspectos que caracterizan al proceso 1.	45
Figura 5: Árbol de construcción estimado con tareas invisibles.....	46
Figura 6: Alineación de las trazas.	55
Figura 7: Árbol de construcción de bloques.	55
Figura 8: Estimación de información ausente.	56
Figura 9: Sugerencia para el descubrimiento de modelo de proceso.	57

Introducción

Introducción.

En la actualidad los procesos de negocio (BP por sus siglas en inglés), están recibiendo mucha atención por parte de la comunidad científica, apoyado en el hecho de que estos son un recurso importante para mantener la competitividad y ayudar al desempeño en organizaciones. Para el perfeccionamiento de estos procesos es necesario llevar a cabo un mejor control sobre ellos, en los cuales se pueden utilizar disímiles métodos, tanto cualitativos como cuantitativos, logrando medirlos, crear valor hacia el cliente y al mismo tiempo mejorar los resultados de la organización (MONARDES, 2011).

Las empresas se apoyan en el uso de sistemas de información que sean capaces de gestionar sus procesos de negocios, trayendo estos innumerables beneficios entre los que se pueden encontrar: automatización de los procesos operativos, suministrar una plataforma de información necesaria para la toma de decisiones y usabilidad de la información. Por su utilidad, se han hecho indispensables en las organizaciones, siendo utilizados por la mayoría de las empresas en el mundo. Estos sistemas de información tienen la capacidad de registrar en forma de trazas las acciones y actividades que se van realizando cuando se ejecutan instancias o casos particulares del proceso de negocio, salvando esta información en un archivo denominado registro de eventos, que permite la persistencia de este historial. Cada entrada del registro de eventos contiene disímiles informaciones sobre las instancias ejecutadas, donde puede variar de acuerdo al sistema que lo genere, por lo general, poseen información acerca del comienzo y finalización de una actividad, incluyendo el agente que la lleva a cabo.

Hasta hace poco tiempo, la información de estos registros de eventos rara vez se utilizaba para analizar los procesos, hoy se pueden utilizar para su descubrimiento (AALST, 2011), a esta nueva disciplina se le denomina Minería de Procesos y tiene como idea el descubrir, monitorear y mejorar los procesos reales (procesos no ficticios), mediante la extracción del conocimiento existentes en las trazas y suministrado por los sistemas de información (AALST, 2011; AALST and WEIJTERS, 2004).

La información proporcionada por los sistemas de información puede ser analizada usando las diferentes técnicas desarrolladas para la Minería de Procesos, las cuales permiten hacer explícito el conocimiento contenido en dichas trazas y responder preguntas acerca de los procesos operativos. Algunas de estas son: ¿Qué sucedió en el pasado?, ¿Por qué sucedió?, ¿Qué es probable que suceda en el futuro?, ¿Cuándo

Introducción

y por qué las organizaciones y las personas se desvían?, ¿Cómo llevar un mejor control sobre un proceso? y ¿Cómo volver a diseñar un proceso para mejorar su desempeño?

La Minería de Procesos en la investigación es un área joven, pero a la vez poderosa, que permite realizar un análisis de los procesos basado en su funcionamiento actual. Tiene como beneficios el descubrimiento de modelos representativos de la realidad, la detección de desviaciones, a partir de la comparación entre el registro de eventos y el modelo representativo del proceso, el descubrimiento de redes sociales y modelos organizacionales relacionado con el proceso que se está analizando; y por último permite extender un modelo existente teniendo en cuenta el análisis de su funcionamiento.

Las técnicas de Minería de Procesos posibilitan, hacer recomendaciones y predicciones, teniendo en cuenta el análisis de los datos actuales y de los históricos. Estas se pueden clasificar en tres grupos tal y como se muestra en la Figura 1. En el primer grupo se encuentran las destinadas al descubrimiento de modelos de procesos, en el segundo las dedicadas al chequeo de conformidad¹ (AALST *et al.*, 2011) y por último, el tercer grupo se ocupa de la extensión de un modelo existente teniendo en cuenta la información almacenada en un registro de eventos (AALST; AALST, 2012).



Figura 1: Representación de los tres grupos de técnicas de la Minería de Procesos

(a) Descubrimiento, (b) Verificación de conformidad, y (c) Mejoramiento (AALST, 2012).

¹Chequeo de conformidad: Se refiere a la comparación un modelo de proceso existente con un registro de eventos del mismo proceso, para chequear si la realidad, tal como está almacenada en el registro de eventos, es equivalente al modelo y viceversa.

Introducción

Uno de los retos actuales en la Minería de Procesos lo constituye el diagnóstico del proceso. Área de conocimiento que abarca el análisis de rendimiento, detección de anomalías e identificación de patrones comunes, visualizando de manera general los aspectos más significativos del proceso (AALST *et al.*, 2011). Para la realización del diagnóstico se han desarrollado diferentes técnicas, entre las cuales se encuentran Chequeo de Conformidad (ROZINAT, A., & AALST, W.M.P. VAN DER, 2008), Fuzzy Clustering (ADRIANSYAH, 2010) y Alineación de trazas (BOSE, R.P.J.C., & AALST, W.M.P. VAN DER., 2012.; R. P. JAGADEESH CHANDRA BOSEA, September 14, 2011).

Las técnicas de descubrimiento permiten revelar un modelo representativo del proceso ejecutado en la empresa, a partir de los registros de eventos (AALST, 2011; AALST *et al.*, 2011). En la literatura estudiada existen varias investigaciones sobre este tema, donde se han desarrollado diversos algoritmos, entre los cuales se pueden mencionar: Alpha (AALST, 2011; MEDEIROS, 2006), Genetic Miner (MEDEIROS, 2006) y Heuristic Miner (AALST, 2011). Por lo general, estos algoritmos manejan un conjunto de constructores de flujo de trabajo, que permiten reconocer una determinada estructura o patrón que aparece de manera común en el proceso (paralelismo, secuencia, selección, lazos, sin libre elección (non-free-choice), tareas invisibles y tareas duplicadas). Aún cuando existen técnicas que cubren la totalidad de los constructores de flujo de trabajo enunciados, la mayoría de estas dan un cubrimiento parcial a los mismos, presentando problemas para manejar determinadas situaciones donde se presentan, debido a que la notación utilizada para representar el modelo de proceso descubierto no es capaz de soportar estas características. Los aspectos que no pueden ser cubiertos en su totalidad por los algoritmos existentes, son: los lazos, sin libre elección, las tareas invisibles y tareas duplicadas (MEDEIROS, 2006).

En la última década, la Minería de Procesos ha evolucionado y se ha ampliado el área de investigación, obteniendo como resultado el desarrollo de varios algoritmos que permiten obtener información a partir de un registro de eventos. Esto demuestra que existen muchas técnicas diferentes para aplicar en esta área y principalmente en el descubrimiento de modelos. Para el análisis de un proceso cuando no se tiene una vista preliminar de sus características, se dificulta la selección de las técnicas de descubrimiento y entendimiento de proceso; obteniéndose como resultado un modelo que no refleja correctamente el comportamiento registrado en las trazas.

Introducción

A partir de lo planteado, sería conveniente que las técnicas de diagnóstico proporcionaran la información necesaria para facilitar el descubrimiento del modelo de proceso, sin embargo, no son claras en la caracterización del proceso analizado y no sugieren qué algoritmos utilizar en la fase posterior. Al no contar con esta información o que en ocasiones es muy escasa, el descubrimiento del modelo se torna engorroso, y conlleva a que se representen de manera incorrecta las relaciones entre las actividades, dificultando su comprensión y utilidad.

Por lo antes mencionado, es necesario obtener una eficiente técnica de diagnóstico, capaz de hacer una correcta caracterización del proceso, identificando los patrones más frecuentes, las posibles anomalías y los aspectos asociados al rendimiento del proceso. Además, esta técnica debe ser capaz de indicar a partir de este diagnóstico, qué técnica de descubrimiento se debe utilizar en la fase posterior, teniendo en cuenta la capacidad de la misma de descubrir constructores (paralelismo, secuencia, selección, lazos, sin libre elección (non-free-choice), tareas invisibles y tareas duplicadas), y de tratar con los efectos o daños causados por el ruido a los registros de eventos.

Los algoritmos existentes en el diagnóstico no facilitan la posterior selección de las técnicas de descubrimiento y entendimiento del proceso.

A partir de un análisis de la situación problemática planteada se define el **problema a resolver**: ¿Cómo detectar durante la etapa de diagnóstico del proceso, los patrones más frecuentes, las posibles anomalías y aspectos asociados al rendimiento del proceso, de forma que se facilite la posterior selección de las técnicas de descubrimiento y entendimiento del proceso. **El objeto de estudio** es la Minería de Procesos y el **objetivo General**: Desarrollar un algoritmo que a partir de las técnicas desarrolladas para el diagnóstico del proceso, facilite la posterior selección de las técnicas de descubrimiento y entendimiento del proceso.

Desglosándose los siguientes **objetivos específicos**:

1. Fundamentar la investigación a partir de la elaboración del marco teórico.
2. Identificar los aspectos que permiten caracterizar un proceso.
3. Desarrollar un algoritmo que a partir de los aspectos definidos sugiera qué técnicas de descubrimiento se deben emplear.

4. Validar usando 3 escenarios diferentes la propuesta realizada.

Atendiendo a situaciones particulares que constituyen parte del propósito general, habitualmente es recomendable plantear un conjunto de **tareas a cumplir**:

1. Análisis del diagnóstico del proceso en el contexto de la Minería de Procesos.
2. Definición de los aspectos que caracterizan un proceso.
3. Evaluación de las técnicas de diagnóstico del proceso más adecuadas en relación con los aspectos antes definidos.
4. Evaluación de las técnicas de descubrimiento del proceso con respecto a los aspectos que caracterizan un proceso.
5. Detección de los aspectos que caracterizan un proceso haciendo uso de técnicas de diagnóstico.
6. Evaluación de los aspectos detectados para la conformación de la propuesta de las técnicas de descubrimiento que se deben emplear.
7. Descripción de los escenarios seleccionados para la validación.
8. Valoración de los resultados de aplicar de las pruebas en los 3 escenarios escogidos.

Como relación entre el problema, los objetivos y el objeto de estudio, limitando del mismo las partes que permitan desarrollar el proceso investigativo con que se alcanza el objetivo, surge como **campo de acción** el diagnóstico del proceso.

Como **posibles resultados**:

- Descripción de las técnicas de diagnóstico del proceso más adecuadas en relación con los aspectos antes definidos.
- Descripción de las técnicas de descubrimiento del proceso con respecto a los aspectos que caracterizan un proceso.
- Descripción de los aspectos que caracterizan un proceso.
- Algoritmo que sugiere qué técnicas de descubrimiento emplear en relación con las características del proceso.

Introducción

Idea a defender: El desarrollo de un algoritmo que permita detectar los patrones más frecuentes, las posibles anomalías y aspectos asociados al rendimiento del proceso, facilitará la posterior selección de las técnicas de descubrimiento y el entendimiento del proceso.

Para el desarrollo de la investigación se utilizaron los siguientes métodos científicos:

Métodos teóricos:

- Histórico lógico.
- Hipotético deductivo.
- Analítico-Sintético.
- Sistémico.

El método histórico lógico se utiliza para enfocar la problemática presentada en la investigación asociada al diagnóstico del proceso; en la primera parte de la investigación se desarrolla un estudio del estado del arte de la problemática analizada, dando detalles de las ventajas y deficiencias de cada uno de los métodos.

La investigación sigue además un método hipotético deductivo, ya que a partir del problema concreto, se plantean objetivos específicos e idea a defender, que en el transcurso de la investigación son resueltos, siguiendo métodos científicamente bien fundamentados.

El método analítico-sintético se utiliza para descubrir los distintos elementos que componen la naturaleza o esencia asociada al fenómeno del diagnóstico del proceso. Definiéndose las causas y los efectos, para posteriormente integrar los elementos en una unidad nueva, en una comprensión total de la esencia de lo que ya se conoce en todos sus elementos y particularidades.

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

Métodos empíricos

- Experimentación.

Introducción

- Medición.

Además de utilizar métodos teóricos se utilizan los métodos empíricos, basando la investigación en la experimentación con datos provenientes de situaciones reales.

Se aplican pruebas estadísticas bien fundamentadas, para analizar la efectividad del modelo desarrollado y la calidad de las respuestas finales. Se establecen estadígrafos e indicadores adecuados que permiten realizar correctas mediciones de los resultados.

El presente documento está estructurado de la siguiente manera: síntesis, introducción y tres capítulos de los cuales a continuación se resume su contenido.

Capítulo 1: Se discuten las ideas básicas y las diferentes estrategias utilizadas en la Minería de Procesos, específicamente las relacionadas con el descubrimiento y diagnóstico del proceso. Se hace una evaluación crítica de las ventajas y desventajas de los diferentes enfoques.

Capítulo 2: Se describe la propuesta de solución. Se presentan los aspectos que se deben considerar para la caracterización de los procesos. También se presentan las técnicas de diagnóstico en correlación con los aspectos que caracterizan al proceso y las sugerencias correspondientes para el posterior descubrimiento del proceso.

Capítulo 3: Se describe el mecanismo de validación de la propuesta y se presentan los resultados obtenidos.

Capítulo 1: Fundamentación teórica:

Introducción al capítulo.

Los sistemas de información permiten facilitar el manejo de los procesos que se llevan a cabo en una organización a través del flujo de trabajo, entre estos se encuentran los de Planificación de los Recursos Empresariales (ERP), los de Gestión de los Procesos de Negocio (BPM por sus siglas en inglés) (WIL M. P. VAN DER AALST, 2003) y los de Gestión de Flujo de Trabajo (WFM por sus siglas en inglés)(AALST, 2004). Estos sistemas proveen herramientas que facilitan la administración de los procesos y de las personas involucradas en cada instancia o tarea que se lleva a cabo. Además, son capaces de proveer todas las etapas en el ciclo de vida de la administración de procesos: (re)diseño, configuración, ejecución, control y diagnóstico.

Partiendo de la información anterior, en este capítulo se realizará un estudio del estado del arte, relacionado con el problema al que se le dará solución y se revisarán los principales conceptos asociados al tema de investigación.

Conceptos asociados al dominio del tema.

Inicialmente se exponen un conjunto de definiciones que son necesarias para una mejor comprensión de la investigación.

1.1. Proceso.

La palabra proceso proviene del latín *proessus*, y significa guía, paso a paso de las actividades a realizar para transformar elementos de entrada en salidas.

Actualmente existen disímiles conceptos relacionado con este término.

Se define un proceso como: “**quién**” está haciendo “**qué**”, “**cuándo**” y “**cómo**” para alcanzar un determinado objetivo (JACOBSON, 1998).

La Norma de Calidad ISO 9000:2000 define un proceso como: “Conjunto de actividades mutuamente relacionadas o que interactúan, las cuales transforman elementos de entrada en resultados” (COLÍN, 2000).

En el año 2000, Huacoto enuncian que un proceso es un conjunto de prácticas que se ejecutan con un propósito determinado, las cuales convierten entradas en salidas que

Capítulo 1: Fundamentación teórica

son de valor para el cliente. El proceso puede incluir herramientas, métodos y/o personas (HUACOTO, 2005).

Ricardo Camacho en el 2008 describe un proceso como el conjunto de actividades o tareas mutuamente relacionadas entre sí, que admite elementos de entrada durante su desarrollo, ya sea al inicio o a lo largo del mismo, los cuales se administran, regulan o autorregulan bajo modelos de gestión particulares para obtener elementos de salida o resultados esperados (CAMACHO, 2008).

A partir de las definiciones anteriores se puede concluir que un proceso es el conjunto de recursos y actividades interrelacionados, que transforman elementos de entrada en salidas. Los recursos pueden incluir personal, finanzas, instalaciones, equipos, técnicas y métodos. También se puede decir que es la sucesión de pasos y decisiones que se siguen para realizar una determinada actividad o tarea. Los elementos de entrada y los resultados previstos pueden ser tangibles (equipos, materiales o componentes), o intangibles (energía o información). Cada proceso tiene clientes y otras partes interesadas (quienes pueden ser internos o externos a la organización) que son afectados por el proceso y quienes definen los resultados requeridos de acuerdo con sus necesidades y expectativas.

1.2. Proceso de negocio.

El proceso de negocio es un tipo específico de proceso, que describe desde un punto de vista orientado al mercado las actividades de una organización. El principal objetivo de los procesos de negocio es satisfacer las necesidades de los clientes.

Los procesos de negocio describen cómo es realizado el trabajo en la empresa y se caracterizan por ser: observables, medibles, mejorables y repetitivos (QUINTANA, 2002). Además pueden ser parte de un proceso más grande que lo comprenda, o bien pueden contener otros procesos de negocio que deban ser incluidos en su función.

Según la ISO 9001-2000, los procesos de negocio son la base para comprender mejor la forma en que opera un negocio en sus diferentes áreas y son una herramienta fundamental para acceder a modelos de calidad y eficiencia en sus actividades (TERALOC, 2004).

Matías Weske, en el 2007 define un proceso de negocio como 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

Capítulo 1: Fundamentación teórica

proceso de negocio es ejecutado por una simple organización, pero con él pueden interactuar procesos de negocios de otras organizaciones (WESKE, 2007).

A partir de los conceptos mencionados se define un proceso de negocio como el conjunto estructurado de actividades, tareas o procedimientos mutuamente relacionados entre sí, que admite elementos de entrada durante su desarrollo, ya sea al inicio o a lo largo del mismo, los cuales se administran, regulan o autorregulan bajo modelos de gestión particulares, para obtener elementos de salida o resultados esperados que son valiosos para un cliente. Estos procesos implican un fuerte énfasis en *cómo* se hace el trabajo en una organización, en contraposición al *qué* de producto.

Subproceso.

Un subproceso es una agrupación de actividades del negocio que representan una compleja y lógica unidad de trabajo. Los subprocesos tienen sus propios atributos y metas, pero contribuyen al objetivo del proceso que los contiene. Un subproceso es también un proceso y su mínima expresión es una actividad.

Un proceso puede descomponerse en varios subprocesos mediante algunos de los patrones de flujo de control conocidos como: secuencia, paralelismo, selección exclusiva y lazos.

1.3. Modelado de procesos de negocio.

El modelado de procesos de negocio es el modelo de uno o más procesos de negocio de una empresa y define las formas en que las operaciones se llevan a cabo para lograr los objetivos deseados. Tiene como objetivo ilustrar un proceso completo, permitiendo a los analistas, gerentes, consultores y otro personal, analizar y mejorar la eficiencia y calidad de los procesos (SMITH, 2003).

Curtis, afirma que existen cuatro puntos de vista en cuanto al modelado de los procesos de negocio: vista funcional (*qué*), la cual representa la dependencia funcional entre los elementos del proceso, vista dinámica (*cuándo, cómo*), que proporciona una secuenciación y control de la información sobre el proceso, vista informacional, que incluye la descripción y relación entre las entidades que son producidas, consumidas o incluso manipuladas por los procesos y la vista organizacional (*quién, dónde*) que describe quién desarrolla cada tarea o función y dónde se desarrolla dentro de la organización (CURTIS, 1992).

Capítulo 1: Fundamentación teórica

El modelado de procesos de negocio es considerado la base para comprender mejor la operación de una organización, documentar y publicar los procesos, buscando una estandarización en la organización y una eficiencia en la integración de soluciones en arquitecturas orientadas a servicios.

Este modelado resulta útil en varias situaciones que pueden ser clasificadas en tres grupos: descripción del proceso, análisis del proceso e implementación del proceso. Actualmente se utilizan diversas técnicas de modelado como: redes de Petri, diagramas de actividades utilizando UML, diagrama de procesos de negocio usando BPMN (abreviado como BPMN-BPD) y otros lenguajes, fundamentalmente gráficos y formales. Los elementos que debe tener un modelo de proceso se pueden encontrar típicamente:

- Objetivo(s) o motivo del proceso.
- Entradas.
- Salidas.
- Recursos utilizados.
- Secuencia de actividades.
- Eventos que dirigen el proceso.
- Roles/participantes involucrados.

De manera general el modelado de procesos de negocio es adoptado por diferentes empresas de todos tipos y tamaños, con 3 objetivos principalmente:

- **Documentar.** Los procesos son parte fundamental de la organización de una empresa y un elemento primordial cuando se intenta implementar modelos de calidad como ISO².
- **Mejorar.** Las empresas buscan una mayor eficiencia en sus procesos. Tratando de: localizar “cuellos de botella” en su gestión e identificar área de oportunidad o mejora, por tanto, recurren al modelado y la simulación de procesos.
- **Agilizar.** En un nivel de mayor sofisticación, las empresas requieren el modelado de procesos como articuladores de los servicios de Tecnologías de información, para poder reaccionar con mayor agilidad a los constantes cambios que exige la competencia actual.

² ISO (International Standard Organization): Es la Organización Internacional para la Estandarización.

Capítulo 1: Fundamentación teórica

A partir de los conceptos antes mencionados, se puede definir que un modelo de procesos de negocio permite registrar mediante alguna notación la esencia del negocio, y hace posible que diversos actores dentro de una empresa lo entiendan y modifiquen con el propósito de incorporar mejoras.

1.4. Minería de Procesos.

Se conoce como Minería de Procesos al conjunto de técnicas que permiten el análisis de diferentes procesos de negocio, basándose en los registros de información proporcionados por sistemas de información transaccionales, de administración de flujo de trabajo (WFM por sus siglas en inglés), de Planificación de los Recursos Empresariales (ERP por sus siglas en inglés), de Gestión de la Relación con los Clientes (CRM por sus siglas en inglés), de Gestión de Cadenas de Abastecimiento (SCM por sus siglas en inglés), etc. La minería tiene como objetivo descubrir, monitorear y mejorar los procesos reales (no los procesos supuestos) a través de la extracción de conocimiento de los registros de eventos ampliamente disponibles en los sistemas de información (AALST, 2012).

Las diferentes técnicas de la Minería de Procesos permiten el descubrimiento de información desde los registros de eventos, permitiendo conocer el flujo de ejecución de un proceso, las redes sociales y sus métricas de desempeño (AALST *et al.*, 2011), y de esta manera aprovechar todos los datos de eventos registrados de una forma significativa para proveer un mejor entendimiento, identificar cuellos de botella, prever problemas, registrar violaciones de políticas, recomendar contramedidas y simplificar procesos, además de disminuir los tiempos de diseño y con ello sus costos. Ellas utilizan la información contenida en un registro de eventos³, el cual contiene un grupo de trazas⁴ (RAYKENLER YZQUIERDO HERRERA, 2011).

Es importante destacar que la Minería de Procesos provee un significativo puente entre la Minería de Datos, el modelado y el análisis de los procesos de negocio. Bajo el área de la Inteligencia de Negocios (Business Intelligence, BI por sus siglas en inglés), se han difundido un grupo de términos que encierran diferentes tipos de análisis en este contexto, tales como, Monitoreo de Actividades de Negocio (Business

³Evento: Cada evento se refiere a una actividad (paso bien definido) y se relaciona con un caso particular o instancia de proceso.

⁴Traza: Es una secuencia de eventos ordenados en el tiempo. Cada traza representa una instancia del proceso ejecutado.

Capítulo 1: Fundamentación teórica

Activity Monitoring, BAM por sus siglas en inglés), que alude a las tecnologías que hacen posible un análisis en tiempo real de los procesos de negocio. El Procesamiento de Eventos Complejos (Complex Event Processing, CEP por sus siglas en inglés), referido a las tecnologías que permiten procesar grandes cantidades de eventos para monitorear, guiar y optimizar el negocio en tiempo real, la Gestión del Desempeño Corporativo (Corporate Performance Management, CPM por sus siglas en inglés) que hace referencia a la medición del funcionamiento del proceso o la organización. Otros términos están vinculados con la gestión, como es el caso del Mejoramiento Continuo de Procesos (Continuous Process Improvement, CPI por sus siglas en inglés), el Mejoramiento de Procesos de Negocio (Business Process Improvement, BPI por sus siglas en inglés), la Gestión de Calidad Total (Total Quality Management, TQM por sus siglas en inglés) y Six Sigma. Las investigaciones en estas áreas tienen en común que los procesos son “puestos bajo el microscopio” con el objetivo de identificar posibles mejoras (AALST, 2012). La Minería de Procesos puede considerarse una tecnología que contribuye a cada una de las áreas antes mencionadas (CPM, BPI, TQM, Six Sigma) (AALST *et al.*, 2011; MEDEIROS, 2006).

En esta última década la información de los eventos de un proceso ha estado disponible y las técnicas de Minería de Procesos han madurado. Como ya se mencionó, las tendencias de gestión relacionadas al mejoramiento de procesos (por ejemplo Six Sigma, TQM, CPI, y CPM) y cumplimiento de normativas (ley Sarbanes-Oxley (SOX), Monitoreo de Actividades de Negocio (BAM)) se pueden beneficiar de la Minería de Procesos (AALST, 2012).

En la actualidad se puede contar con diversos sistemas académicos y comerciales, donde se han implementado los diferentes algoritmos o técnicas que se han desarrollado para la Minería de Procesos. Ejemplo de productos de software con capacidades de Minería de Procesos: ARIS Process Performance Manager (Software AG), Comprehend (Open Connect), Discovery Analyst (StereoLOGIC), Flow (Fourspark), FuturaReect (Futura Process Intelligence), Interstage Automated Process Discovery (Fujitsu), OKT Process Mining suite (Exeura), Process Discovery Focus (Iontas/Verint), Process Analyzer (QPR), ProM (TU/e), Rbminer/Dbminer (UPC), y Reectjone (Pallas Athena) (AALST, 2012).

1.5. Diagnóstico de procesos.

Según el diccionario de la real academia, **diagnóstico** (del griego *diagnostikós*, a su vez del prefijo *día-*, "a través", y *gnosis*, "conocimiento" o "apto para conocer"), alude al

Capítulo 1: Fundamentación teórica

análisis que se realiza para determinar cualquier situación y cuáles son las tendencias. Esta determinación se realiza sobre la base de datos, hechos recogidos y ordenados sistemáticamente, que permiten juzgar mejor qué es lo que está pasando y evaluar problemas de diversas naturalezas.

En el tema estudiado, Minería de Procesos, uno de los desafíos actuales lo constituye el diagnóstico del proceso. Área que abarca el análisis de rendimiento, detección de anomalías, identificación de patrones comunes y también puede ser útil al tratar con procesos desestructurados. El diagnóstico ayuda a tener una visión general del proceso, de los aspectos más significativos del mismo y de las técnicas que pueden ser más eficientes durante su análisis (BOSE, R.P.J.C. and AALST, 2012b).

Un diagnóstico realizado a un registro de eventos ayuda a responder preguntas como las que se enuncian a continuación:

1. ¿Cuál es el comportamiento más común que se ejecuta del proceso?

Dado un registro de eventos sería interesante conocer qué componentes del proceso son críticos o esenciales. Esos componentes esenciales forman la estructura principal del proceso y deberían ser conservados. Los esfuerzos de mejoras del proceso deben enfocarse a dichos componentes.

2. ¿Dónde las instancias del proceso se desvían y que tienen estas en común?

En la práctica, a menudo existe una brecha significativa entre el modelo prescrito o lo que se supone debe suceder, y lo que realmente sucede. Existe una necesidad de aumentar el diagnóstico del proceso con técnicas que pueden ayudar a encontrar desviaciones analizando las trazas en el registro de eventos. Hay muchas aplicaciones que tienen este requerimiento, por ejemplo, diagnóstico de fallas, detección de anomalías y diagnóstico de seguros fraudulentos. Dado un registro de eventos que contiene un conjunto de trazas que se generaron por funcionamiento normal del proceso y por mal funcionamiento del proceso, un análisis de estas trazas para encontrar desviaciones en trazas anómalas de trazas normales puede dar señales en el entendimiento de la anomalía o mal funcionamiento.

3. ¿Hay patrones comunes de ejecución en las trazas?

Un analista puede estar interesado en conocer cuando hay patrones de ejecución interesantes en el registro de eventos. La ausencia o presencia de estos patrones de

Capítulo 1: Fundamentación teórica

ejecución puede ser el indicador de una anomalía o una violación de seguridad o un malfuncionamiento.

4. ¿En qué contexto se ejecuta una actividad o un conjunto de actividades en un registro de eventos?

Existen dependencias entre las actividades en un proceso, la ejecución de actividades se espera que suceda en ciertos contextos. Puede haber dependencias de corto y largo rango entre actividades, las dependencias de largo rango son difíciles de descubrir. Un analista puede estar interesado en entender los contextos de ejecución de actividades y/o secuencias de actividades.

5. ¿Cuáles son las instancias del proceso que comparten/capturan un comportamiento deseado?

Comúnmente en el diagnóstico un analista puede estar interesado en encontrar instancias del proceso que comparten un comportamiento deseado. El comportamiento deseado puede ser expresado como una manifestación de un patrón de secuencia de actividades o alguna combinación compleja de estos patrones. Aunque los enfoques temporales lógicos pueden ayudar a abordar este problema en cierta medida, mediante el descubrimiento de casos de procesos que capturan el comportamiento deseado exactamente, uno también podría estar interesado en el descubrimiento de instancias de procesos que comparten el comportamiento deseado aproximadamente.

6. ¿Hay algunos patrones particulares en el proceso?

Los patrones de flujo de trabajo se refieren a las formas y estructuras recurrentes que abordan los requerimientos del negocio. Por ejemplo, los hitos indican especificaciones de puntos de ejecución en el modelo de proceso y proporcionan un mecanismo para apoyar la ejecución condicional de una tarea o sub-proceso. Un analista estaría interesado en descubrir la presencia y el análisis de los patrones de los hitos en el proceso de registro de eventos. De manera similar, el descubrimiento de los modelos de procesos con simultaneidad es uno de los problemas más difíciles en la Minería de Procesos. La presencia de actividades concurrentes crea permutaciones diferentes de actividades en el registro de eventos que se suma a la complejidad de los algoritmos de descubrimiento. La detección de la presencia de actividades simultáneas también podría ayudar en el pre-procesamiento de los registros.

1.6. Trabajos relacionados sobre diagnóstico de procesos.

En el área del diagnóstico de procesos se han desarrollado un conjunto de técnicas, entre las cuales se pueden mencionar:

Chequeo de conformidad.

El Chequeo de conformidad es un análisis que permite saber hasta qué punto, el registro de eventos se corresponde con el modelo de procesos y viceversa. Esta técnica (AALST, 2012; ROZINAT, A. and AALST, 2008) compara un modelo teórico inicial con el comportamiento observado en las trazas y apunta a la detección de inconsistencias y/o desviaciones entre el modelo del proceso y su registro de eventos.

Esta técnica tiene inherentes limitaciones en su aplicación especialmente en el diagnóstico de procesos. Primeramente asume que hay un modelo del proceso, cuando en la realidad el modelo puede no estar presente o puede contar con problemas. Se puede argumentar que el modelo puede ser descubierto a partir del registro de eventos y con este se podría hacer el Chequeo de conformidad. Sin embargo, este enfoque no es adecuado para el análisis de los procesos altamente complejos y/o flexibles, la clase de modelos que representan la mayoría de los registros de eventos de la vida real y que culminan en el descubrimiento de modelos estereotipados como "espagueti".

Incluso en los casos donde el modelo del proceso está disponible como red de Petri, es muy difícil mirar adentro del proceso para identificar y localizar problemas especialmente en el caso de estos que sean muy grandes.

Stream Scope Visualization.

Es una técnica de visualización de trazas que está basada en las correlaciones de clases de eventos. Usando Corriente de Visualización y Alcance (Stream Scope Visualization) los patrones de los eventos Co-ocurrentes pueden ser detectados por su cercanía. Sin embargo, Stream Scope Visualization está limitado en el hecho de que visualiza cada traza por separado y no provee una visión general del registro de eventos (GÜNTHER, 2009).

Minería difusa (en inglés Fuzzy Miner).

La Minería difusa es el primer intento de dejar a un lado la suposición de que las actividades se producen en el mismo nivel de abstracción.

Capítulo 1: Fundamentación teórica

Esta produce un modelo basado en grafos, donde se utilizan dos tipos de nodos, los nodos que se refieren a una actividad, y los nodos que se refieren a un conjunto de actividades, o clústeres. Por lo tanto, el modelo es capaz de proporcionar una vista de alto nivel de un proceso abstrayendo detalles no deseados.

Sin embargo, la Minería difusa todavía supone que cada evento en el registro de eventos pertenece a uno de estos nodos, es decir, hay una relación de uno a muchos ya que cada nodo puede representar muchas actividades, pero cada actividad está representada por un nodo exactamente.

Minería basada en teoría de regiones.

En (AALST *et al.*, 2009) se presenta una técnica donde las relaciones entre nodos y actividades es realmente de muchos a muchos.

Inicialmente esta teoría solo podía ser aplicada a un reducido conjunto de sistemas de transición. Sin embargo, al transcurrir el tiempo, este enfoque fue extendido para permitir la síntesis de cualquier sistema de transición finito.

Es necesario tener en cuenta, que solo un grupo reducido de personas han aplicado las regiones basadas en la Minería de Procesos, aunque recientemente algunos trabajos acerca de la teoría de regiones basadas en lenguaje han aparecido (BERGENTHUM *et al.*, 2007), y en este mismo trabajo se muestra como esta teoría se puede aplicar a la Minería de Procesos. Estos enfoques son muy interesantes y construyen directamente una red de Petri y no construyen un sistema de transición intermedio.

Esta técnica, primeramente construye un espacio utilizando el registro de eventos y luego usa la teoría de regiones para construir una red de Petri a partir del espacio creado anteriormente. Este enfoque está limitado en dos sentidos, primeramente solo considera redes de Petri como modelos del proceso, aunque se sabe que hay muchos otros modelos tales como redes de flujos de trabajo o redes elementales. Al aplicar esta teoría se puede generar grandes tráficó de información (cuello de botella).

Agrupación difusa (en inglés Fuzzy Clustering).

La agrupación difusa, otro de los enfoques en el diagnóstico de procesos, es un algoritmo para agrupar los eventos automáticamente a un nivel deseado de la abstracción (DONGEN, B. F. and ADRIANSYAH, 2010). Los diagramas obtenidos en

Capítulo 1: Fundamentación teórica

este trabajo no permiten conocer las relaciones existentes entre las diferentes actividades que conforman el clúster.

Alineación de trazas (en inglés Trace Alignment).

La alineación de trazas alivia el problema de determinar patrones interesantes en medianos y grandes registros, determinándolos automáticamente y mostrándolos al usuario (BOSE, R. P. JAGADEESH CHANDRA and AALST, 2012a).

La alineación de trazas es una técnica que utiliza la programación dinámica para tabular las trazas de forma tal que se simplifiquen los problemas de entendimiento de las relaciones entre las actividades. Está inspirada en la alineación de múltiples secuencias (MSA, por sus siglas en inglés), herramienta fundamental en el área de la bioinformática pues ayuda a desentrañar las estructuras secundarias y terciarias de las proteínas y moléculas, la evolución y funciones de estas, y a inferir las relaciones taxonómicas y filogenéticas entre los organismos y para el diagnóstico de enfermedades genéticas. Sin embargo se consideró un desafío adaptar esta última técnica a la alineación de trazas.

Este es un tema muy poco tratado en estos días aunque es conocida su utilidad en el diagnóstico de procesos, resolviendo muchas interrogantes y apoyando a otras herramientas que trabajan en este sentido. Una de las aplicaciones de la alineación de trazas es el descubrimiento de las desviaciones entre las trazas. Esta técnica también permite la visualización de múltiples trazas a la vez y es capaz de descubrir los patrones comunes de ejecución dentro y a través del registro de eventos.

Análisis de diagrama de puntos.

El Análisis de diagrama de puntos (DCA, por sus siglas en inglés), es una de las técnicas de visualización del registro de eventos más usada. El DCA, análogo a los diagramas de Gantt, ayuda en el análisis de rendimiento del proceso representando los eventos del proceso de una forma gráfica, como puntos en un plano donde una dimensión hace referencia a los casos en el registro de eventos, y la otra dimensión se refiere al tiempo en el que se ejecutaron estos. Se utiliza un color distinto para cada actividad que se encuentre almacenada en el registro de eventos. Se puede notar que con esta forma de graficar los puntos o eventos no se encuentran alineados por lo tanto es muy difícil determinar patrones comunes entre casos distintos.

Capítulo 1: Fundamentación teórica

El DCA calcula algunos indicadores de desempeño, tales como el intervalo de mínimo, máximo y promedio entre eventos. Un analista de negocios necesita investigar manualmente el DCA para identificar los problemas potenciales de rendimiento. Para un registro de eventos con un gran número de actividades (del orden de unas pocas decenas hasta cientos), la inspección manual y la comprensión de la tabla de puntos se vuelve muy complicada y con frecuencia es imposible identificar patrones interesantes.

Tandem Arrays.

Una rama fundamental en el diagnóstico del proceso es el descubrimiento de patrones comunes de ejecución. Bose y Van Der Aalst (BOSE, R.P. JAGADEESH CHANDRA and AALST, 2009) proponen el uso de Tandem Arrays y repeticiones máximas para capturar patrones recurrentes dentro y a través de las trazas, aunque al utilizar esta técnica, estos patrones se pueden determinar en tiempo lineal, ellos sufren dos limitaciones: primeramente el número de patrones descubiertos puede ser realmente grande, sin embargo usando métricas así como las propuestas en (LI *et al.*, 2010) se pueden filtrar patrones insignificantes y por último, los patrones descubiertos son atómicos y las dependencias/correlaciones entre los patrones deben ser descubiertas por separado; en otras palabras, el contexto de sus manifestaciones se pierde y necesita ser restablecido por separado.

Descubrimiento de patrones usando árbol de sufijos.

Las repeticiones que existen entre las trazas en el registro de eventos se puede determinar aplicando los algoritmos de identificación de patrones en las secuencias obtenidas, al concatenar las trazas en el registro de eventos con un delimitador distinto de cualquier elemento en el alfabeto original, es decir, las actividades del registro de eventos. Tal concatenación de trazas puede incurrir en una secuencia muy larga, para esto se puede adoptar una construcción de árbol de sufijos eficiente como para manejar secuencias muy grandes (CHEUNG, 2005). Repeticiones aproximadas pueden ser encontradas primeramente identificando repeticiones exactas y luego buscar todas las sub-secuencias a una distancia determinada de las repeticiones exactas.

Conclusiones.

Capítulo 1: Fundamentación teórica

Después de haber realizado un profundo estudio sobre las diferentes técnicas de diagnóstico, se llegó a la conclusión que presentan problemas para caracterizar un proceso de negocio. Esto conlleva a que en ocasiones sea complicado comprender el proceso ejecutado.

Debido a esta situación se está desarrollando en la Universidad de las Ciencias Informáticas UCI, un algoritmo para el diagnóstico, que permite construir un árbol de matrices representativas de los subprocesos que componen al proceso analizado. Para su desarrollo se hace uso de la alineación de trazas, técnica propuesta por Bose y Van der Aalst en (BOSE, R.P.J.C. and AALST, 2012b). Este algoritmo permite agrupar las tareas automáticamente y resalta los aspectos más significativos del proceso en cada momento. El descubrimiento de los subprocesos que componen al proceso analizado, sus dependencias y correlaciones permiten una mayor precisión en el diagnóstico realizado. La herramienta ProD define tres pasos fundamentales:

1- Alinear las trazas: El primer paso es el agrupamiento de las trazas y la alineación de las mismas. Las trazas alineadas constituyen una representación de las actividades de acuerdo a un orden relativo y una estructuración de las mismas. El orden establecido entre las actividades permite identificar los patrones de flujo de control que se manifiestan entre los subprocesos. Como resultado de la alineación de un grupo de trazas se obtiene una matriz A.

2- Pre-procesar las trazas alineadas: A partir de las trazas alineadas se pueden determinar los casos incompletos (casos que no terminan con las actividades finales identificadas para el proceso). Estos casos pueden ser tratados o eliminados según se considere, así se puede volver a alinear las trazas.

3- Determinar el árbol de bloques de construcción: En este paso se construye a partir de la matriz A un árbol de bloques de construcción que representan los subprocesos que componen al proceso analizado.

Este algoritmo tiene como objetivo construir un árbol de bloques de construcción que representa la descomposición del proceso analizado. La Figura 2 muestra un ejemplo de un árbol de bloques de construcción. Los bloques de construcción C_{A}^2 y C_{A}^3 representan subprocesos ordenados secuencialmente, C_{A}^4 y C_{A}^5 representan subprocesos ordenados como opciones de una selección. C_{A}^6 y C_{A}^7 representan subprocesos en paralelo.

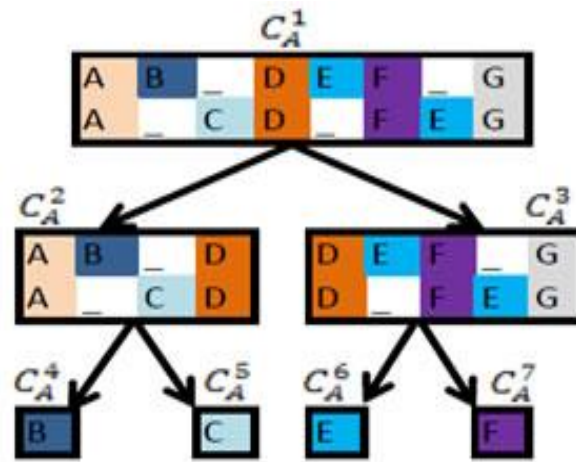


Figura 2: Árbol de bloques de construcción

1.7. Descubrimiento.

Según el diccionario de la real academia se puede decir que descubrimiento hace referencia a un hallazgo, encuentro, manifestación de lo que estaba oculto, secreto o era desconocido (_ACADEMIA, 2010).

Proceso de descubrimiento.

El proceso de descubrimiento es una de las tareas más difícil y destacada en la Minería de Procesos (AALST, 2012). Está basada en la construcción de un modelo de proceso a partir de la información obtenida de un registro de eventos al ejecutar un proceso o instancia de él (AALST, 2011; AALST *et al.*, 2011; AALST, 2012; COOK, 1996), proveniente de determinados sistemas de información, por ejemplo: sistemas de Planificación de los Recursos Empresariales (ERP por sus siglas en inglés) y sistemas de Gestión de la Relación con los Clientes (CRM por sus siglas en inglés).

Para el descubrimiento del modelo de procesos se han desarrollado diferentes técnicas capaces de realizar el descubrimiento sin ninguna información a-priori, meramente basado en la información obtenida de los registros de eventos. Los modelos descubiertos pueden ser (por ejemplo: Redes Petri, BPMN, o un diagrama de actividades UML⁵), aunque también podría describir otras perspectivas (ejemplo una red social⁶).

⁵UML: Lenguaje de Modelado Unificado.

⁶Las redes sociales permiten visualizar el comportamiento de los recursos, sus métricas y su gobernabilidad.

Capítulo 1: Fundamentación teórica

Trabajos relacionados sobre descubrimiento de procesos.

A continuación se muestra una visión general de los enfoques más relevantes en esta área.

Alpha.

Es un algoritmo para el descubrimiento en la Minería de Procesos, destinado a la reconstrucción de la causalidad de un conjunto de secuencias de eventos. Fue presentada por primera vez por Van der Aalst, Weijter y Mărușter. Esta técnica recibe un registro de flujo de trabajo como entrada y los resultados se van construyendo en una red de flujo de trabajo. Esto se hace mediante el examen de las relaciones causales observadas entre las tareas encontradas en el registro (AALST, 2011; MEDEIROS, 2006; VAN DER AALST, 2003).

El algoritmo Alpha ilustra muy bien algunas de las ideas principales detrás del proceso de descubrimiento. Es simple y muchas de sus ideas se han incorporado en algoritmos más complejos. No posee problema en el descubrimiento de modelos que presenten bucles de longitud igual o mayor que tres, patrones básicos de flujo de trabajo como secuencia, paralelismo, entre otros.

Limitaciones:

- Tiene problemas al tratar con bucles de longitud uno o dos.
- Las frecuencias no son tomadas en cuenta.
- Es muy sensible al ruido.
- No descubre el constructor (non-free-choise (sin-libre-elección) de tipo no local).
- No se tiene garantía de que se produzca un modelo correcto cuando el proceso subyacente que puede ser descrito por un WF-red⁷ contiene actividades duplicadas (dos transiciones con la etiqueta de la misma actividad) y las transiciones en silencio o tareas invisibles (actividades que no se guardan en el registro de eventos).

Por ser un algoritmo sencillo y donde se aplica muy bien la mayoría de las ideas del descubrimiento de un modelo de proceso, ha servido como base para el desarrollo de otros algoritmos más fuertes o extensiones como Alpha⁺ y Alpha⁺⁺, donde algunas de sus limitaciones han sido resueltas. Por ejemplo el Alpha⁺ soluciona la limitación del descubrimiento de lazos de longitud uno y dos (A.K.A. DE MEDEIROS, 2004; WIL M.P 2011). En el caso de Alpha⁺⁺ (desarrollado por Wen et al, 2007) introdujo mejoras para

⁷ WF-red: Este término se refiere a red de flujo de trabajo (workflow).

Capítulo 1: Fundamentación teórica

el tratamiento del constructor sin-libre-selección (en inglés non-free-choice).

Minería heurística.

La minería heurística (HM) es una práctica aplicable a la Minería de Procesos, este algoritmo es capaz de tratar con el ruido y la baja frecuencia, además es poco sensibles al carácter incompleto de los registros, y puede utilizarse para expresar el comportamiento principal (es decir, no todos los detalles y excepciones) almacenado en un registro de eventos. La minería heurística también se puede ver como una extensión del algoritmo Alpha (A.J.M.M. WEIJTERS, W.M.P. VAN DER AALST, AND A.K. ALVES DE MEDEIROS, 2003b; A.J.M.M. WEIJTERS, W.M.P. VAN DER AALST, 2003a; AALST, 2011; MEDEIROS, 2006).

Los algoritmos heurísticos en la Minería de Procesos utilizan una representación similar a las redes causales. Por otra parte, estos algoritmos toman en cuenta las frecuencias y secuencias de eventos al construir un modelo de proceso. La idea básica es que las rutas poco frecuentes no deben ser incorporadas en el modelo. Tanto la representación parcial proporcionada por las redes causales y el uso de frecuencias hace que el enfoque sea mucho más robusto que otras técnicas desarrolladas (AALST, 2011).

Limitaciones:

- No es capaz de descubrir dependencia a larga distancia no locales.
- No puede tratar con situaciones donde estén presentes bucles cortos, del tipo de las relaciones de dependencia (división AND/XOR-/ Join), tiene problemas con las actividades invisibles que se presenten en Split/Join (AALST, 2011).

Múltiples fases (Multi-phase miner).

Estos autores (B.F. VAN DONGEN, AND W.M.P. VAN DER AALST, 2005) introducen como resultado de la minería de las trazas, un modelo en cadenas de procesos, controladas por eventos EPCs (siglas de Event-driven Process Chains⁸). Este algoritmo se estructura en dos pasos, en el primero de estos se obtiene un modelo de procesos por cada una de las trazas analizadas. En este paso se asegura que cada actividad aparezca solo una vez en la traza, a cada instancia de una tarea

⁸ EPCs: Significa en español cadenas de procesos impulsado por eventos.

Capítulo 1: Fundamentación teórica

en la traza se le asigna un identificador único. Ya en el segundo paso se mezclan los modelos obtenidos en el paso anterior y se definen relaciones de división y unión en sus diferentes variantes: AND, OR y XOR (DONGEN, B.F. VAN and AALST, 2004). Este algoritmo está basado en Alpha (B.F. VAN DONGEN, AND W.M.P. VAN DER AALST, 2005).

Limitaciones:

- Presenta problemas con la duplicidad de tareas, tareas invisibles, los registros de eventos que estén afectados por ruido y las dependencias que no son locales (AALST, 2011).

Algoritmos genéticos.

Los algoritmos genéticos son métodos adaptativos de búsqueda que tratan de imitar el proceso de la evolución.(AALST, 2011); presentan 4 pasos: (a) la inicialización,(b) la selección, (c) la reproducción, y(d) la terminación (AALST, 2011; MEDEIROS, 2006).

Estos algoritmos comienzan con una población inicial de los individuos. A cada individuo se le asigna la medida de su estado de forma, para indicar su calidad. En este caso, un individuo es un modelo de proceso posible y el estado de forma es una función la cual evalúa qué tan bien un individuo es capaz de reproducir el comportamiento en el registro.

Con este enfoque se desarrollaron dos algoritmos el GA y el DGA. El GA provee al proceso de descubrimiento de Minería de Procesos la extracción de modelos con las construcciones estructurales (como secuencias, selección, paralelismo, bucles estructurados, no-local sin-libre-elección y las tareas invisibles). En el caso del DGA, es una extensión del algoritmo genético básico (GA), que permite descubrir todas las construcciones estructurales, incluyendo la duplicidad de actividades. Ambas técnicas pueden tratar con registros incompletos y ser robustas al ruido.

Nota: Los algoritmos genéticos se utilizan para mejorar un modelo de proceso obtenido mediante la técnica minería heurística.

Limitaciones:

Capítulo 1: Fundamentación teórica

- No se puede descuidar el tiempo de cómputo. La situación es más crítica para el DGA, debido a que utiliza un espacio de búsqueda más grande, por lo tanto, necesita más iteraciones (ejecutar más generaciones) para converger a soluciones buenas (ejemplo, modelos de procesos que son completos, precisos y plegado).
- Estas técnicas no son eficientes para extraer grandes modelos. Se pueden tomar un largo tiempo para descubrirlo.
- Poseen dificultad al extraer modelos con construcciones que permiten muchas situaciones de entrelazado.
- El DGA sólo puede capturar modelos en los que los duplicados no comparten elementos de entrada/salida
 1. Ejemplo: construcciones paralelas con más de cinco ramas.
 2. Los puntos AND-Split/Join desequilibrados, especialmente cuando se combinan con largas ramas en paralelo.
 3. Más de tres bucles de longitud dos conectados al mismo punto.

Una vez más, la situación es peor para el DGA porque las intercalaciones tienden a conducir a la consecuencia de más duplicados de tareas y en consecuencia, el DGA requiere más iteraciones para converger a soluciones buenas que el GA (MEDEIROS, 2006).

Fuzzy Miner (Minería difusa).

Este enfoque proporciona un conjunto ampliable de los parámetros para determinar qué actividades y arcos deben ser incluidos. Además, puede construir modelos jerárquicos, es decir, las actividades menos frecuentes se pueden mover a los subprocesos (B.F. VAN DONGEN, N. BUSI, G.M. PINNA, ANDW.M.P. VAN DER AALST. , 2007).

Günther et al en este enfoque dirige su atención al análisis de las trazas que reflejan el comportamiento de procesos poco estructurados. En consideración se especifica que se deben resaltar en estos casos, los aspectos importantes del comportamiento analizado y ocultar los que pueden no ser significativos. Para ello se desarrollaron dos métricas que guían el proceso de descubrimiento del modelo de procesos. La primera de las métricas es *significación*, que permite medir la importancia relativa que tiene un evento y las relaciones de precedencia binarias que existen con respecto a él, esta especifica el nivel de interés que se puede tener sobre un evento o la ocurrencia de

Capítulo 1: Fundamentación teórica

este después de otro; un aspecto a considerar puede ser la frecuencia de aparición de un evento, mientras mayor es la frecuencia, mayor es la significación del mismo. La otra métrica es la *correlación*, determinada por la medida en la que una relación de precedencia entre los eventos puede ser relevante. Permite tener una medida de cuan estrechamente relacionados están dos eventos. Basado en estas dos métricas, el algoritmo estructura el modelo de procesos, incorporando los conceptos de registro de filtrado, es decir, la eliminación de los eventos menos significativos de los registros, siendo capaz de deducir y extraer la estructura de lo que es caótico al limpiar una gran cantidad de comportamiento confuso (B.F. VAN DONGEN, N. BUSI, G.M. PINNA, ANDW.M.P. VAN DER AALST. , 2007; GOEDERTIER, 2008).

El algoritmo es robusto ante el ruido, posee carácter interactivo y de exploración, distinguiéndose de otras técnicas de minería de anteriores procesos.

Limitaciones:

- La capacidad de configurabilidad y sus métricas hace este enfoque universal, siendo también una de sus debilidades (B.F. VAN DONGEN, N. BUSI, G.M. PINNA, ANDW.M.P. VAN DER AALST. , 2007).

✚ Minería basada en Región.

Los métodos de síntesis basados en región son utilizados para construir Redes de Petri de un sistema en transición. Para ello se consideran las trazas como un lenguaje finito y el modelo obtenido es una red minimal que recoge el comportamiento del lenguaje dado (AALST, 2011; R. YZQUIERDO-HERRERA, 2011). Esta técnica trata de evitar el comportamiento adicional, reflejando de la manera más precisa las huellas reflejadas en las trazas. Para aplicar el estado basado en las regiones, se necesita primero crear un sistema de transición. Después de la transformación de un registro de eventos en un sistema de transición de bajo nivel, se puede sintetizar una red de Petri, a su vez, esta red de Petri se puede utilizar para construir un modelo de procesos en otros de alto nivel de denotación (por ejemplo, BPMN, diagramas de actividad UML, y EPCs). El desafío consiste en doblar un sistema de transición de gran tamaño en una pequeña red de Petri mediante la detección de simultaneidad. La idea central es descubrir las regiones que corresponden a los lugares. La característica principal

Capítulo 1: Fundamentación teórica

de este enfoque es que puede ser utilizado para descubrir las más complejas estructuras de control de flujo.

El algoritmo puede ser aplicado iterativamente, reduciendo así la complejidad del espacio, que es comúnmente aceptado para ser el cuello de botella para muchas técnicas de Minería de Procesos (B.F. VAN DONGEN, N. BUSI, G.M. PINNA, ANDW.M.P. VAN DER AALST. , 2007).

Limitantes:

- El algoritmo no es robusto ante el ruido.
- Presenta problema con el rendimiento. (B.F. VAN DONGEN, N. BUSI, G.M. PINNA, ANDW.M.P. VAN DER AALST. , 2007)

Conclusiones del capítulo.

En este capítulo se explicaron los diferentes conceptos asociados a una nueva disciplina llamada Minería de Procesos, en inglés Process Mining, método para obtener el modelo de un proceso estructurado, a partir de un conjunto de ejecuciones reales obtenidas de un registro de eventos que contiene información sobre el orden en que se realizaron las actividades, el ejecutor, tiempo de duración y otras informaciones proporcionadas por un sistema de información.

Dentro de otros aspectos mencionados en este capítulo se puede encontrar las diferentes técnicas para el descubrimiento o extracción de un modelo de proceso a partir de un registro de eventos y sus limitaciones, además de algunos algoritmos que se han desarrollado para diagnosticar un proceso, entre los cuales se encuentran: Alineación de trazas, Agrupación difusa y ProD.

De acuerdo al estudio realizado sobre las diferentes técnicas de diagnóstico, se pudo apreciar que estas no posibilitan una correcta caracterización del proceso, o al menos no definen todas las características del proceso que se necesitan para facilitar la posterior selección de las técnicas de descubrimiento y entendimiento de procesos. Entre el grupo de técnicas estudiadas, la que más se destaca es ProD, que permite visualizar los diferentes aspectos que posea el proceso que se esté analizando. También se pudo percibir que los algoritmos de descubrimiento se basan en algunos patrones de control de flujo como: secuencia, paralelismo, selección, sin libre elección y lazos. Además tienen en cuenta la ausencia de información o tareas invisibles en los

Capítulo 1: Fundamentación teórica

registros de eventos, tareas duplicadas y por último la presencia de ruido en los registros de eventos.

Capítulo 2: Propuesta de solución.

Introducción.

En este capítulo se abordará sobre los diferentes aspectos que caracterizan un proceso de negocio. Se hará una evaluación de las diferentes técnicas de diagnóstico y descubrimiento de procesos de acuerdo a los aspectos que lo caracterizan y finalmente se tratará la propuesta de solución.

2. Propuesta de solución.

En el capítulo anterior se realizó un análisis de las técnicas de descubrimiento y diagnóstico, para entender cuáles podrían ser las características más adecuadas para el análisis de un proceso de negocio.

En el estudio realizado anteriormente, se evidenció que ninguna de las técnicas de diagnóstico realiza una evaluación integral del proceso; por tanto existe la necesidad de realizar una propuesta, donde la técnica seleccionada permita efectuar una correcta caracterización del mismo y a partir de esas características hacer la sugerencia de la técnica de descubrimiento más efectiva en el análisis del proceso.

Inicialmente para la realización de esta propuesta es necesario conocer cuáles son los aspectos que se definieron para caracterizar un proceso.

2.1 Aspectos que caracterizan un proceso de negocio.

En la Minería de Procesos existen tres perspectivas para analizar un proceso, teniendo en cuenta los flujos de control, los recursos (organizacional) y por último los casos. Se escoge una o varias perspectivas a analizar en dependencia de la información contenida en las trazas. La más común en cuanto a su uso es la de control de flujo y es la que posibilita determinar cómo se organizan las tareas que dan lugar al proceso. Desde esta perspectiva se debe tener en cuenta la información relacionada con la identificación del caso o instancia ejecutada, identificación de la tarea y el tiempo en que se produjo la misma. Desde la perspectiva del recurso se debe tener en cuenta la información referente al usuario del sistema que ejecutó la tarea, y es posible determinar la estructura organizacional con respecto a la cual se realizan las diferentes tareas. Lo antes expuesto permite conocer las jerarquías que existen entre roles y usuarios durante el proceso. En la tercera perspectiva, la cual está relacionada con los casos, se debe tener en cuenta la información en detalle de las tareas, como la edad y el sexo de los usuarios implicados. Desde esta perspectiva se pueden descubrir casos excepcionales, como la probabilidad de que personas de determinado sexo realicen en tiempo determinada tarea.

Capítulo 2: Propuesta de solución

Teniendo en cuenta la perspectiva de flujo de trabajo antes mencionada y la revisión de las diferentes técnicas de descubrimiento se definen las siguientes características para analizar un proceso:

- Secuencia.
- Paralelismo.
- Selección.
- Lazos.
- Sin libre elección.
- Tareas invisibles.
- Duplicidad de información o Tareas duplicadas.

Adicionalmente es necesario fijarse en la capacidad de las técnicas para tratar con registros de eventos afectados por el ruido.

Secuencia: Es el patrón básico encontrado en un registro de eventos. Se evidencia cuando hay una dependencia entre dos actividades, de tal forma que una actividad no pueda iniciarse antes de que otra haya terminado (una actividad será habilitada, solo hasta que la actividad anterior sea ejecutada).

Distribución en paralelo o paralelismo: La distribución en paralelo es necesaria cuando dos o más actividades deben ejecutarse de forma concurrente o en paralelo, permitiendo la ejecución de actividades simultáneamente y en cualquier orden.

Selección exclusiva: Ocurre cuando en un punto del flujo de trabajo se escoge sólo una de varias ramas del proceso (XOR-Split, OR exclusiva-Split).

Lazos: Este patrón se ejemplifica cuando en el proceso una o más actividades pueden ser ejecutadas repetidamente.

Sincronización: Una actividad es iniciada cuando dos o más hilos completan la ejecución de sus actividades. En un punto del proceso, dos o más hilos de control convergen en un solo hilo (And- Join).

Sin libre elección: Es una mezcla de la sincronización y la elección donde se evidencia una dependencia de larga distancia.

Tareas invisibles: Las tareas invisibles corresponden a pasos silenciosos que se utilizan para fines solamente de enrutamiento, por lo tanto, ellos no están presentes en los registros de eventos

Tareas duplicadas: La duplicación de tareas se refiere a situaciones en las que múltiples tareas en el proceso tienen la misma etiqueta. Los duplicados son

Capítulo 2: Propuesta de solución

normalmente en contextos diferentes (alrededor de las tareas) en un proceso (MEDEIROS, 2006).

Ruido: El ruido puede aparecer en dos situaciones: alguna de las huellas de eventos se registran de manera incorrecta (por ejemplo, debido a la temporal configuración del sistema) o situaciones excepcionales.

2.2. Evaluación de las técnicas de descubrimiento de acuerdo a los aspectos que caracterizan un proceso.

A partir de las técnicas de descubrimiento definidas en el capítulo uno y la descripción de los aspectos que caracterizan un proceso, en esta sección se hace un análisis para conocer cómo estas técnicas responden ante las diferentes características.

En la Tabla 1 se muestra una evaluación a partir de un grupo de técnicas desarrolladas para el descubrimiento de procesos. En la parte lateral izquierda se encuentran los aspectos que caracterizan un proceso y en la parte superior los algoritmos de descubrimiento, los cuales serán evaluados en dependencia de si cubren parcial, total o si no cubren el aspecto seleccionado. La evaluación dada estará entre 0 y 1, en dependencia del cumplimiento del parámetro en ese algoritmo. Los aspectos a evaluar son: *secuencia*, *paralelismo*, *lazos*, *selección*, *si libre elección*, las 8 situaciones donde se pueden presentar las *tareas invisibles* (ver Tabla 10), *tareas duplicadas*, *la capacidad de los algoritmos para trabajar con ruido* y *si el procesos es o no desestructurado*.

Tabla 1: Caracterización de las técnicas de descubrimiento.

Aspectos	Alpha	Alpha+	Alpha++	DWS	Heuristic Miner	Multi phases	Schimm	ILP Miner
Secuencias	1	1	1	1	1	1	1	1
Paralelismo	1	1	1	1	1	1	1	1
Lazos - > 3 - < 3	0.5	1	1	1	0.5	1	1	1
Selección	1	1	1	1	1	1	1	1
Sin libre elección -local			1					1

Capítulo 2: Propuesta de solución

-no local	0.5	0.5		0.5	0.5	0.5	0.5	
Tareas invisibles	0.125	0	0.25	0.375	0.25	0.50	0.125	0.50
Tareas duplicadas	0	0	0	1	0	0	0	1
Robusto al ruido	0	0	0	0	1	0	1	0
Proceso desestructurado	0	0	0	0	0	0	0	0
Total	4.125	4.50	5.25	5.835	5.25	5.0	5.625	6.50

Aspectos	Región Miner	AGNEs	Goedertier	Fuzzy Miner	Splitpar	Genetic Miner	Markov	Transitions Systems
Secuencias	1	1	1	1	1	1	1	1
Paralelismo	1	1	1	1	1	1	1	1
Lazos - > 3 - < 3	1	1	1	1	1	1	1	1
Selección	1	1	1	1	1	1	1	1
Sin libre elección -local -no local	1	1	1	1	1	1	0.5	1
Tareas invisibles	0.25	0.125	0.125	0.625	0.25	0.75	0.25	0.375
Tareas duplicadas	1	1	0	1	1	1		1
Robusto al ruido		1	1	1	1	1	1	0
Proceso desestructurado	0	0	0	1	0	0	0	0
Total	6.25	7.125	6.125	8.625	7.25	7.75	5.75	6.375

Teniendo en cuenta los aspectos que se definieron para la caracterización de un proceso y los diferentes algoritmos existentes en el descubrimiento se realizó una evaluación con el objetivo de construir el vector característico de cada una de las técnicas, que sirven como entrada a la propuesta de solución.

Capítulo 2: Propuesta de solución

2.3. Evaluación de las técnicas de diagnóstico de acuerdo a los aspectos que caracterizan un proceso.

En esta sección se hace un análisis sobre las diferentes técnicas para el diagnóstico de procesos a partir de los aspectos que caracterizan un proceso.

En la investigación se hizo mención de algunas técnicas de diagnósticos, de las cuales no todas estarán evaluadas según los aspectos que caracterizan un proceso, debido a que no permiten la identificación de patrones de control de flujo o no están desarrolladas para realizar este tipo de tareas, sino que se encargan de ofrecer otros datos para el análisis del proceso. Estas técnicas son las que se enuncian a continuación: Análisis de Diagrama por Puntos (Dotted Chart Analysis) y Corriente de Visualización y Alcance (Stream Scope Visualization).

Tabla 2: Evaluación de las técnicas de diagnóstico.

	Tendem Arrays	Agrupación Difusa	Alineación de trazas	Minería Difusa	Chequeo de conformidad	ProD
Secuencias		*	*			*
Paralelismo	*	*	*	*		*
Lazos	*					*
Selección	*	*	*			*
Sin libre elección						
Tareas invisibles		*	*		*	*
Tareas duplicadas					*	*
Ruido		*				*

Capítulo 2: Propuesta de solución

Proceso desestructurado	*	*				*
-------------------------	---	---	--	--	--	---

A partir de la Tabla 2, se pudo identificar que ProD es la técnica que permite detectar mayor cantidad de parámetros, ayudando a tener una visión general de los aspectos más significativos de un proceso a partir de la información contenida en un registro de eventos.

2.4. Detención de las características del proceso analizado.

Para poder sugerir cuales son los algoritmos más adecuados para el descubrimiento, es necesario un mecanismo que identifique la manifestación de los patrones representativos de dichas situaciones en las trazas. A partir de la identificación de los diferentes aspectos se puede hacer la sugerencia correspondiente.

La detección de patrones comunes es parte del diagnóstico del proceso (BOSE, R.P.J.C. and AALST, 2012b), para el apoyo de la investigación se utilizó el algoritmo ProD.

El algoritmo ProD permite la detención de las diferentes características de un proceso de negocio a través de un árbol de bloques de construcción, el cual representa la descomposición del proceso analizado. A partir de esta descomposición, es capaz de mostrar alguno de los aspectos que caracterizan al proceso en análisis (secuencia, paralelismo, lazos, selección (XOR, OR) y secuencia oculta), como se muestra en la Figura 3.

Capítulo 2: Propuesta de solución

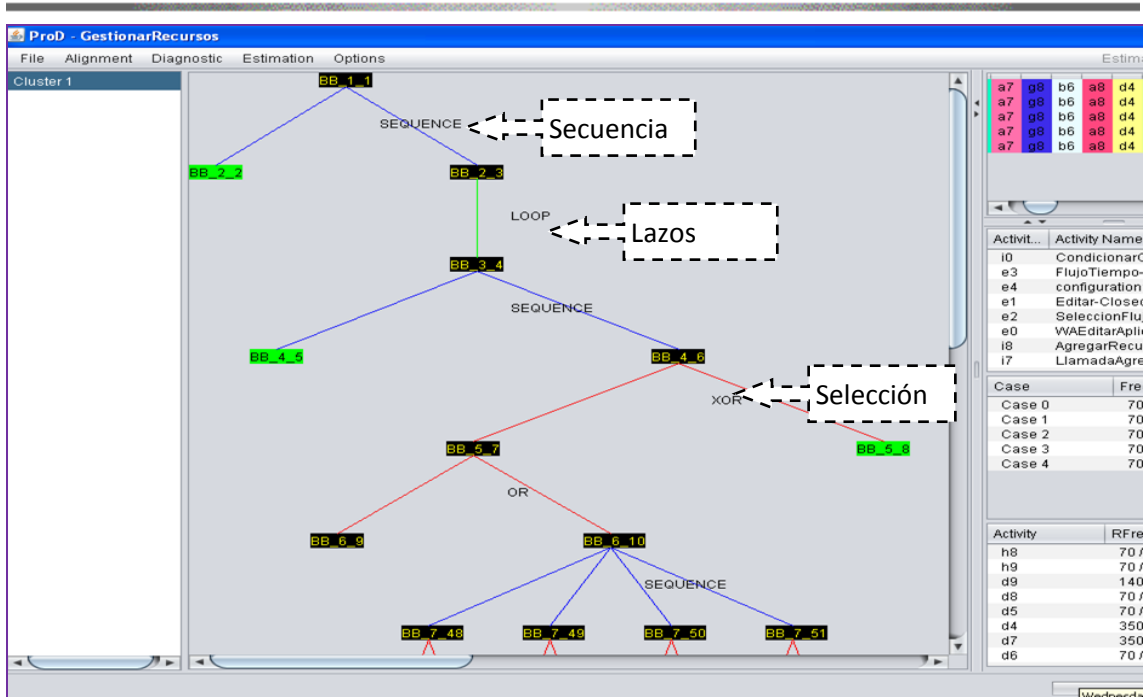


Figura 3: Árbol de construcción con los aspectos que caracterizan al proceso.

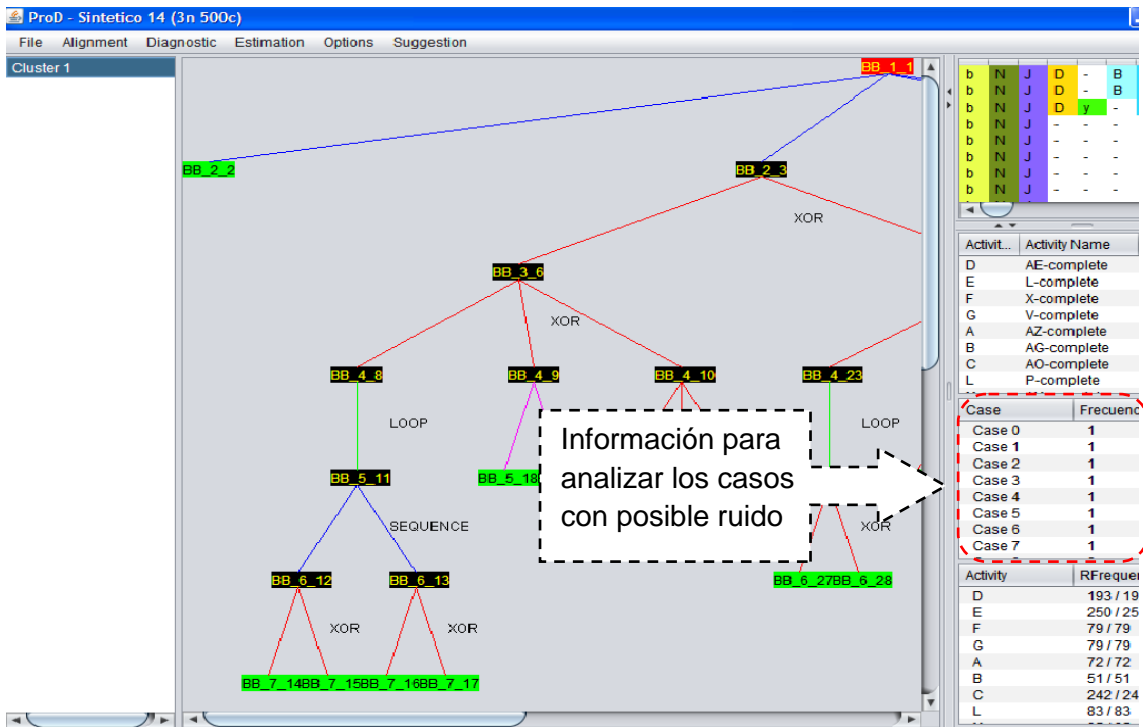


Figura 4: Árbol de construcción con los aspectos que caracterizan al proceso 1.

Capítulo 2: Propuesta de solución

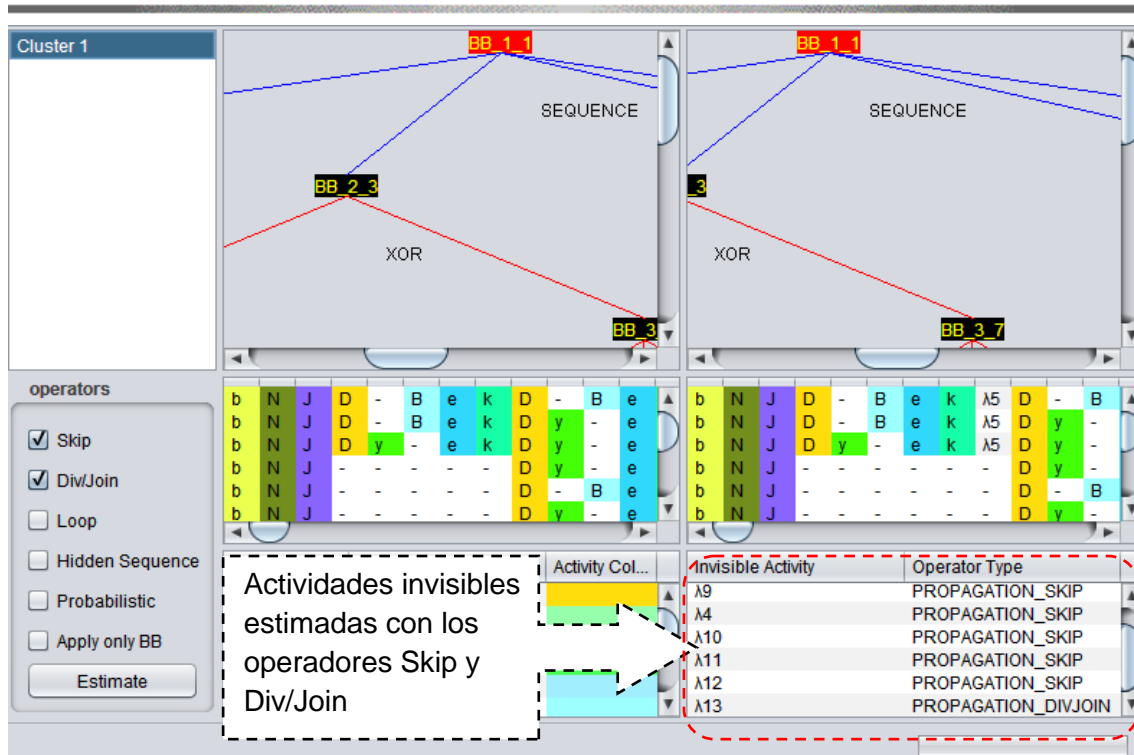


Figura 5: Árbol de construcción estimado con tareas invisibles.

Este algoritmo además de identificar los aspectos anteriores (ver Figura 3) brinda información para la identificación de casos con posibles situaciones de ruido (ver Figura 4), haciendo un análisis de las frecuencias de aparición por casos. En la sesión de estimación se identifican posibles ausencias de información al aplicar algunos de los operadores existentes para realizar este tipo de operación (ver Figura 5).

2.5. Sugerencia de algoritmos de descubrimiento.

La sugerencia de algoritmos de descubrimiento, permite obtener un listado de las técnicas más eficientes para extraer un modelo de procesos, utilizando la información contenida en un conjunto de trazas. Esta sugerencia tiene como entrada el resultado obtenido al aplicar el algoritmo para determinar el árbol de bloques de construcción publicado en (RAYKENLER YZQUIERDO-HERRERA, 2012) con las diferentes características por las que se puede descomponer el proceso, también se utilizó el resultado al aplicar los diferentes operadores de estimación al árbol de bloques de construcción para detectar si existe o no tareas invisibles en el proceso, se analizó la frecuencia de aparición de los casos para descubrir posibles ruidos en el registro de evento y por último, se hizo uso de los vectores obtenidos a partir de la evaluación realizada a las técnicas de descubrimiento.

2.5.1. Determinar patrones en un árbol de bloques de construcción.

Capítulo 2: Propuesta de solución

Para obtener algunas de las características presente en un árbol de construcción se propuso un algoritmo (Recorrer Árbol) que tiene como objetivo encontrar los patrones que posee el proceso al descomponerse en bloques de construcción. Los patrones a encontrar son: secuencia, paralelismo, selección (OR y XOR) y secuencia oculta. Este algoritmo recibe como entrada un Árbol T y una lista vacía, devolviendo al final una lista de todos los tipos de patrones encontrados en T .

Algoritmo 1: Recorrer árbol

Entrada: Un árbol de bloques de construcción (t) y una lista vacía Aux

Salida: Una lista de patrones Aux .

1. **Si** Tipo de patrón de $t \neq$ ningún tipo \wedge tipo de patrón de t no está en Aux
Entonces
2. $Aux \leftarrow$ tipo de patrón;
FinSi
3. **Para cada** Hijos de T **Hacer**
4. **Recorrer Árbol** (t , Aux)
FinPara
5. **Devolver** Aux ,

FinProceso

2.5.2. Detectar posibles casos afectados por el ruido.

Para detectar posibles ruidos en los registros de eventos se planteó el algoritmo Buscar ruido que tiene como objetivo encontrar el o los posibles casos afectados por esta característica, teniendo en cuenta su frecuencia de aparición. Este algoritmo tiene como entrada un bloque de construcción y devuelve la menor frecuencia de aparición junto a su identificador.

Algoritmo 2: Buscar ruido

Entrada: Un bloque de construcción.

Salida: Una lista con el identificador y la frecuencia del menor caso.

1. Crear dos listas vacía $ListaNombre$ y $ListaFrecuencia$.

Capítulo 2: Propuesta de solución

2. Crear una lista (*ListaAuxiliar*) que va a contener la cantidad de veces que se repite cada caso dentro de un bloque de construcción.
3. Crear una lista vacía (*ListaCasos*).
4. **Para** $i < 0$ **Hasta** n **Con Paso 1 Hacer**
5. $ListaNombre(i) \leftarrow " " + "Case " + i$;
6. $ListaFrecuencia(i) \leftarrow " " + ListaAuxiliar(i)$;
- FinPara**
7. **Para** $i < 0$ **Hasta** n **Con Paso 1 Hacer**
8. **Para** $j < 0$ **Hasta** k **Con Paso 1 Hacer**
9. Ordenar de menor a mayor *ListaNombre* y *ListaFrecuencia* por los valores de *ListaFrecuencia*.

FinPara

FinPara

10. **Si** $ListaFrecuencia(0) < ListaFrecuencia(0)$ **Entonces**
11. $ListaCasos(0) \leftarrow ListaFrecuencia(0)$;
12. $ListaCasos(1) \leftarrow ListaNombre(0)$;

FinSi

13. **Devolver** *ListaCasos*

FinAlgoritmo

2.5.3. Determinar características del proceso.

Para completar el vector de características del proceso analizado se desarrolló un algoritmo (Completar vector) que tiene como objetivo devolver una lista con las características del proceso en análisis. Esta lista contiene ceros y unos. El cero indica que (no) y el uno que (si) posee una característica en específico. Las características pueden ser (secuencia, paralelismo, lazos, selección, sin libre elección, tareas invisibles, tareas duplicadas, ruido y si el proceso es desestructurado o no).

Algoritmo 3: Completar vector

Entradas: Un árbol de bloques de construcción (t) y las siguientes variables que pueden ser Verdadero o Falso.

Invisible: Indica si existe o no tareas invisibles en el proceso.

TD: Especifica si el proceso posee tareas duplicadas.

Capítulo 2: Propuesta de solución

Ruido: Indica si el proceso ha sido afectado por el ruido.

SLE: Especifica si en el proceso está presente el constructor sin libre elección.

Salida: Una lista de elementos en ceros o unos.

1. Crear dos listas vacía *ListaVector* y *ListaPatrones*. *ListaVector* recogerá valores de ceros y unos.
2. Llamar al algoritmo **Recorrer árbol** (*t*, *ListaPatrones*).
3. **Mientras** *ListaPatrones* <> \emptyset **Hacer**
4. Completar la *ListaVector* con uno, siempre y cuando la característica se encuentre en *ListaPatrones*.

FinMientras.

5. Completar la lista *ListaVector* con uno si las variables son Verdadero o en el caso de que sean Falso completar con cero, si existe otras características que no se han completado al verificar estas características rellenar con cero.
6. **Devolver** *ListaVector*

FinAlgoritmo

2.5.4. Sugerir técnica de descubrimiento.

Para poder hacer una correcta sugerencia de técnicas, es necesario encontrar los aspectos que tiene el proceso analizado, pero para completar ese vector de características es necesario buscar las posibles afectaciones que puedan aparecer en el registro de evento por el ruido. Luego de tener el vector creado, entonces es posible diseñar un algoritmo que a partir de lo mencionado sea capaz de sugerir un conjunto de algoritmos.

La sugerencia de técnicas, tiene como objetivo proponer los algoritmos más adecuados para el descubrimiento del proceso en análisis, a partir de una lista de valores que obtiene a partir del algoritmo anterior. A continuación se calcula la distancia entre la evaluación de cada uno de los algoritmos almacenada en la (*MatrizEvaluación*) y la lista de valores (*ListaValor*) obtenida del Algoritmo 3. Luego se ordena esa distancia junto a los nombres de los algoritmos y se obtiene como resultado una lista de sugerencia, ordenados de menor a mayor por la distancia entre cada técnica.

Selección del criterio de distancia.

Capítulo 2: Propuesta de solución

Para elaborar la lista de sugerencias se utiliza la distancia Euclidiana la cual se define a continuación:

Definición de distancia Euclidiana: Se denomina distancia euclídea entre dos puntos $A(x_1, y_1)$ y $B(x_2, y_2)$ a la longitud del segmento de recta que tiene por extremos A y B .

Se expresa matemáticamente como: $d(A, B) = \sqrt{(X_B - X_A)^2 + (Y_B - Y_A)^2}$.

Se revisaron además las distancias de Mahalanobis y Manhattan.

No se selecciona la distancia de Mahalanobis puesto que su utilidad radica en que es una forma de determinar la similitud entre dos variables aleatorias. Se expresa matemáticamente como: $d_m(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T \Sigma^{-1}(\vec{x} - \vec{y})}$. En el caso que se está

analizando, entre las variables (aspectos que caracterizan un proceso) no existe correlación alguna, además las dimensiones de las variables no poseen un límite superior ya que sus valores están comprendidos entre cero y uno.

En el caso de la distancia de Manhattan: Se expresa matemáticamente como: $d_M(A, B) = \sum_{k=1}^p |b_k - a_k|$ y al aplicarla en el caso que se está analizando se obtienen iguales resultados que la Euclidiana en cuanto al orden de la propuesta. Esto se debe a que la distancia buscada depende directamente de la resta absoluta de los valores de cada atributo. Por lo que también pudo ser una buena selección para calcular la distancia existente entre el vector del proceso que se quiere analizar y cada uno de los vectores característicos de las técnicas de descubrimiento.

A parte de la distancia de Mahalanobis y Manhattan, se revisaron otros criterios de distancias como: Hamming, Hausdorff, Levenshtein y Minkowski las cuales no se corresponden con lo que necesita esta investigación. Por ejemplo: Hamming se basa en diferencia entre una palabra de código válido y otra, Hausdorff mide cuán lejos están uno de otros dos subconjuntos compactos de un espacio y Levenshtein mide la distancia entre palabras (el número mínimo de operaciones requeridas para transformar una cadena de caracteres en otra).

Algoritmo 4: Sugerencia de técnicas

Entradas: Una *ListaValor* y una *MatrizEvaluación*.

Salida: Una lista (*ListaNombres*).

1. Inicializar variables *aux*, *temp* y *resultado*.

Capítulo 2: Propuesta de solución

2. Crear dos listas *ListaNombres* y *ListaAspectos*. *ListaNombre* almacena los nombres de las técnicas y *ListaAspectos* almacena los diferentes aspectos o características que puede tener un proceso.
3. Crear dos listas vacías *ListaDistancia* y *ListaAuxiliar*. *ListaDistancia* almacena la distancia entre las técnicas y el proceso analizado. *ListaAuxiliar* guardará valores modificados de las técnicas.
4. **Para $k < 0$ Hasta n Con Paso 1 Hacer**
5. **Para $j < 0$ Hasta k Con Paso 1 Hacer**
6. **Si $ListaValor(j) == 0 \wedge MatrizEvaluación < > 0$ Entonces**
7. *ListaAuxiliar(i) < -0;*
8. **Sino**
9. *ListaAuxiliar(i) < -MatrizEvaluación FinSi*
- FinPara**
10. **Para $j < 0$ Hasta k Con Paso 1 Hacer**
11. *resultado+ < -(ListaValor(j)-ListaAuxiliar(j))²;*
- FinPara**
12. *ListaDistancia < -√resultado;*
13. *resultado < -0;*
- FinPara**
14. **Para $k < 0$ Hasta n Con Paso 1 Hacer**
15. **Para $j < 0$ Hasta k Con Paso 1 Hacer**
16. Ordenar simultáneamente de menor a mayor *ListaDistancia* y *ListaNombre*, a partir de los valores en la *ListaDistancia*.
- FinPara**
- FinPara**
17. **Devolver *ListaDistancia***

FinAlgoritmo

Conclusiones del capítulo.

En este capítulo se le dio cumplimiento a los objetivos específicos dos y tres planteados para la investigación. Luego de hacer un estudio sobre los algoritmos de descubrimiento se identificaron los aspectos a tener en cuenta para caracterizar un proceso. A partir de estas características se realizó una evaluación de las técnicas, lo cual posibilitó la construcción de los vectores característicos de cada una de ellas, que sirven como entrada a la propuesta de solución.

Capítulo 2: Propuesta de solución

El desarrollo del algoritmo para la sugerencia de técnicas de descubriendo posibilitará que la herramienta ProD además de diagnosticar y estimar información ausente, sea capaz de identificar otras características como tareas duplicadas, ruido, saber si el proceso es desestructurado o no y por último, sugerir que técnicas de descubrimiento son las más adecuadas según las características que posea el proceso en análisis.

Capítulo 3: Validación de la solución.

Introducción.

Cuando se está realizando una aplicación es inevitable cometer errores. Errores que si no se les da tratamiento pueden afectar el correcto funcionamiento del sistema o de la propuesta de solución. Una forma de comprobar que el trabajo que se está realizando está libre de errores es usando métricas.

En este capítulo se realizan los tipos de pruebas que son necesarias aplicar para poder afirmar que se cuenta con una buena propuesta de solución. Para ello se cuenta con la métrica Fitness y una evaluación cualitativa de los resultados propuestos.

4.1. Implementación del algoritmo propuesto.

Se desarrolla un algoritmo que permite sugerir las técnicas de descubrimiento que se pueden utilizar en la fase posterior, a partir de un diagnóstico realizado a un registro de eventos en la Minería de Procesos.

3.1.1. Herramienta y tecnología utilizada.

Para la implementación del algoritmo se utilizó como IDE de desarrollo el Netbeans en su versión 6.9 y como lenguaje de programación Java.

Netbeans.

Netbeans es un entorno de desarrollo, que permite a los programadores escribir, compilar, depurar y ejecutar programas. Está escrito en Java, pero puede servir para cualquier otro lenguaje de programación como C y C++. Netbeans IDE es un producto libre y gratuito, sin restricciones de uso (NETBEANS, 2012).

Lenguaje de programación Java.

Java es un lenguaje de programación con el que se puede realizar cualquier tipo de aplicación. Una de las principales características que favoreció la difusión y el crecimiento de este lenguaje es su capacidad de que el código funcione sobre cualquier plataforma de software y hardware. Esto quiere decir que si se realiza un programa en Java podrá funcionar en cualquier ordenador del mercado. Otras características importantes de este lenguaje son: lenguaje Orientado a Objetos (OO) y altamente fiable (CAÑETE, 2007).

3.1.2. Aplicación ProD.

La aplicación está enmarcada en cuatro momentos fundamentales: cargar el registro de eventos, alinear las trazas de este registro de eventos, estimar información ausente y por último la sugerencia para el descubrimiento.

En un primer momento la aplicación permite cargar un registro de eventos y configurar los parámetros necesarios para la alineación de las trazas. Para el desarrollo del componente de alineación de las trazas se utilizó el plugin desarrollado por Bose y Van der Aalst.

En la Figura 6 se muestra una imagen de la aplicación en la que se evidencia la alineación de un registro de eventos. En la parte derecha de la Figura aparece el panel correspondiente a los parámetros para la configuración de la alineación.

Posterior a la alineación de las trazas se realiza la descomposición del proceso. La Figura 7 muestra el árbol de bloque de construcción producto de la descomposición del proceso analizado. En la Figura 7 se muestran tres paneles verticales, el primero (de izquierda a derecha) muestra las agrupaciones creadas en la alineación. A partir de seleccionar una agrupación se muestra en el panel del centro el árbol de bloques de construcción producto de su descomposición. En cada nodo del árbol se muestra el nombre del bloque de construcción creado, las relaciones que se establecen entre los bloques de construcción se reflejan en los arcos de dos formas, mediante un mensaje de texto y utilizando colores diferentes para cada tipo de relación. Los nodos de color verde representan hojas del árbol, los azules bloques de construcción que no pudieron descomponerse y los negros nodos intermedios. El árbol se puede expandir o contraer según se desee.

En el panel derecho se muestran cuatro secciones. La primera (de arriba hacia abajo) muestra el contenido del bloque de construcción seleccionado en el panel del centro. En la segunda sección se muestra la leyenda construida durante la alineación de las trazas. En la siguiente sección se muestra la frecuencia de aparición de los casos que conforman el bloque de construcción seleccionado. En la última sección se muestra la frecuencia relativa de ocurrencia de las actividades que conforman el bloque de construcción seleccionado.

Capítulo 3: Validación de la solución

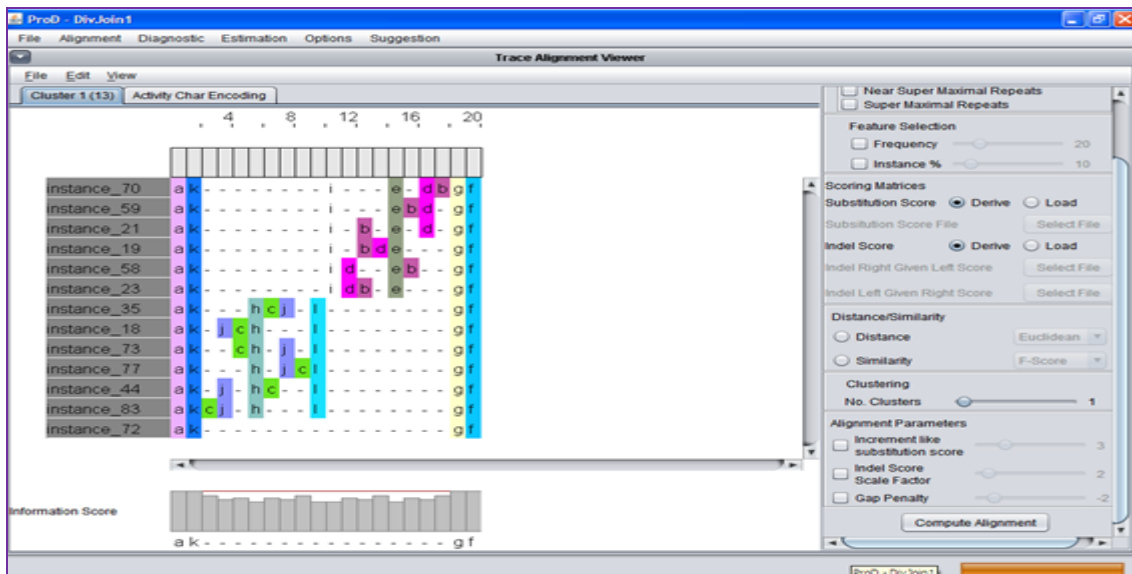


Figura 6: Alineación de las trazas.

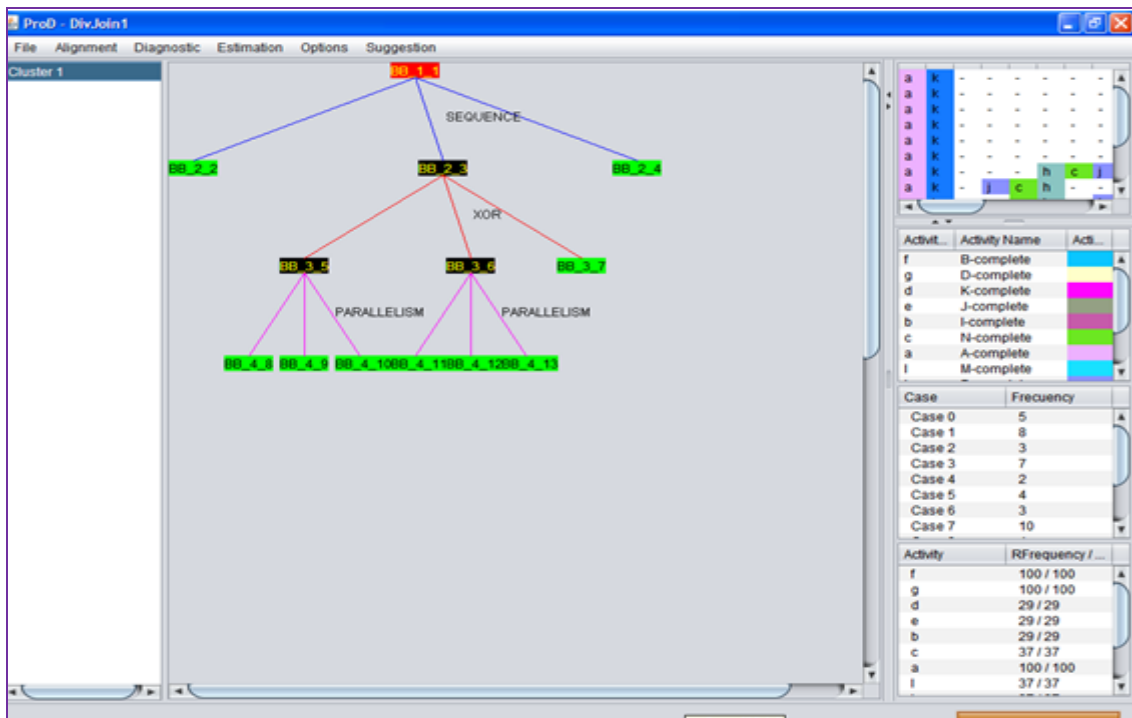


Figura 7: Árbol de construcción de bloques.

En un tercer momento, se realiza la estimación de información ausente a partir del árbol de bloque de construcción obtenido anteriormente. La Figura 8 muestra la estimación realizada a un proceso analizado.

Capítulo 3: Validación de la solución



Figura 8: Estimación de información ausente.

En la Figura 8 se muestran tres paneles verticales. El primero (de izquierda a derecha) muestra en la sección superior el identificador del árbol asociado a cada grupo obtenido en la alineación. En la sección inferior aparecen los operadores que pueden aplicarse al bloque de construcción seleccionado o todos los bloques de construcción que conforman el árbol obtenido anteriormente.

En el panel del centro se muestra el árbol de bloques de construcción obtenido, así como los detalles del bloque seleccionado y la leyenda.

En el panel de la derecha se muestra en su sección superior el árbol de bloques de construcción estimado. En la segunda sección aparecen los detalles del bloque de selección seleccionado en la sección anterior y en la última sección, se muestra un resumen de las actividades invisibles que fueron insertadas y la razón por la cual se adicionó.

En la parte superior derecha de la Figura 8 aparece un menú que permite salvar la solución obtenida a partir de la estimación de información ausente.

En un último momento se realiza la sugerencia de los algoritmos de descubrimiento más adecuados de acuerdo a las características del proceso en análisis. Este momento se apoya en la información obtenida del árbol de construcción y la estimación de información ausente.

Capítulo 3: Validación de la solución

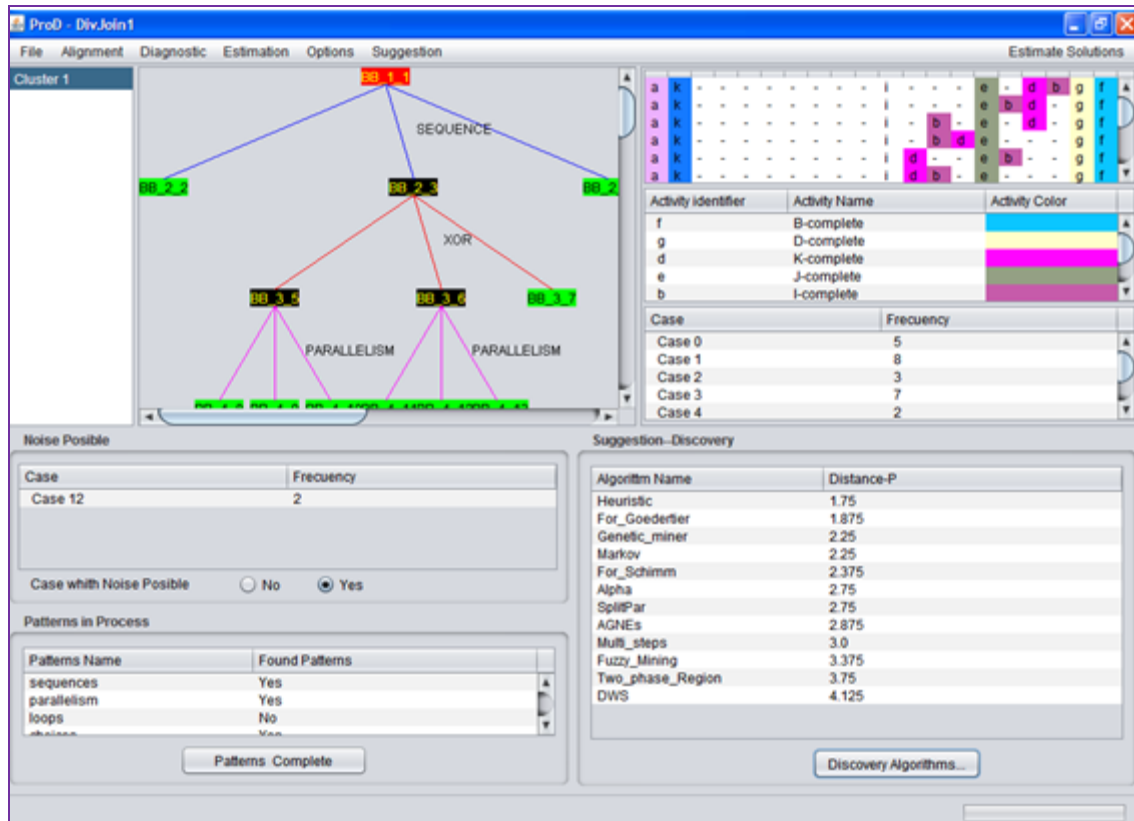


Figura 9: Sugerencia para el descubrimiento de modelo de proceso.

En la Figura 9 se muestran 2 paneles horizontales, los cuales están divididos en 3 partes. El primero (de izquierda a derecha) muestra en la sección superior el identificador del árbol asociado a cada grupo obtenido en la alineación. A partir de seleccionar una agrupación, se muestra en el panel del centro el árbol de bloque de construcción producto de su descomposición.

En el panel superior derecho se muestran tres secciones. La primera (de arriba hacia abajo) muestra el contenido del bloque de construcción seleccionado en el panel del centro. En la segunda sección se muestra la leyenda construida durante la alineación de las trazas. En la siguiente sección se muestra la frecuencia de aparición de los casos que conforman el bloque de construcción seleccionado.

En el segundo panel horizontal se encuentran dos paneles verticales. El primer panel está dividido en dos secciones. La primera representa los posibles casos que pueden ser tratados como ruido, en el cual se muestra el identificador del caso y la frecuencia de aparición.

Capítulo 3: Validación de la solución

La segunda sección muestra una tabla resumen con los diferentes aspectos que se tienen en cuenta para analizar un proceso, esta dada por *Yes* o *No*, en dependencia si aparece o no el constructor en el proceso.

Por último, en el panel derecho se muestran los nombres de los algoritmos de descubrimiento y la distancia que existe entre el vector del algoritmo mencionado y el proceso en análisis.

3.2. Resultados experimentales.

Se definieron un conjunto de procesos para probar la efectividad del algoritmo desarrollado y para ello se emplea la aplicación informática expuesta anteriormente.

3.2.1. Métrica.

A partir de la sugerencia de algoritmos se obtiene un grupo de técnicas de descubrimiento, que posibilitan la obtención de modelos de procesos comprensibles y más completos. En consecuencia, es necesario determinar la forma en la que se debe evaluar la estructura y comprensión de dichos modelos, por lo que se necesita medir el grado de correspondencia entre el modelo descubierto y el comportamiento observado en el registro de eventos. En este trabajo la métrica seleccionada es el *Fitness*, esta utiliza modelos representados por redes de Petri y Heurísticas; es usada para medir la correspondencia entre un registro de eventos y la estructura del modelo descubierto.

3.2.2. Características de los procesos analizados.

Para realizar las pruebas se utilizaron tres procesos que fueron generados recurriendo a la herramienta *Process Log Generator* en la versión 1.4 beta. Se decidió utilizar una herramienta que generara de manera aleatoria un registro de eventos artificial, debido a que en el área de Minería de Procesos no se detectó una base de datos que contenga procesos que puedan ser utilizados para la validación del algoritmo desarrollado.

Aunque existen algunos registros de eventos asociados a procesos reales, no todos cuentan con las características adecuadas para realizar una correcta validación de la propuesta. Los problemas están relacionados con el reflejo parcial de los patrones de flujo de trabajo y la cantidad de casos.

Capítulo 3: Validación de la solución

En un registro de eventos generado de manera artificial, por una de las herramientas desarrolladas para realizar esta operación, se pueden reflejar los patrones de flujo de control conocidos, tener diferentes niveles de patrones anidados y la cantidad de casos puede variar en correspondencia con los elementos enunciados. Las características de los registros de eventos generados son las siguientes:

Tabla 3: Descripción de los registros de eventos.

Atributos	Proceso1	Proceso2	Proceso3
Cantidad de casos	1000	100	50
Eventos	10400	693	553
Cantidad de patrones anidados	3	2	
Clases de eventos	34	16	13
Refleja el patrón Secuencia	Si	Si	Si
Refleja el patrón Lazos	No	No	Si
Refleja el patrón AND Split/join, XOR Split/Join	Si	Si	Si
Refleja el patrón Ausencia de información	No	Si	No
Refleja el patrón Sin libre elección	No	Si	No
Es desestructurado	No	No	No
Afectado por el ruido	No	No	No

Para la confección de los registros de eventos se fueron variando las características distintivas en los procesos, como son: el ruido, las tareas invisibles, los lazos, etc., que son necesidades que pueden poseer los registros, pero que no todos los algoritmos de descubrimiento son capaces de satisfacer, además se combinaron con otras variables como: cantidad de casos y clases de eventos.

3.2.3. Análisis de los resultados.

Capítulo 3: Validación de la solución

Para el análisis de los resultados obtenidos se utilizó la métrica Fitness, junto a un análisis cualitativo a partir de la observación de los resultados alcanzados con la sugerencia de técnicas para el descubrimiento de procesos.

La métrica utilizada aunque en ocasiones ofrece resultados satisfactorios, en otros casos no es el mejor criterio de medición para saber cuándo un modelo de proceso se corresponde con el registro de eventos.

La métrica Fitness presenta algunos inconvenientes cuando trata con modelos que poseen ausencia de información, debido a que no tiene como referente cuánta información falta en el registro de eventos. Al no poder hacer una correcta verificación entre el registro de evento y el modelo, se obtiene por lo general un Fitness igual a uno, porque no refleja este comportamiento o simplemente la métrica no es capaz de llegar a un resultado determinado. Otra desventaja de esta métrica es que no es posible aplicarla a todas las notaciones utilizadas por los algoritmos de descubrimiento, en este caso se puede encontrar la notación utilizada por Multi fases y Fuzzy Miner. Por tanto, para poder darle una evaluación a estos algoritmos se realizaron algunas comparaciones entre el modelo resultante al aplicar Multi fases o Fuzzy Miner y el resto de los modelos obtenidos al aplicar los demás algoritmos, en dependencia de la similitud entre dichos modelos, se escogió una evaluación correspondiente al Fitness del modelo que más se le parecía. Para apoyar esta métrica, se hará uso de una valoración cualitativa, con el objetivo de comprobar que los resultados que se van obteniendo son satisfactorios.

A partir de lo expuesto anteriormente se realizaron algunas pruebas para medir la efectividad de la propuesta.

Análisis del Proceso 1.

Con el análisis del Proceso 1 se obtuvieron las siguientes recomendaciones, Tabla 4. Este proceso tiene como características principales: secuencia, paralelismo y selección, situaciones que no afectan la evaluación de los algoritmos, Tabla 5.

Este es un caso donde cualquiera de los algoritmos que se recomiende puede ser capaz de descubrir el modelo sin ninguna complicación, puesto que las técnicas existentes son capaces de satisfacer estas tres necesidades (secuencia, paralelismo y selección).

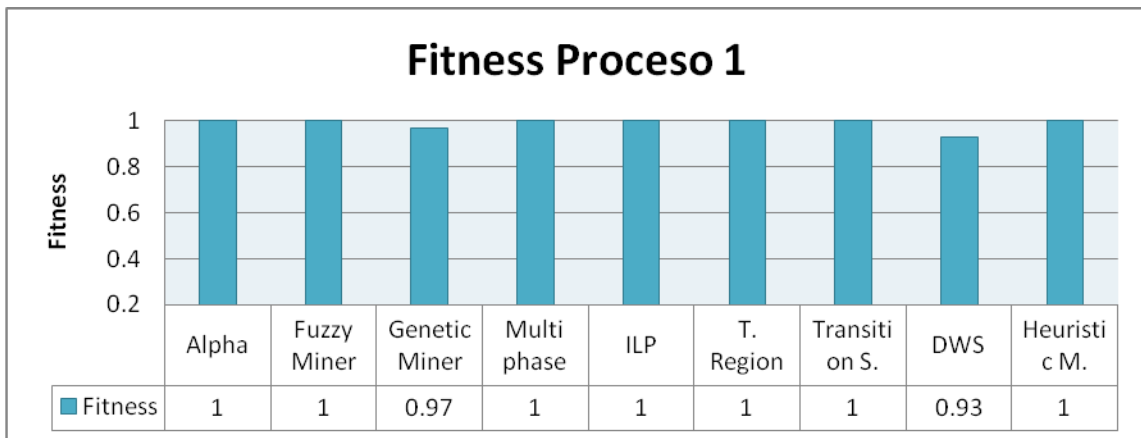
Tabla 4: Sugerencias para el Proceso 1.

Técnica	Distancia
Alpha	0.0
DWS	0.0
Multi phases	0.0
T. Region	0.0
ILP Miner	0.0
Transitions Systems	0.0
Heuristic Miner	0.0
Genetic Miner	0.0
Fuzzy Miner	0.0

Por ende, al analizar los resultados se obtiene una lista donde los 9 algoritmos propuestos tienen una distancia de (0.0), es decir, que cualquiera puede descubrir el modelo. Al considerar el Fitness de los algoritmos Alpha, DWS, Multi phases, T. Region, ILP Miner, Transitions Systems, Fuzzy Miner y Heuristic Miner se aprecia que tienen un valor de 1.0; pero en el caso del Genetic Miner el Fitness es igual a 0.98, este resultado es un aproximado que está en el intervalo de correspondencia, ya que al aplicar este algoritmo nunca hay un valor exacto, es decir un Fitness de 1.0, por esta razón se decidió que los valores que estén entre este intervalo (0.97-0.99) se tomen como un Fitness de 1.0.

La Tabla 5 muestra un resumen de los resultados de la métrica Fitness para los modelos obtenidos del Proceso 1.

Tabla 5: Fitness para los modelos obtenidos del Proceso 1.



En este proceso se evidencia que el Fitness no siempre es exacto o no es la mejor opción para medir correspondencia entre el modelo y el registro de eventos, por lo que en algunas ocasiones hay que apoyarse en otras vías para realizar una evaluación más específica, como el análisis cualitativo.

Análisis del Proceso 2.

Para el análisis del Proceso 2 se muestran los resultados obtenidos en la Tabla 6. Este proceso tiene como característica principal la ausencia de información, además de poseer los patrones secuencia, paralelismo, selección y sin libre elección. Al aplicarle al proceso el algoritmo desarrollado, se obtiene como resultado una lista ordenada de técnicas, donde el primer elemento es el más recomendado, por ser la técnica que mejor se corresponde con las necesidades del proceso analizado.

En la lista obtenida aparece en primer lugar la técnica Minería Genética (Genetic Miner) con una distancia de 0.25, esta permite el descubrimiento de las características siguientes: secuencia, lazos, paralelismo, selección, sin libre elección, tareas duplicadas, tareas invisibles y es robusto al ruido. Además se precisa en esta posición por su capacidad para tratar con la ausencia de información en los registros de eventos, ya que es capaz de descubrir seis de ocho situaciones posibles, siendo la mejor en este aspecto. Al comprobar la correspondencia entre el registro de eventos y el modelo descubierto (el Fitness) se obtiene como resultado un valor de 0.99, lo que significa que esta técnica es capaz de reproducir en el modelo todos o casi todos los rastros en el registro de eventos.

Capítulo 3: Validación de la solución

A continuación del algoritmo Genetic Miner aparece en la lista el algoritmo Fuzzy Miner, con una distancia de 0.375, siendo capaz de soportar cinco de ocho situaciones posibles donde se puede manifestar la ausencia de información, además es capaz de descubrir las siguientes características: secuencia, lazos, paralelismo, selección, sin libre elección, tareas duplicadas, tareas invisibles, es robusto al ruido y permite trabajar con procesos desestructurados. Debido a la notación que utiliza el Fuzzy Miner para el descubrimiento de modelos de proceso no es posible aplicar la métrica propuesta (Fitness), por lo que para comprobar dicha correspondencia se hizo comparaciones entre el modelo obtenido y otros modelos descubiertos, y se concluyó un Fitness con valor 1.0 para la evaluación del algoritmo.

En tercer lugar aparece la técnica ILP, capaz de descubrir cuatro de ocho situaciones posibles donde se puede manifestar la ausencia de información. El resultado al aplicarle la métrica (Fitness) fue de 1.0.

Tabla 6: Sugerencias para el Proceso 2.

Técnica	Distancia
Genetic Miner	0.25
Fuzzy Miner	0.375
ILP Miner	0.5
Transitions Systems	0.625
Alpha	0.75
T. Region	0.75
Multi phases	1.0
DWS	1.125
Heuristic Miner	1.25

En el cuarto lugar están las técnicas Transitions Systems y Alpha con una distancia de 0.75, capaces de soportar tres de ocho situaciones posibles donde se puede manifestar la ausencia de información, con Fitness 1.0 cada una. Y en los últimos tres lugares aparecen Multi phases, DWS y Heuristic Miner con una distancia de 1.0, 1.125 y 1.25 respectivamente. Estas técnicas son capaces de descubrir cuatro, tres, y dos situaciones de ausencia de información respectivamente de las ocho existentes. En el caso de la técnica Multi phases, debido a la notación utilizada para la representación del modelo no es posible aplicarle el Fitness, por lo que se hizo una comparación entre

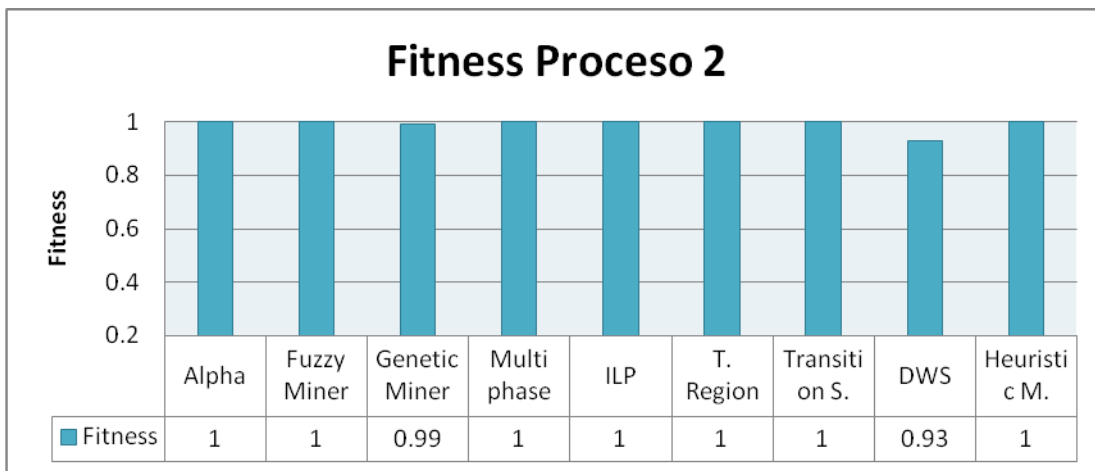
Capítulo 3: Validación de la solución

este modelo y otro modelo descubierto, al ver la similitud entre ambos modelos se le dio un Fitness de 1.0. En el caso de DWS y Heuristic Miner obtuvieron un Fitness de 0.93 y 1.0 respectivamente.

Después de analizar los resultados del Fitness, los modelos obtenidos y la evaluación cualitativa, se puede decir que la sugerencia fue buena. La evaluación cualitativa coincide con la capacidad de los algoritmos de soportar o descubrir la ausencia de información en las diferentes situaciones donde puedan presentarse, además de poder descubrir las restantes características que presenta el proceso.

La Tabla 7 muestra un resumen de los resultados de la métrica Fitness para los modelos obtenidos del Proceso 2. Los valores del Fitness oscilan entre cero y uno, mientras más cerca del uno esté, mejor será la información que se obtenga del modelo, debido a que este será capaz de reproducir la mayoría de los rastros en el registro de eventos.

Tabla 7: Fitness de los modelos obtenidos del Proceso 2.



Análisis del Proceso 3.

En el análisis realizado al Proceso 3 se puede apreciar que los patrones comunes que aparecen en este proceso son secuencia, lazos y selección, pero el patrón distintivo en este proceso es lazos. Al aplicarle al proceso el algoritmo desarrollado se obtuvo como resultado una lista de técnicas que permiten el manejo de los patrones que se encuentran en el proceso. En esta lista los algoritmos desde el Alpha hasta el Fuzzy Miner tienen una distancia de 0.0 debido a que cubren totalmente todas las características que presenta el proceso. En el caso del algoritmo Genetic Miner se encuentra en la octava posición debido a que tiene una distancia de 0.199 ya que

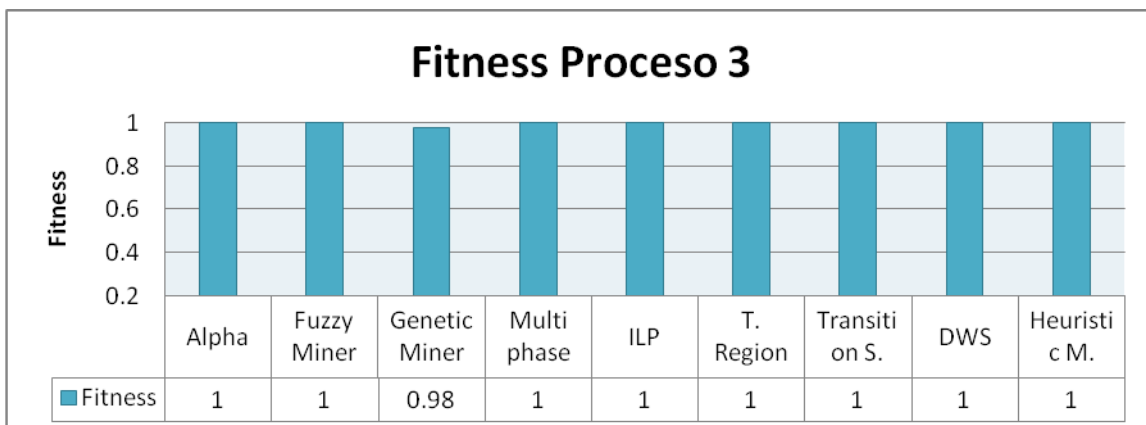
Capítulo 3: Validación de la solución

posee dificultad al extraer modelos con construcciones que permiten muchas situaciones de entrelazado y a menudo utiliza un espacio de búsqueda más grande, por lo tanto, necesita más iteraciones para converger a soluciones buenas. En la última posición se encuentra el Heuristic Miner con una distancia de 0.5 el cual cubre parcialmente la característica distintiva, que en este proceso específicamente son los lazos, pudiendo soportar de dos situaciones sólo una.

Tabla 8: Sugerencia del Proceso 3.

Técnicas	Distancia
Alpha	0.0
DWS	0.0
Multi phases	0.0
T. Region	0.0
ILP Miner	0.0
Transitions System	0.0
Fuzzy Miner	0.0
Genetic Miner	0.199
Heuristic Miner	0.5

Tabla 9: Fitness de los modelos obtenidos del Proceso 3.



Conclusiones del capítulo.

En este capítulo se le dio cumplimiento al objetivo 4 planteado para la investigación. En función de darle cumplimiento a este objetivo se definieron tres procesos que

Capítulo 3: Validación de la solución

fueron generados utilizando la herramienta Process Log Generator, contando con las características adecuadas para realizar una correcta validación de la propuesta.

Además se seleccionó la métrica Fitness, la cual permite medir la correspondencia entre un registro de eventos y la estructura del modelo descubierto, usando las técnicas desarrolladas en la Minería de Procesos.

Debido a que esta métrica no es del todo eficiente ya que en ocasiones a la notación utilizada por algunos algoritmos para representar el modelo no se le puede aplicar dicha métrica, se hizo necesario hacer un análisis cualitativo a partir de la observación de los resultados alcanzados con la sugerencia de técnicas para el descubrimiento.

De acuerdo a los resultados que se obtuvieron al analizar el Fitness, los modelos obtenidos y la evaluación cualitativa, se puede decir que la sugerencia fue satisfactoria. La evaluación cualitativa coincide con la capacidad de los algoritmos de soportar o descubrir los patrones comunes en las diferentes situaciones donde puedan presentarse en el proceso.

Conclusiones generales.

Para el desarrollo de la investigación se plantearon un conjunto de objetivos que fueron cumplidos gradualmente.

Se desarrolló un algoritmo para el diagnóstico de procesos y a partir de la investigación realizada se permitió llegar a las siguientes conclusiones:

Las técnicas de descubrimiento permiten obtener un modelo de proceso a partir de un registro de eventos proporcionado por los sistemas de información.

Actualmente la selección de una técnica de descubrimiento es complicada, debido a que no se conoce con antelación cuáles son las características que posee el proceso en análisis.

De las técnicas analizadas para el diagnóstico de procesos, la mayoría no hacen un examen completo del proceso, donde se pueda identificar las características principales del mismo y se propongan que técnicas de descubrimiento son las más adecuadas para descubrir el modelo de procesos.

Para la sugerencia de algoritmos de descubrimiento se tuvo en cuenta las características derivadas de un árbol de bloque de construcción, obtenidas a través del algoritmo ProD, la estimación de información teniendo en cuenta los diferentes operadores para estimar y descubrir ausencia de información en el registro de evento y por último se utilizó la información referente a la frecuencia de aparición de los diferentes casos en un árbol de bloques de construcción.

A partir de una evaluación realizada al algoritmo desarrollado, se determina su efectividad para realizar las sugerencias de técnicas de descubrimiento, teniendo en cuenta las características detectadas en el proceso analizado.

Referencias Citadas

Recomendaciones.

Aplicar la solución en escenarios reales que permitan su validación.

Mejorar la detención de ruido utilizando otras variables como la verificación de restricciones.

Flexibilizar la herramienta para que permita la incorporación de nuevos algoritmos de descubrimiento.

Referencias citadas:

- _ACADEMIA, D. D. L. R. *descubrimiento*, 2010. [2012]. Disponible en: http://buscon.rae.es/drael/SrvltConsulta?TIPO_BUS=3&LEMA=descubrimiento
- A.J.M.M. WEIJTERS, W. M. P. V. D. A. Rediscovering Workflow Models from Event-Based Data using Little Thumb, 2003a.
- A.J.M.M. WEIJTERS, W. M. P. V. D. A., AND A.K. ALVES DE MEDEIROS Process Mining with the HeuristicsMiner Algorithm, 2003b.
- A.K.A. DE MEDEIROS, B. F. V. D., W.M.P. VAN DER AALST, AND A.J.M.M. WEIJTERS. *Process Mining: Extending the Alpha-algorithm to Mine Short Loops*, 2004.
- AALST, W. M. P. V. D. *Process Mining. Discovery, Conformance and Enhancement of Business Processes*. Springer Heidelberg Dordrecht London New York, 2011. p. 978-3-642-19344-6
- AALST, W. M. P. V. D.; A. ADRIANSYAH, *et al.* *Process Mining Manifesto Business Process Management Workshops 2011, Lecture Notes in Business Information Processing. Springer-Verlag*, 2011, 99.
- AALST, W. M. P. V. D.; V. RUBIN, *et al.* *Process Mining: A Two-Step Approach to Balance Between Underfitting and Overfitting Software and Systems Modeling*, 2009, 9(1): 87-111.
- AALST, W. M. P. V. D. and A. J. M. M. WEIJTERS *Process Mining: A Research Agenda Special Issue of Computers in Industry, Elsevier Science Publishers, Amsterdam*, 2004, 53(3).
- AALST, W. V. D. *Kees van Hee: Workflow Management: Models, Méthodos y Systems*, 2004.
- AALST, W. V. D. *Manifiesto sobre la Minería de Proceso.*, 2012.
- ADRIANSYAH, B. F. V. D. A. A. *Process Mining: Fuzzy Clustering and Performance Visualization*, 2010.
- B.F. VAN DONGEN, A. W. M. P. V. D. A. *Multi-Phase Process Mining: Aggregating Instance Graphs into EPCs and Petri Nets*, 2005.
- B.F. VAN DONGEN, N. B., G.M. PINNA, AND W.M.P. VAN DER AALST. *An Iterative Algorithm for Applying the Theory of Regions in Process Mining*, 2007.
- BERGENTHUM, R.; J. DESEL, *et al.* *Process Mining Based on Regions of Languages Lecture Notes in Computer Science*, 2007, 4714: 375-383.

Referencias Citadas

- BOSE, R. P. J. C., & AALST, W.M.P. VAN DER. "Process Diagnostics Using Trace Alignment: Opportunities, Issues, and Challenges" en *Information Systems*, 37(2), 117-141 2012.
- BOSE, R. P. J. C. and W. M. P. V. D. AALST. Abstractions in Process Mining: A Taxonomy of Patterns. en. DAYAL, U.;EDER, J.*et al*, Springer Berlin / Heidelberg, 2009. 5701: 159-175.p.
- BOSE, R. P. J. C. and W. M. P. V. D. AALST Process diagnostics using trace alignment: Opportunities, issues, and challenges *Inf. Syst.*, 2012a, 37(2): 117-141.
- BOSE, R. P. J. C. and W. M. P. V. D. AALST Process Diagnostics Using Trace Alignment: Opportunities, Issues, and Challenges *Information Systems*, 2012b, 37(2): 117-141.
- CAMACHO, R. *¿Qué es un proceso? – definición y elementos* 2008. [Disponible en: CAÑETE, S. C. *Tecnologías Java utilizando NetBeans 6*, 2007.
- COLÍN, L. *Las normas ISO 9000:2000 de Sistemas de Gestión de la Calidad*, 2000.
- COOK, J. E. *Process Discovery and Validation Through Event-Data Analysis*, 1996. p.
- CURTIS, B., M.I. KELLNER Y J. OVER;. *Process modeling, Communications ACM: 35(9)*,, 1992. 75- 90.
- CHEUNG, C. F., YU, J.X., LU, H. Constructing Suffix Tree for Gigabyte Sequences with Megabyte Memory *IEEE Trans. Knowl. Data Eng.*, 2005, 17(1).
- DONGEN, B. F. and A. ADRIANSYAH. Process Mining: Fuzzy Clustering and Performance Visualization
- Business Process Management Workshops. en. RINDERLE-MA, S.;SADIQ, S.*et al*, Springer Berlin Heidelberg, 2010. 43: 158-169.p.
- DONGEN, B. F. V. and W. M. P. V. D. AALST Multi-phase Process Mining: Building Instance Graphs *Lecture Notes in Computer Science*, 2004, 3288: 362-376.
- GOEDERTIER, S. *Declarative Techniques for Modeling and Mining Business Processes*, 2008.
- GÜNTHER, C. W. *Process Mining in Flexible Environments, Eindhoven University of Technology*, 2009.
- HUACOTO, N. E. Propuesta para implantar CMMI en una empresa con múltiples unidades desarrolladoras de software. Lima-Perú: Universidad Nacional Mayor de San Marcos, 2005.
- JACOBSON, I. "Applying UML in The Unified Process" 1998.

Referencias Citadas

- LI, J.; R. P. J. C. BOSE, *et al.* *Mining Context-Dependent and Interactive Business Process Maps using Execution Patterns*. BPM 2010 Workshops, Springer-Verlag,, 2010. 109 -121 p.
- MEDEIROS, A. K. A. D. *Genetic Process Mining*, 2006. p.
- MONARDES, J. M. *Control y Optimización de Procesos con Process Mining* 2011. [2012]. Disponible en: <http://bpmknowledge.bligoo.com/>
- NETBEANS. *Netbeans*, 2012. [Disponible en: <http://netbeans.org/community/releases/71/>
- QUINTANA, C. J. *Indicadores de Alineamiento entre Procesos de Negocios y Sistemas Informáticos*”, Universidad de Concepción, 2002. p.
- R. P. JAGADEESH CHANDRA BOSEA, W. M. P. V. D. A. *Process Diagnostics Using Trace Alignment: Opportunities, Issues, and Challenges*, September 14, 2011.
- R. YZQUIERDO-HERRERA, R. S.-C., M. LAZO-CORTÉS *Análisis de la Ausencia de Información en las Trazas Usadas en la Minería de Procesos*, 2011.
- RAYKENLER YZQUIERDO-HERRERA, R. S.-C., MANUEL LAZO-CORTÉS, ADRIAN TORRES-GRAÑA. *Diagnóstico de proceso basado en el descubrimiento de subprocesos*, 2012. Vol. XXXIII: p. 133-141.
- RAYKENLER YZQUIERDO HERRERA, A. T. G., ROGELIO SILVERIO CASTRO, MANUEL LAZO CORTÉS *Minería de proceso como herramienta para la auditoría*, 2011.
- ROZINAT, A., & AALST, W.M.P. VAN DER *Conformance Checking of Processes Based on Monitoring Real Behavior*, 2008.
- ROZINAT, A. and W. M. P. V. D. AALST *Conformance Checking of Processes Based on Monitoring Real Behavior*, 2008.
- SMITH, H. A. F., PETER. *Business Process Model*, 2003. [Disponible en:
- TERALOC. *Modelado de procesos de negocio* 2004. [2012]. Disponible en: http://www.teraloc.com/portal/index.php?option=com_content&view=article&id=51&Itemid=92
- VAN DER AALST, W. M. P. A. W., A J M M AND MARUSTER, L (2003). . *"Workflow Mining: Discovering process models from event logs"*, *IEEE Transactions on Knowledge and Data Engineering*, 2003. 16.
- WESKE, M. *BPM Concepts-Languages-Architectures*, 2007.
- WIL M. P. VAN DER AALST, A. H. M. T. H. A. M. W., Ed. *Business Process Management: A Survey*, 2003.

Referencias Citadas

WIL M.P , A. *Process Mining:Discovery, Conformance and Enhancement of Business Processes*. Springer Heidelberg Dordrecht London New York, 2011.

Tabla 10: Caracterización de las técnicas de descubrimiento a partir del constructor Tareas invisibles.

Algoritmo	Situación de salto	Situación de división/unión	Actividades invisibles contra Actividades duplicadas	Actividades invisibles contra lazos	Actividades invisibles	Lazos contra actividades invisibles junto a	Secuencia oculta de subprocesos	Opciones equiprobables	% de cumplimiento
Heuristic Miner	*		*						0.25
Alpha				*					0.125
Multi-phase	*	*	*	*					0.50
DWS	*		*	*					0.375
Alpha ++			*	*					0.25
Genetic Miner	*	*	*	*	*	*			0.75
Región Miner	*	*							0.25
Fuzzy Miner	*	*	*	*	*				0.625
Markov	*	*							0.25
Splitpar	*	*							0.25
Schimm	*								0.125
AGNEs	*								0.125
Goolestier	*								0.125
ILP Miner				*	*	*	*		0.50

Transitions System	*	*			*				0.375
---------------------------	---	---	--	--	---	--	--	--	-------