



Ingeniería Industrial

ISSN: 1815-5936

Facultad de Ingeniería Industrial, Instituto Superior
Politécnico José Antonio Echeverría, Cujae.

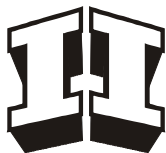
Pérez-Alfonso, Damián; Yzquierdo-Herrera, Raykenler;
Pupo-Hernández, Eudel; Orellana-García, Arturo
Minería de proceso para la comprensión del proceso en la etapa de diagnóstico
Ingeniería Industrial, vol. XXXIX, núm. 2, 2018, pp. 146-159
Facultad de Ingeniería Industrial, Instituto Superior Politécnico José Antonio Echeverría, Cujae.

Disponible en: <https://www.redalyc.org/articulo.oa?id=360458488005>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UNAM  redalyc.org

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto



Minería de proceso para la comprensión del proceso en la etapa de diagnóstico

Process mining for process comprehension on diagnosis stage

Damián Pérez-Alfonso, Raykenler Yzquierdo-Herrera, Eudel Pupo-Hernández, Arturo Orellana-García

Universidad de las Ciencias Informáticas (UCI), La Habana, Cuba

Correo Electrónico: dalfonso@uci.cu, ryzquierdo@uci.cu, epupo@uci.cu, aorellana@uci.cu

Recibido: 30 de septiembre de 2014

Aprobado: 25 de enero de 2018

RESUMEN

La minería de proceso tiene como objetivo descubrir, monitorear y mejorar los procesos partiendo de los eventos registrados durante su ejecución en los sistemas de información. Sin embargo, las técnicas de minería de proceso disponibles para el diagnóstico de procesos brindan poca información sobre la completitud del registro de eventos o los comportamientos infrecuentes del proceso, afectando su comprensión. Partiendo de las limitaciones de las técnicas disponibles respecto a la comprensión se propone la minería de variantes. Esta técnica permite construir variantes de modelos del proceso, a partir de su descomposición en subprocesos. Se describe la aplicación de la minería de variantes y otras técnicas en un entorno real. Los resultados obtenidos evidencian que la propuesta disminuye la carga cognitiva del modelo e incide positivamente en las dimensiones aptitud, generalización y precisión, al considerar la ausencia de información y el ruido, mejorando la comprensión en el diagnóstico del proceso.

Palabras clave: diagnóstico de proceso, minería de procesos, comprensión del proceso, variantes modelos.

Abstract

Process mining aims to discovering, monitoring and improving process, starting from events logged during process execution on information systems. Nevertheless, available process mining techniques for process diagnosis offer insufficient information about event log completeness or infrequent process behavior, affecting process comprehension. A new technique, named Variants Miner is presented, starting from limitations of process mining techniques regarding process comprehension. This technique aims to construct different variants of process models through process decomposition on sub-processes. An application of Variants Miner and other techniques in a real environment is presented. The obtained results show that the proposal decreases the cognitive load of the model and improve the fitness, generalization and precision, considering missing information and noise, improving process comprehension on process diagnosis stage.

Keywords: process diagnosis, process mining, process comprehension, models variants.capital, management, tendency, challenges.

I. INTRODUCCIÓN

La ejecución de los procesos de negocio en una empresa, habitualmente son registrados por sistemas de información tales como los sistemas de Planificación de Recursos Empresariales (ERP) u otros de propósitos más específicos [1]. Los registros de estos sistemas almacenan las trazas de ejecución de los procesos de negocio y se denominan registros de eventos.

La minería de proceso es una disciplina que impulsa el desarrollo de técnicas y herramientas para analizar los procesos, partiendo de los registros de eventos. Estos análisis tienen como objetivo proveer un mejor entendimiento de los procesos y los elementos para mejorar su rendimiento [2]. Los registros de eventos están conformados por trazas, donde cada traza contiene una secuencia de eventos ordenados según su ocurrencia. Estos reflejan la ejecución de las actividades del proceso al cual pertenece la traza. Asociado a los eventos se registran el instante de tiempo en el cual ocurrió, el recurso que ejecutó la actividad y datos relevantes del negocio [3].

Las técnicas de minería de proceso analizan automáticamente la información contenida en los registros de eventos, ofreciendo resultados en menor tiempo y con mayor fiabilidad que las técnicas tradicionales de análisis de procesos, lo cual ha aumentado el interés en esta disciplina [4]. La minería de proceso es utilizada en diferentes etapas de la gestión de procesos de negocio, incluyendo el diagnóstico, la ejecución, el ajuste y la reconfiguración [5]. El diagnóstico del proceso ayuda a tener una visión general del proceso, de los aspectos más significativos del mismo y de las técnicas que pueden ser más útiles en su posterior análisis [6]. Esta etapa de la gestión de procesos de negocio abarca el análisis de rendimiento, la detección de anomalías, la identificación de patrones comunes y su contextualización [7]. La comprensión del proceso en la etapa de diagnóstico es fundamental, debido a que es la primera dentro de la gestión de procesos de negocio.

La comprensión de los procesos, a través de los modelos que los representan, depende de las características del modelo y de las características del usuario que interpreta el modelo [8]. La comprensión del modelo debe ser entendida como una escalera [9]: la sintaxis (cómo se combinan los elementos gramaticales en un modelo de proceso) debe estar clara antes de que la semántica pueda ser discutida y la semántica (qué significan los elementos gramaticales del modelo de proceso) debe estar clara antes de discutir los elementos pragmáticos. Por tanto, lograr la comprensión del proceso en la fase de diagnóstico requiere que se representen adecuadamente los aspectos relevantes (sintaxis), que estos sean veraces (semántica), y conduzcan a interpretaciones y acciones prácticas (pragmática).

Se han desarrollado un conjunto de técnicas de minería de proceso útiles en la etapa de diagnóstico. Sin embargo, determinadas características de los registros de eventos condicionan el impacto de estas técnicas en la comprensión del proceso. Las dimensiones y/o la complejidad del registro de eventos, así como el número de tareas involucradas provocan que las representaciones de las técnicas Dotted Chart Analysis [10] y Trace Alignment [7] sean complejas, dificultando la comprensión en su dimensión sintáctica. La Descomposición en Bloques de Construcción [11] al no considerar el ruido presente en el registro de eventos, genera descomposiciones incorrectas de los subprocesos, afectando la comprensión semántica del proceso en este tipo de situaciones. FuzzyMiner [12] maneja adecuadamente el ruido, sin embargo, no considera la ausencia de información en el registro de eventos, lo cual puede conducir a representaciones incorrectas en el modelo que ofrece. Adicionalmente, las técnicas Dotted Chart Analysis, FuzzyMiner y Trace Alignment no identifican los patrones de control de flujo presentes. Al no identificar los patrones se dificulta la ubicación en contexto de las anomalías y patrones comunes que ofrecen, limitando la comprensión en el ámbito pragmático.

De forma general, las técnicas disponibles para el diagnóstico suelen ser afectadas por el ruido, el número de tareas involucradas, las dimensiones y/o la complejidad del registro de eventos [13]. Además, solo permiten una comprensión fragmentada del proceso, desde la perspectiva única de cada herramienta, brindando poca información sobre los comportamientos infrecuentes del proceso o la completitud del registro de eventos.

En este trabajo se identifican los elementos que afectan la comprensión del proceso en la etapa de diagnóstico. A partir de estos elementos se propone una nueva técnica de minería de proceso. La técnica propuesta permite superar limitaciones de las técnicas disponibles, ante la presencia de ruido y/o ausencia de información en los registros de eventos, mejorando la comprensión del proceso.

II.MÉTODOS

La comprensión de un proceso pasa por la comprensión de sus representaciones, es decir, por la comprensión del o de los modelos que lo representan. Por su parte, la comprensión del modelo del proceso es una función de las características del modelo y de las características del usuario que interpreta el modelo [8,14] Este trabajo se centra en las propiedades del modelo de proceso que intervienen en la comprensión.

Para establecer los elementos que determinan la comprensión de un modelo de proceso se ha utilizado la semiótica y la teoría de carga cognitiva. La teoría semiótica establece que la comprensión debe ser entendida como una escalera, incluyendo la sintaxis, la semántica y la pragmática, en orden ascendente [9] La teoría de carga cognitiva distingue entre la carga cognitiva intrínseca y extrínseca. La intrínseca está determinada por la complejidad de la información, es decir, la cantidad de elementos y sus relaciones e interacciones. En el dominio del proceso, la carga intrínseca se refiere a la complejidad del proceso modelado y la carga extrínseca es determinada por la forma en la que la información es representada [15]. Por ende, la carga cognitiva extrínseca está sujeta a las decisiones de diseño que se realizan al describir el proceso en un modelo [16].

Comprensión desde la modelación de procesos

La comprensión ha sido abordada en la literatura como un criterio determinante en la calidad de la modelación de procesos. El *framework* de calidad de modelos SEQUAL[17], se basa en la semiótica, estableciendo como dimensiones de la calidad de un modelo: la sintáctica, semántica y pragmática. La calidad sintáctica se asocia al modelo y al lenguaje de modelado; la calidad semántica relaciona el modelo, el dominio y el conocimiento, y por último, la calidad pragmática relaciona el modelo, la modelación y su habilidad para facilitar el aprendizaje y la acción [18]. La calidad pragmática, según SEQUAL, es la relación entre el modelo y la interpretación de la audiencia del modelo, teniendo como meta la comprensión, atendiendo la habilidad del modelo para facilitar el aprendizaje y la acción. Se proponen, como medios para elevar la calidad pragmática, la capacidad de ejecución, simulación y animación del modelo, así como su transformación y filtrado para presentar abstracciones del modelo desde diferentes puntos de vista [9].

La Guía de Modelación (GOM) es otro *framework* para evaluar la calidad de la modelación y se basa en los principios de "correctitud", claridad, relevancia, comparabilidad, eficiencia económica y diseño sistemático [19]. Utilizando el principio de claridad de GOM y la calidad pragmática de SEQUAL se ha determinado que, desde el punto de vista de las características del modelo, el grado de conexión promedio entre los elementos y la densidad del modelo son relevantes, pero el tamaño constituye el factor primario que incide en la comprensión [20].

Adicionalmente, se han identificado como características de interés para la comprensión, el diseño y uso de los constructores gramaticales [16] el resaltado de colores[21], la división en subprocesos [22] y el estilo gramatical de las etiquetas de texto [23]. Los resultados de estas y otras investigaciones que relacionan los estilos de modelación de procesos por una parte y el entendimiento del modelo, así

como la propensión a errores por otra, han conducido al planteamiento de una guía de siete principios básicos para la modelación de procesos: **7PMG**[24]. En esta guía se recomienda que se debe modelar estructuradamente, usando tan pocos elementos en el modelo como sea posible y descomponer cualquier modelo con más de 50 elementos.

En el caso particular de la comprensión sintáctica del modelo, esta puede ser analizada desde dos ángulos: el rendimiento (en qué medida la interpretación del modelo contribuye a comprender el contenido formal del modelo) y la eficiencia (qué recursos son utilizados para comprender el modelo). Considerando que la semántica del modelo se expresa en sus etiquetas textuales se ha demostrado que la información semántica adicional obstaculiza la comprensión sintáctica [8]. Esto se debe a que una etiqueta de texto larga, implica una mayor carga cognitiva con respecto a una única letra al utilizar una mayor parte de la memoria de trabajo para procesar la información textual y la información de dominio que representa.

En la comprensión semántica, por otra parte, influye la selección del idioma utilizado para las etiquetas textuales presentes en el modelo. El impacto de este factor sobre la comprensión semántica es significativamente superior al que genera la selección de la notación de modelado [25].

Comprensión desde la minería de proceso

Las investigaciones sobre comprensión, en el ámbito de la modelación, asumen que el modelo refleja fielmente el proceso, de tal forma que la interpretación del modelo conduce a la comprensión del proceso. Sin embargo, la experiencia en la aplicación de técnicas de minería de proceso indica, que los modelos construidos con técnicas tradicionales de modelación suelen presentar diferencias notables con la ejecución real del proceso que representan [5]. Estas diferencias son expresión de la veracidad del modelo e impactan los niveles semántico y pragmático de la comprensión del proceso.

En la minería de proceso, la relación entre un modelo de proceso y las trazas de su ejecución es considerada un indicador de la calidad del modelo. Este criterio se utiliza para evaluar la calidad de los modelos construidos a través de algoritmos de descubrimiento de procesos [26], donde el objetivo es lograr un balance entre cuatro dimensiones [2]:

- **Aptitud:** el modelo debe permitir el comportamiento presente en el registro de eventos.
- **Precisión:** el modelo no debe permitir comportamiento completamente sin relación con el que se puede apreciar en el registro de eventos.
- **Generalización:** el modelo debe generalizar el comportamiento observado en el registro de eventos.
- **Simplicidad:** el modelo debe ser tan simple como sea posible, reflejando el comportamiento del proceso.

Las dimensiones *aptitud*, *precisión* y *generalización* se refieren a la alineación entre el modelo y la ejecución real del proceso. Por otra parte, la *simplicidad* considera la carga cognitiva que implican los modelos complejos. Para cada dimensión se utilizan métricas, que son aplicables sobre modelos cuya notación sea Redes de Petri [27].

El enfoque utilizado en la minería de proceso para analizar la comprensión del proceso, complementa el de la modelación, al incluir la alineación entre el modelo y la realidad. No obstante, aun cuando considera aspectos de representación en la dimensión *simplicidad*, descarta otros ya mencionados de formato, estructuración y presentación del modelo, que también inciden en la comprensión del proceso.

Comprensión en la etapa de diagnóstico del proceso

Considerando los elementos mencionados, sobre los enfoques utilizados para analizar la comprensión en la modelación y la minería de proceso, se ha utilizado en esta investigación un enfoque híbrido. Esta propuesta combina los objetivos del diagnóstico y los factores que influyen en la comprensión del proceso, ya sean relativos al modelo o a su alineación con la ejecución real del proceso. Por tanto, se

considera que las herramientas de minería de proceso, para mejorar la comprensión del proceso en la fase de diagnóstico, deben poseer las características contenidas en la tabla 1.

Tabla 1. Características deseadas para la comprensión del proceso en el diagnóstico

	Características	Factores para la comprensión
1	Identificación de patrones de control de flujo	Constructores gramaticales / Aptitud
2	Vista jerárquica del proceso	Separación en subprocesos
3	Vista global del proceso	Complejidad del modelo
4	Visualización y manejo del ruido	Complejidad del modelo / Precisión - Generalización
5	Visualización y manejo de la ausencia de información	Complejidad del modelo / Precisión - Generalización

Se considera la identificación de los patrones de control de flujo, ya que estos constituyen elementos fundamentales dentro del proceso porque definen el orden en que las actividades son ejecutadas. También constituyen puntos de referencia para la identificación de anomalías y desviaciones, el manejo del ruido y la ausencia de información. Adicionalmente, pueden ser utilizados para la descomposición en subprocesos y la identificación de comportamientos comunes. Al ser constructores gramaticales, su identificación y diseño tienen un impacto en los niveles sintáctico y semántico de la comprensión del proceso. Su correcta identificación impacta además en la dimensión aptitud.

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

La complejidad del modelo es otro elemento que impacta directamente en la comprensión. Obtener una vista global del proceso disminuye la complejidad del modelo porque permite la abstracción de los patrones específicos concentrándose en los comportamientos y patrones generales y de más alto nivel.

Se incluyen además la visualización y manejo del ruido y la ausencia de información, ya que definen la cantidad de relaciones entre las actividades que se muestran. Esto puede conducir a modelos con muchos nodos y alta densidad o a modelos sencillos pero que no representan adecuadamente al proceso. Por tanto, estas características inciden directamente en las dimensiones de calidad *precisión* y *generalización*.

Técnica propuesta. Minería de variantes

Para mejorar la comprensión en la etapa de diagnóstico se propone una nueva técnica, nombrada minería de variantes. Al igual que el resto de las técnicas para el diagnóstico utiliza como entrada un registro de eventos. Sin embargo, la técnica desarrollada sigue un enfoque diferente a las técnicas antes referenciadas, al proponer descomposiciones alternativas para el mismo subproceso. Esto permite controlar el impacto del ruido y la ausencia de información en la construcción de las alternativas, al descartar determinados comportamientos presentes en el registro de eventos o considerar determinados comportamientos ausentes.

Las diferentes alternativas de descomposición que pueden existir en cada subproceso conforman variantes del proceso. Las variantes de un modelo de procesos o variantes del proceso, son modelos que describen el mismo proceso de negocio, y poseen algunas diferencias estructurales. Estas diferencias pueden estar dadas por los patrones de control de flujo que se utilizan en secciones equivalentes del proceso y la presencia de determinadas actividades, tal como se puede apreciar en la Fig.1.

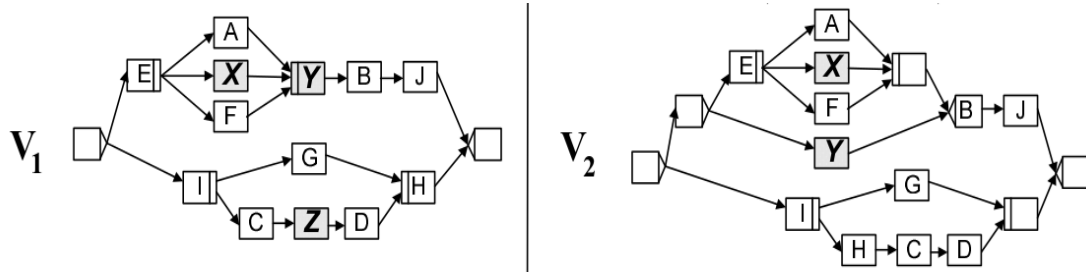


Fig.1. Variantes del proceso [28]

Los elementos antes mencionados permiten que la técnica ofrezca un modelo jerárquico en forma de árbol, que representa diferentes variantes del proceso cuya ejecución está contenida en el registro de eventos. Para esto se propone un tipo de modelo de proceso denominado *árbol de variantes*. En este modelo por cada subproceso se obtienen tantas variantes como posibles descomposiciones se hayan identificado para el subproceso. Dentro de una variante, un nodo representa una descomposición del subproceso al que se refiere, de acuerdo con un patrón de control de flujo y posee como hijos los subprocesos que se obtienen por dicha descomposición. Las hojas están conformadas por las actividades codificadas o por subprocesos para los cuales no se obtuvo ninguna variante de descomposición.

En la Fig. 2 se muestran las variantes de descomposición para un mismo subproceso, que forman parte de un *árbol de variantes*. La primera variante corresponde al patrón de control de flujo selección no exclusiva. La segunda variante representa una descomposición por el patrón *secuencia*. Por último, se muestra la descomposición del subproceso utilizando el patrón *lazo*. Las hojas representan las actividades que conforman el proceso. Los nodos muestran el patrón de control de flujo utilizado para descomponer el subproceso en la variante correspondiente. Cada nivel en el árbol indica un nivel de abstracción.

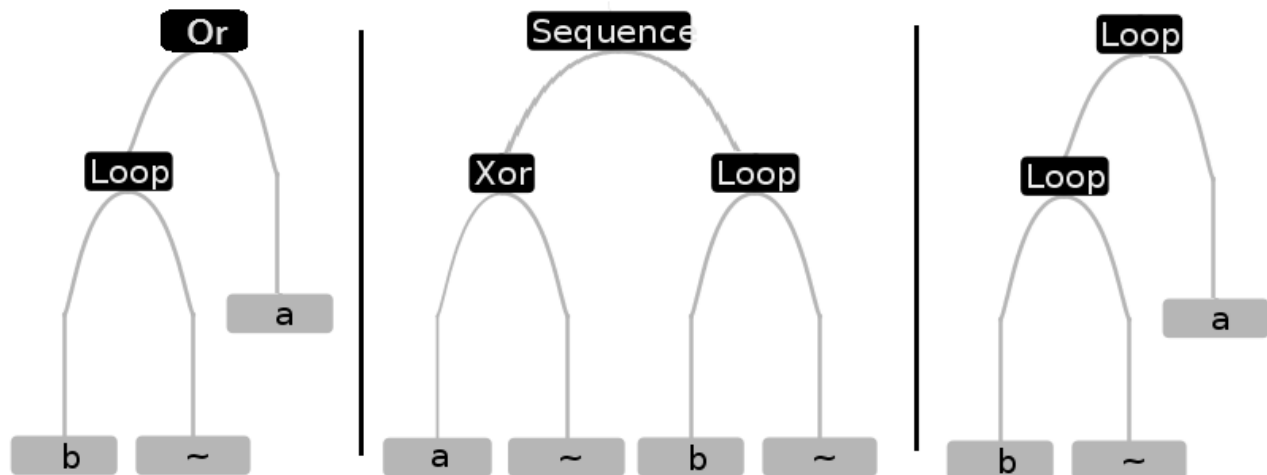


Fig. 2. Variantes contenidas en un árbol de variantes

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

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

Etapas de la técnica propuesta

1. Preprocesamiento del registro de eventos.

En esta etapa realiza la codificación de los eventos, asignando una letra a cada evento contenido en el registro de eventos. La codificación tiene como objetivo disminuir la carga cognitiva durante la comprensión sintáctica del proceso, al eliminar la información semántica de las etiquetas de los eventos. Posterior a la codificación se unifican las trazas cuyas secuencias coinciden, disminuyendo el número de trazas a procesar y agrupando los comportamientos coincidentes. Las secuencias resultantes de la agrupación son ordenadas descendientemente, por su frecuencia de aparición, para expresar su importancia relativa dentro del proceso.

2. Extracción de comportamiento.

El propósito de esta etapa es la extracción de comportamientos representativos de los patrones de control de flujo, a partir del registro de eventos preprocesado. Se considera comportamiento todas las sucesiones directas e indirectas entre actividades, así como los eventos repetidos y los que inician y terminan trazas. Por cada comportamiento extraído se obtiene la frecuencia de su aparición en las trazas codificadas. Esta frecuencia es utilizada para establecer cuán representativo es un comportamiento del proceso que lo genera.

3. Obtención de variantes.

Las variantes se obtienen a partir de la combinación de los comportamientos extraídos en la fase anterior y la asunción de ciertos comportamientos ausentes del registro de eventos. En esta fase se ejecutan dos algoritmos iterativamente, uno para la descomposición en subprocesos y otro para la identificación de patrones de control de flujo. Las posibles variantes de descomposición de un subproceso se determinan dentro de un ámbito definido en términos de umbrales para ruido y completitud o ausencia de información. A partir de un registro de eventos y un ámbito determinado se genera el *árbol de variantes*. La identificación de los patrones de control de flujo se realiza combinando comportamientos presentes en el registro de eventos y comportamientos asumidos bajo el umbral de ausencia de información. Para encontrar la combinación de comportamientos que puede conducir a la descomposición de cada subproceso por un determinado patrón de control de flujo se utiliza un algoritmo basado en búsqueda de costo uniforme.

4. Generación del perfil de diagnóstico.

El algoritmo implementado en esta etapa recorre el árbol de búsqueda generado y recopila los comportamientos considerados como ruido y ausencia de información durante la identificación de los patrones. A partir de estos comportamientos asumidos o descartados se establecen la estimación de *aptitud y precisión*. El perfil se conforma con la información de diagnóstico de cada nodo en el *árbol de variantes* obtenido.

III.RESULTADOS

Para la comparación integral de las técnicas de diagnóstico, a partir de su aporte a la comprensión del proceso, se analizó un proceso que pertenece a un entorno real. Para esto se seleccionó un registro de eventos correspondiente al proceso *solicitar producto*, perteneciente al módulo *almacén* del sistema XAVIAHIS. El Sistema de Información Hospitalaria XAVIAHIS permite la recolección, almacenamiento, procesamiento, recuperación y comunicación de información de atención al paciente para todas las actividades relacionadas con una institución de salud.

El registro de eventos utilizado recopila 1212 ejecuciones del proceso en un lapso de año y medio y un total de 2400 ocurrencias de las 8 actividades que conforman el proceso. El análisis del modelo de

referencia, presentado en la

Fig. 3, permite identificar múltiples actividades de fin del proceso y lazos en diferentes puntos. Después del inicio aparece el nodo de tipodecision, *tipo_solicitud*, expresando una *selección* exclusiva entre cuatro alternativas. La tercera no produce evidencias en el registro de eventos al vincular el evento *tipo_solicitud* directamente con el fin del proceso. Las restantes alternativas corresponden a los subprocesos *ver_detalle_sol_licitacion*, *ver_detalle_sol_almacen* y *ver_detalle_sol_bq*.

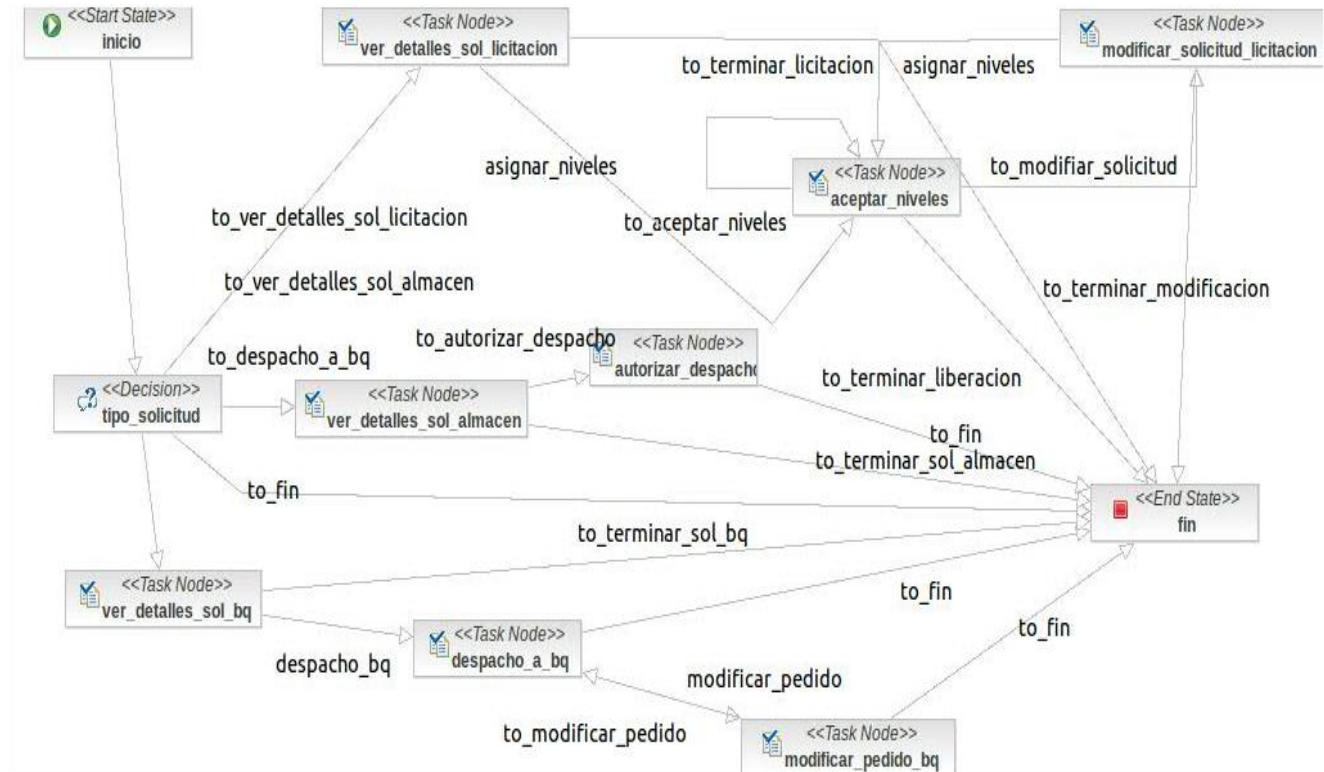


Fig. 3. Modelo del proceso solicitar producto

Aplicación de la técnica propuesta

La aplicación de la minería de variantes, permite la obtención de dos variantes del proceso, como se muestra en la Fig. 4. Las variantes obtenidas corresponden a las descomposiciones mediante los patrones de control de flujo *selección exclusiva* y *secuencia*, respectivamente.

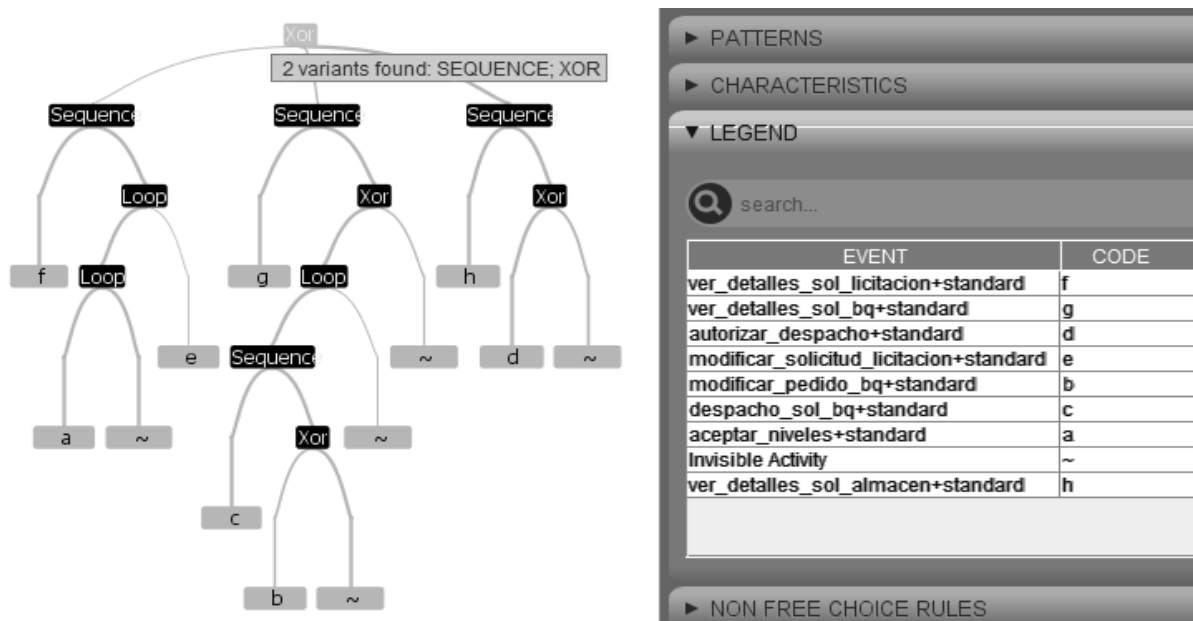


Fig. 4. Variante del proceso mediante la descomposición por selección exclusiva

En la primera variante se identifica una selección exclusiva entre tres subprocesos, esto es coherente con el modelo de referencia de la

Fig. 3. En el *perfil de diagnóstico* asociado a esta variante se obtuvo una evaluación de 0.0 para el ruido y 1.0 para la completitud, corroborando que en el registro de eventos se expresa el comportamiento del patrón de control de flujo *selección exclusiva* en su totalidad y no se ha descartado ningún comportamiento. Además, en el *árbol de variante* de la Fig. 4, se refleja en el grosor de la última arista del primer nivel de abstracción, que la alternativa más ejecutada del proceso es la del subproceso, conformado por las actividades *ver_detalle_sol_almacen* y *autorizar_despacho*, codificadas como *h* y *d* respectivamente.

En la primera sección del *árbol de variantes* mostrada en la Fig. 5, se identifica una *secuencia* entre la actividad *ver_detalle_sol_almacen* y un subproceso descompuesto mediante *selección exclusiva*. La traza que contiene únicamente a la actividad *autorizar_despacho* representa una situación de ausencia de información para el patrón de control de flujo *secuencia*. La técnica identifica esta situación e inserta una actividad invisible, lo que permite la identificación del patrón de *selección exclusiva*. En el grosor de las aristas del subproceso cuya descomposición se realizó mediante *selección exclusiva* se aprecia que la alternativa que contiene la actividad invisible es menos frecuente, lo cual es coherente con la frecuencia de las trazas en cuestión.

En la segunda sección de la Fig. 5 se presenta el fragmento del *árbol de variantes* que contiene la *secuencia* entre *ver_detalle_sol_licitacion* (codificada como *f*) y un subproceso compuesto por *modificar_solicitud_licitacion* y *aceptar_niveles* (codificadas como *e* y *a* respectivamente). Esta *secuencia* indica que *ver_detalle_sol_licitacion* no puede cerrar la ejecución del proceso. Por su parte, el *lazo* representado en esta sección expresa que toda repetición comienza y termina con el subproceso que contiene a la actividad *aceptar_niveles*. Por tanto, se concluye que las alternativas de terminar el proceso después de *ver_detalle_sol_licitacion* o de *modificar_solicitud_licitacion*, no han sido ejecutadas en el ámbito temporal que abarca el registro de eventos. Esto contradice lo expresado en el modelo de referencia de la

Fig. 3, pero es coherente con las trazas del registro de eventos.

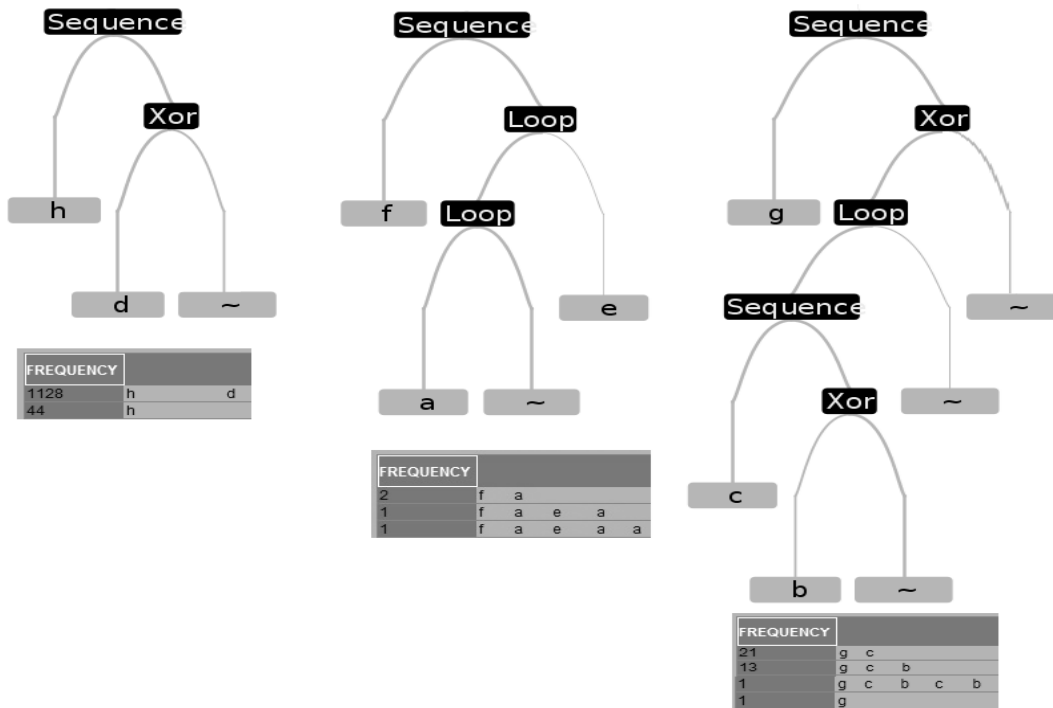


Fig. 5. Secciones del árbol de variantes para cada subprocesso, con sus respectivas trazas de eventos codificados

En el modelo de referencia se muestra un *lazo* que involucra a *despacho_sol_bq* y *modificar_pedido* además, de una *secuencia* entre *ver_detalles_sol_bq* y un subprocesso compuesto por el resto de las actividades antes mencionadas. En la tercera sección de la Fig. 5, se puede observar que el *árbol de variantes* coincide con el modelo de referencia y con el registro de eventos. La técnica propuesta identifica una situación de ausencia de información en la *secuencia* que se inicia con *ver_detalles_sol_bq* (codificada como *g*) y en la que se inicia con *despacho_sol_bq* (codificada como *c*). En ambos casos se inserta una actividad invisible. Se evidencia en el grosor de la arista que representa la alternativa de terminar el proceso con *ver_detalles_sol_bq*, que esta ejecución es muy poco frecuente, lo cual sugiere una anomalía en el proceso.

Comparación con otras técnicas de minería de proceso para el diagnóstico

En la Fig. 6 se muestra el resultado de la aplicación de la técnica alineación de trazas [7] al registro de eventos del sistema XAVIAHIS. En la alineación de trazas las actividades codificadas con *h* y *e* nunca coinciden en las mismas trazas con *c*, *f* y *g*. Esto también sucede con el grupo formado por las actividades codificadas como *a*, *b* y *d* respecto a los grupos antes mencionados. La situación descrita sugiere una *selección exclusiva* entre tres subprocessos, lo cual es coherente con el modelo de referencia. Sin embargo, la identificación en la alineación de comportamientos que sugieran determinados patrones de control de flujo, se complica al no tener una vista jerárquica del proceso.

11397 (12)	--c-----f g
11339 (20)	--c-----f-
4411	--c-----f g f g
4039 (1)	--ba-----
3280	--c-----
100564 (112)	he-----
110862 (43)	h-----
3801	--badaa----
3462	--bada----

Activity Name	Char Encoding
despacho_sol_bq-standard	f
modificar_pedido_bq-standard	g
modificar_solicitud_licitacion-standard	d
autorizar_despacho-standard	e
ver_detalles_sol_licitacion-standard	b
ver_detalles_sol_bq-standard	c
aceptar_niveles-standard	a
ver_detalles_sol_almacen-standard	h

Fig. 6. Alineación de trazas y su codificación, relativas al proceso solicitar producto

La ejecución de la minería difusa [12] sobre el registro de eventos en cuestión genera el mapa de proceso de la Fig. 7, donde se aprecian tres subprocesos, lo cual está alineado con las tres alternativas de ejecución del proceso. Sin embargo, no se muestra la relación entre estos subprocesos y no se brinda información sobre la alta relevancia de la alternativa que incluye las actividades *ver_detalles_sol_almacen* y *autorizar_despacho*, ya que el 96 % de las trazas siguen este camino.

Esta técnica no considera la información ausente en el registro de eventos, ni reconoce patrones de control de flujo, por lo que no se modela la alternativa de que únicamente se ejecute *ver_detalles_sol_almacen* y el proceso termine sin ejecutar la actividad *autorizar_despacho*, lo cual sucede en 44 ejecuciones del proceso contenidas en el registro de eventos. Por otra parte, no se modela que todas las ejecuciones del subproceso *ver_detalles_sol_licitacion* terminan en la actividad *aceptar_niveles*, comportamiento presente en el registro de eventos.

En el mapa obtenido no se puede identificar que el registro de eventos posee ejecuciones del proceso que terminan frecuentemente en la actividad *despacho_sol_bq*. A diferencia del *árbol de variantes*, en el mapa obtenido tampoco se refleja la anomalía que representa la ejecución del proceso que contiene únicamente a la actividad *ver_detalles_sol_bq*. Además, se muestra una relación de repetición sobre la actividad *despacho_sol_bq*, que puede interpretarse erróneamente como un *lazo* que se inicia y termina con esa actividad. Sin embargo, ninguna traza contiene un comportamiento de este tipo, ni el modelo de referencia contempla esta alternativa.

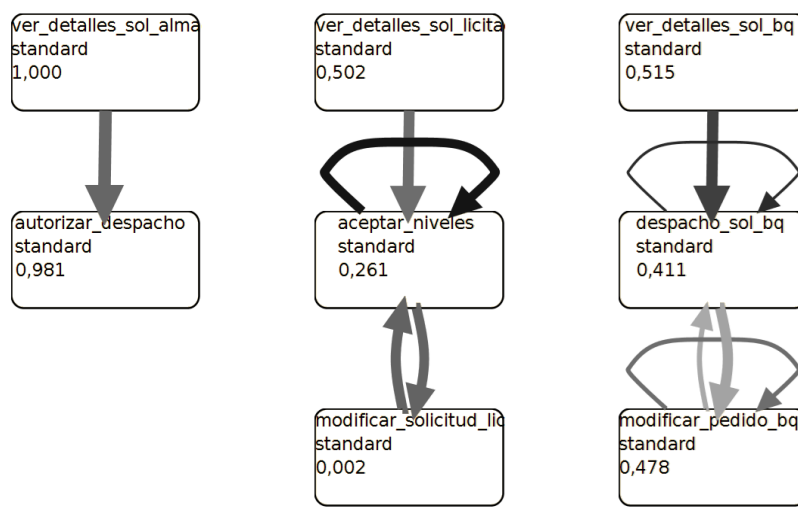


Fig. 7. Mapa del proceso solicitar producto

En la Fig. 8 se muestra el resultado de la aplicación de la técnica descomposición en subprocesos al registro de eventos proveniente del sistema XAVIAHIS. En el primer nivel de descomposición se identifica una selección exclusiva. Las alternativas involucradas corresponden a los subprocesos *ver_detalles_sol_bq* (BB_2_2), *ver_detalles_sol_licitacion* (BB_2_3) y *ver_detalles_sol_almacen* (BB_2_4). Este resultado es coherente con el modelo de referencia del proceso. El primer subproceso presenta diferencias visuales respecto a los resultados obtenidos mediante la aplicación de la técnica propuesta, esto se debe a que para algunos patrones de control de flujo no se manejan los mismos conceptos.

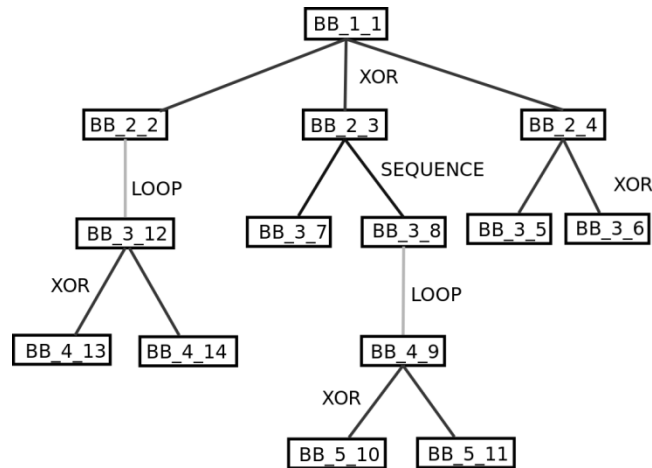


Fig. 8. Descomposición en bloques de construcción

En el subproceso *ver_detalles_sol_licitacion* existe una incongruencia en la identificación del patrón de control correspondiente al último nivel. En este caso mediante la minería de variantes se identifica un *lazo* y la descomposición en bloques de construcción determina una *selección exclusiva*. Debido a que en ambos casos el subproceso identificado en el nivel anterior se descompuso mediante el patrón de control de flujo *lazo* ambas variantes explican el comportamiento presente en el registro de eventos. No obstante, la descomposición propuesta por la minería de variantes tiene una mayor precisión al explicar el comportamiento presente en el modelo de referencia, sin admitir comportamiento adicional.

IV.DISCUSIÓN

La aplicación de las técnicas disponibles para el diagnóstico evidenció sus respectivas limitaciones para la comprensión del proceso. En la alineación de trazas aunque se ofrece la frecuencia de ejecución de

las trazas, esta no se jerarquiza, lo cual dificulta la identificación del comportamiento que puede ser considerado como ruido. Adicionalmente, cualquier inferencia se basa únicamente en el comportamiento contenido en el registro de eventos, ya que no se maneja la ausencia de información. Ambas situaciones limitan la comprensión en las dimensiones *precisión* y *generalización*.

Un problema similar existe con la minería difusa, ya que el modelo obtenido no representa todas las ejecuciones reales del proceso. En este caso se afecta además la comprensión en la dimensión *aptitud*. De forma general, al no identificar formalmente los patrones de control de flujo, el modelo obtenido puede generar interpretaciones incorrectas del proceso, afectando su comprensión sintáctica y semántica.

La aplicación de la técnica descomposición en bloques de construcción evidencia el impacto positivo que genera en la comprensión del proceso la identificación de los patrones de control de flujo y una vista jerárquica del proceso. No obstante, la existencia de ruido y la representación del patrón de control de flujo *lazo* conducen, en ciertos contextos, a la identificación de patrones erróneos. Esto afecta la comprensión en la dimensión *precisión*, al crear un modelo que, en ciertos elementos, se distancia de la ejecución real del proceso.

Por otra parte, la minería de variantes incide positivamente en las perspectivas de generalización y *precisión* de la comprensión, al manejar la ausencia de información e identificar correctamente los patrones de control de flujo. El *árbol de variantes* obtenido proporciona una vista jerárquica del proceso, posibilitando el análisis de cada uno de los subprocesos que componen el proceso por separado y de acuerdo con diferentes niveles de abstracción. Esta jerarquía disminuye la carga cognitiva del modelo mediante su división en secciones. Mientras que la codificación de las

actividades facilita la comprensión a nivel sintáctico, mediante la abstracción de la semántica asociada a las etiquetas de las mismas. Adicionalmente, la representación de las frecuencias de ejecución de los caminos dentro del proceso, en el grosor de las aristas, brinda información sobre la aptitud del modelo, minimizando la carga cognitiva adicional.

V.CONCLUSIONES

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

1. La identificación de patrones de control de flujo, manejo y visualización del ruido y la ausencia de información, así como la obtención de una vista global y jerárquica del proceso son características que impactan en la comprensión del proceso durante su diagnóstico.
2. El perfil de diagnóstico obtenido mejora la comprensión del proceso al ubicar en contexto el comportamiento considerado como ruido y el asumido como ausente, durante la identificación de los patrones de control de flujo.
3. Los resultados obtenidos al aplicar la minería de variantes y otras técnicas de diagnóstico, sobre un registro de eventos proveniente de un entorno real, permiten afirmar que la propuesta mejora la comprensión en el diagnóstico de procesos. 📄

VI.REFERENCIAS

1. Hendricks KB, Singhal VR, Stratman JK. The impact of enterprise systems on corporate performance: A study of ERP, SCM, and CRM system implementations. *Journal of Operations Management*. 2007 2007.25:65–82. ISSN 02726963. DOI 10.1016/j.jom.2006.02.002
2. van der Aalst WMP. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*: Springer Science & Business Media; 2011 2011-04-05. 355 p. ISBN 9783642193453.
3. Verbeek HMW, Buijs JCAM, Dongen BFv, et al. XES, XESame, and ProM 6. En: *Information Systems Evolution*. Springer Berlin Heidelberg; 2011. p. 60-75. ISBN 978-3-642-17721-7, 978-3-642-17722-4.
4. van der Aalst WMP. *Process mining manifesto*. Clermont-Ferrand2012 2012. ISBN 18651348 (ISSN); 9783642281075
5. van der Aalst WMP. A Decade of Business Process Management Conferences: Personal Reflections on a Developing Discipline. En: *Proceedings of the 10th International Conference on Business Process Management*. Berlin, Heidelberg: Springer-Verlag. p. 1–16. ISBN 978-3-642-32884-8. DOI 10.1007/978-3-642-32885-5_1
6. Yzquierdo-Herrera R, Silverio-Castro R, Lazo-Cortés M, et al. Diagnóstico de proceso basado en el descubrimiento de subprocesos. *Ingeniería Industrial*. 2012 2012.33:133–41. ISSN 1815-5936.
7. Jagadeesh Chandra Bose RP, van der Aalst WMP. Process diagnostics using trace alignment: Opportunities, issues, and challenges. *Information Systems*. 2012 2012.37:117–41. ISSN 03064379. DOI 10.1016/j.is.2011.08.003.
8. Mendling J, Strembeck M, Recker J. Factors of process model comprehension—Findings from a series of experiments. *Decision Support Systems*. 2012 2012.53:195–206. ISSN 0167-9236. DOI 10.1016/j.dss.2011.12.013
9. Krogstie J, Sindre G, Jørgensen H. Process models representing knowledge for action: a revised quality framework. *European Journal of Information Systems*. 2006 2006.15:91–102. ISSN 0960-085X. DOI 10.1057/palgrave.ejis.3000598
10. Molka T, Gilani W, Zeng X-J. Dotted Chart and Control-Flow Analysis for a Loan Application Process. En: *Business Process Management Workshops*. Springer Berlin Heidelberg; 2013. p. 223–4. ISBN 978-3-642-36284-2, 978-3-642-36285-9.
11. Yzquierdo-Herrera R, Silverio-Castro R, Lazo-Cortés M. Sub-process Discovery: Opportunities for Process Diagnostics. En: *Enterprise Information Systems of the Future*. Springer; 2013. p. 48-57. ISBN 3642366104.

12. Günther C, van der Aalst W. Fuzzy mining - Adaptive process simplification based on multi-perspective metrics. En: Business Process Management. Brisbane: Lecture Notes in Computer Science. Springer, Berlin. p. 328–43. ISBN 03029743
13. López-Jiménez R, Pupo-Hernández E, Pérez-Alfonso D, et al. Técnicas de minería de proceso para el diagnóstico de procesos de negocio. En: I Conferencia Científica Internacional UCIENCIA. La Habana, Cuba: Universidad de las Ciencias Informáticas. p. ISBN 978-959-286-026-1.
14. Recker J, Reijers HA, van de Wouw SG. Process Model Comprehension: The Effects of Cognitive Abilities, Learning Style, and Strategy. Communications of the Association for Information Systems. 2014 2014.34:9. ISSN 1529-3181.
15. Kirschner PA. Cognitive load theory: Implications of cognitive load theory on the design of learning. Learning and instruction. 2002 2002.12:1–10. ISSN 0959-4752.
16. Figl K, Recker J, Mendling J. A study on the effects of routing symbol design on process model comprehension. Decision Support Systems. 2013 2013.54:1104–18. ISSN 0167-9236. DOI 10.1016/j.dss.2012.10.037.
17. Lindland OI, Sindre G, Solvberg A. Understanding quality in conceptual modeling. Software, IEEE. 1994 1994.11:42–9. ISSN 0740-7459. DOI 10.1109/52.268955.
18. Mendling J, Reijers HA, Cardoso J. What makes process models understandable? En: Business Process Management. Springer; 2007. p. 48–63. ISBN 978-3-540-75182-3,
19. Becker J, Rosemann M, von Uthmann C. Guidelines of business process modeling. En: Business Process Management. Springer; 2000. p. 30–49. ISBN 978-3-540-67454-2,
20. Reijers HA, Freytag T, Mendling J, et al. Syntax highlighting in business process models. Decision Support Systems. 2011 2011.51:339–49. ISSN 0167-9236. DOI 10.1016/j.dss.2010.12.013.
21. Reijers HA, Mendling J. A study into the factors that influence the understandability of business process models. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on. 2011 2011.41:449–62. ISSN 1083-4427. DOI 10.1109/TSMCA.2010.2087017.
22. Reijers HA, Mendling J, Dijkman RM. Human and automatic modularizations of process models to enhance their comprehension. Information Systems. 2011 2011.36:881–97. ISSN 0306-4379. DOI 10.1016/j.is.2011.03.003.
23. Mendling J, Reijers HA, Recker J. Activity labeling in process modeling: Empirical insights and recommendations. Information Systems. 2010 2010.35:467–82. ISSN 0306-4379. DOI 10.1016/j.is.2009.03.009.
24. Mendling J, Reijers HA, van der Aalst WMP. Seven process modeling guidelines (7PMG). Information and Software Technology. 2010 2010.52:127–36. ISSN 0950-5849. DOI 10.1016/j.infsof.2009.08.004.
25. Recker JC, Dreiling A. The effects of content presentation format and user characteristics on novice developers' understanding of process models. Communications of the Association for Information Systems. 2011 2011.28:65–84. ISSN 1529-3181.
26. vanden Broucke SK, Delvaux C, Freitas J, et al. Uncovering the relationship between event log characteristics and process discovery techniques. En: Business Process Management Workshops. Springer. p. 41–53. ISBN 978-3-319-06257-0.
27. De Weerd J, Vanthienen J, Baesens B, et al. A comprehensive benchmarking framework (CoBeFra) for conformance analysis between procedural process models and event logs in ProM. En: Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on. IEEE. p. 254–61. ISBN 978-1-4673-5895-8. DOI 10.1109/CIDM.2013.6597244.
28. Li C, Reichert M, Wombacher A. Mining business process variants: Challenges, scenarios, algorithms. Data & Knowledge Engineering. 2011.70:409–34. ISSN 0169-023X. DOI 10.1016/j.datak.2011.01.005.