

Universidad de las Ciencias Informáticas

Facultad 7

Procedimiento para la detección de eventualidades
aplicando técnicas de minería de proceso en el Sistema de
Información Hospitalaria xavia HIS

Trabajo de Diploma para optar por el Título de
Ingeniero en Ciencias Informáticas

Autores: Yuliet Oliva Romero

Jandro González García

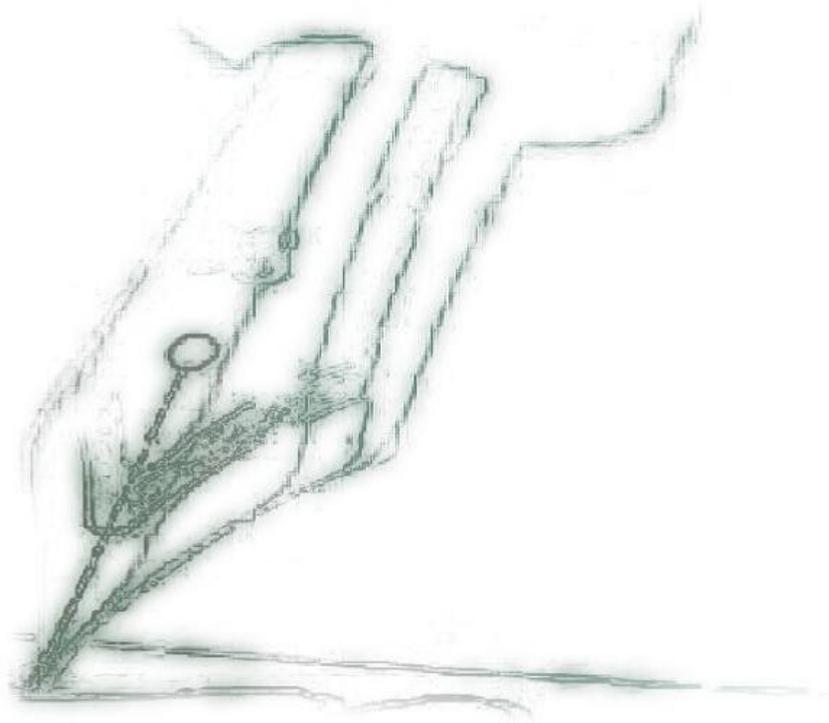
Tutor: MSc. Yovannys Sánchez Corales

Co-Tutor: Ing. Frank Dávila Hernández

Consultante: Ing. Dunia Santos Curbelo

La Habana, Junio de 2013

“Año 55 de la Revolución”



"La ciencia se compone de errores, que a su vez, son los pasos hacia la verdad"

Julio Verne

Jandro

A mis padres Alina y Enrique, por ser mis padres pero también mis amigos, porque no son más pero tampoco menos, porque son, sencillamente, mis padres.

A Jennifer, por quererme sin pensarlo y por tener la difícil tarea de ser la mujer de mi vida.

A mis suegros Bárbara y Manuel, por ser como unos padres para mí.

A mis dos abuelas, Eloina y Mercedes, porque no hay nada mejor que tenerlas a ellas como abuelas.

A mi tío Ezequiel, que se olvidó de ser mi tío para casi poder llamarlo "papá".

A mi primo Dario, por ser un primo, un amigo y un hermano.

A mi primita Aracy, por ser la pequeña de la familia, y la próxima ingeniera.

A mi tía, por estar ahí cuando la he necesitado.

A mi prima Michely porque no dejó que la distancia nos separara.

A mi hermano Idris, que, aunque nunca aparece, siempre me ha apoyado.

A Mirna, mi profesora de toda la vida.

A Laura, mi ingeniera favorita, por toda una vida de tenerla conmigo.

A mi pandilla Darián, Ale y José, gracias por aguantarme 5 años, sé que no fue fácil.

A los profesores que a lo largo de mi carrera me ayudaron a ser mejor.

A mi tutor, por darme la oportunidad de superarme.

A Yuliet, mi compañera de tesis, por dejarme someterla al año más difícil de su carrera.

A mis amigos de la universidad, que siempre los voy a recordar con cariño: Celio, David, Arian, Alejandro García, Reynier Germán y Reynier Arias, Daryel, Yaneslearser, Yisel, Dunia, Zaillet, Dianita, y a todos los que me quedaron, que no me alcanzaría la vida para mencionarlos.

También a Stephen y Fabry por ser amigos sin importar los kilómetros.

A dos seres muy especiales: a mi abuelo Ezequiel, por cuidarme siempre desde allá arriba, y a Odin, por enseñarme que un gran amigo puede tener cuatro patas, EPD.

Y finalmente a mí, por soportarme toda la vida sin explotar.

Yuliet

Agradezco a todas aquellas personas que de una forma u otra aportaron un granito de arena para que este sueño se hiciera realidad.

A la Revolución y a Fidel, por darme la oportunidad de ser mejor cada día.

A mi mamá, Idalmis, por su amor incondicional, dedicación y sacrificio.

A mi papá, Félix, por estar siempre presente, aunque lo vea poco.

A mi tía, Lili, por ser mi ejemplo a seguir.

A mi Harold, por su amor, comprensión, paciencia y por estar presente las 24 horas del día.

A mi mejor amiga, Mayi, por estos 5 años de amistad tan maravillosos.

A mi abuelo, Arsenio, por ser más que abuelo, un padre para mí.

A mi abuela, Cata, por guiarme desde el cielo.

A mis abuelos paternos, Juana y Sergio, por acogerme siempre con cariño.

A mis tías y tíos, en especial a Dori, Kenia, Esperanza, Alexander, Pedri, Chino.

A Rider, por ser como un hermano para mí.

A mis primas y primos, especialmente a Yaima, Yainery, Yasiel y Yasmany.

A mis suegros, Anita y Toni, por acogerme como una hija.

A Alicia, por su cariño y sus consejos. A David R., por su amistad incondicional.

A mi compañero de tesis, Jandro, sin él este trabajo no hubiese sido posible.

A Mirna, por su ayuda en todo momento y las horas frente a este trabajo.

A mi tutor, Yovannys, por darme la oportunidad de superarme.

Al tribunal y oponente por ayudar a graduarme con resultados satisfactorios.

A mis compañeras del apto, por las buenas y las malas, Yisel, Zaillet, Dunia, Yiray, y las que ya no están.

A mis compañeros del aula, en especial a Alex, Darián, Suau, Alejandro, Hectico, Reynier, Yoandry, Daryel.

Jandro

A mis padres Alina y Enrique.

A mi novia Jenni.

A mis suegros Baby y Manolo.

A mi familia a mis amigos (los de verdad, de aquí y de allá) y a todos los que tengan la voluntad y el coraje de seguir adelante y “no aferrarse al queso viejo”.

Yuliet

A mi mamá, Idalmis, por darme la vida, por su cariño diario y sus consejos, por confiar en mí y apoyarme siempre, por ser madre y padre, amiga y compañera, por ser ejemplo de sacrificio. Sin ella este camino hubiese sido muy difícil. Te amo mucho.

A mi familia en general por su ayuda incondicional en todo momento y por depositar en mí confianza suficiente para lograr mis metas. Sin ustedes no hubiera sido posible llegar hasta aquí.

A mi Harold, por ser como es, por apoyarme siempre, por su amor y por ser muy importante para mí.

RESUMEN

La minería de proceso constituye una alternativa novedosa para mejorar los procesos en una variedad de dominios de aplicación. Tiene como objetivo extraer información a partir de los datos almacenados en los registro de trazas de los sistemas de información, en busca de errores inconsistencias y vulnerabilidades.

Actualmente, el Sistema de Información Hospitalaria xavia HIS no posee un procedimiento de análisis y aprovechamiento de los datos almacenados en el registro de trazas, por tanto se desconoce el comportamiento adecuado de los procesos implementados, así como errores de programación que comprometen el desempeño del sistema.

La investigación se centró en el diseño de un procedimiento para la detección de eventualidades mediante la aplicación de técnicas de minería de proceso en el sistema xavia HIS. Se realizó la validación del procedimiento, en el módulo emergencias del sistema xavia HIS, ya que este presenta mayor volumen de datos almacenados.

Se emplearon las herramientas Pentaho Data Integration para la extracción, transformación y carga de los datos; el servidor de bases de datos PostgreSQL, MonoDevelop para el desarrollo de la herramienta de apoyo, JyX para convertir al formato estándar de registro de eventos y el ProM para realizar el análisis con minería de proceso, además se utilizaron los modelos propuestos por van der Aalst, Bozkaya y el caso de estudio de Jans.

La aplicación de minería de proceso en el módulo emergencias posibilitó la detección de errores de programación que comprometen el desempeño del sistema. Se detectó poca consistencia entre los procesos programados y los artefactos de ingeniería, ya que los modelos obtenidos presentan un alto nivel de ajuste (*fitness*) y muestran poca correspondencia con el modelo de negocio.

Palabras clave: Gestión basada en procesos, minería de proceso, registro de eventos, registro de trazas, modelos, sistemas de información.

ABSTRACT

Process mining is a novel approach to improve processes in a variety of application domains. It has as its main goal to extract information from data stored in the trace log of the information systems, looking for inconsistencies, errors and vulnerabilities.

Currently, the Hospital Information System xavia HIS does not have a method of analysis and use of data stored in the trace log, so the proper behavior of the implemented processes is unknown and programming errors that compromise the system performance.

The research focused on the design of a screening procedure for eventualities by applying process mining techniques in xavia HIS system. Validation procedure was performed in the emergencies module of xavia HIS system, as this has a greater volume of stored data.

There were used Pentaho Data Integration tools for the extraction, transformation and loading of data, the database server PostgreSQL, MonoDevelop for the development of support tool, JyX to convert the standard format of the event log and ProM to perform the analysis with process mining, besides the models proposed by van der Aalst, Bozkaya and Jans case study were also used.

The process mining application in the emergency module enabled detection of programming errors that compromise the system performance. It was detected little consistency between scheduled processes and engineering artifacts because the models obtained present a high level of suitability and show little correspondence with the business model.

Keywords: Process based management, process mining, event log, trace log, models, information systems.

TABLA DE CONTENIDOS

INTRODUCCIÓN.....	1
CAPÍTULO 1. FUNDAMENTOS TEÓRICO-METODOLÓGICOS SOBRE MINERÍA DE PROCESO.....	6
1.1 Gestión basada en procesos	6
1.2 Breve reseña de la minería de proceso.....	7
1.3 Minería de proceso	9
1.4 Minería de proceso en el sector de la salud.....	14
1.5 Técnicas de minería de proceso.....	15
1.6 Modelos para aplicar técnicas de minería de proceso	20
1.6.1 Modelo de van Giessel	21
1.6.2 Modelo de Rozinat.....	22
1.6.3 Modelo de Bozkaya	24
1.6.4 Modelo de van der Aalst.....	26
1.6.5 Caso de estudio de Jans	28
1.6.6 Valoración de los modelos anteriores	29
1.7 Investigaciones sobre minería de proceso en la salud.....	30
1.7.1 Aplicación de la minería de proceso en el área de salud.....	30
1.7.2 Descubrimiento automatizado de modelos de flujo de trabajo a partir de datos hospitalarios	31
1.7.3 BPM EHR: de minería de proceso a mejora de procesos a usabilidad de procesos	31
1.7.4 Desafíos de la minería de proceso en los sistemas de información hospitalarios.....	32
1.8 Herramientas seleccionadas.....	33
1.8.1 Pentaho Data Integration.....	33
1.8.2 PostgreSQL	34
1.8.3 ProM	35
1.8.4 MonoDevelop.....	36

CAPÍTULO 2. PROCEDIMIENTO PARA LA DETECCIÓN DE EVENTUALIDADES APLICANDO TÉCNICAS DE MINERÍA DE PROCESO.....	37
2.1 Características del procedimiento	37
2.2 Procedimiento para la detección de eventualidades aplicando minería de proceso.....	39
2.2.1 Fase 1. Planificación y justificación	39
2.2.2 Fase 2. Caracterización del proceso de negocio	39
2.2.3 Fase 3. Preparación del registro de eventos.....	40
2.2.4 Fase 4. Inspección del registro de eventos	45
2.2.5 Fase 5. Análisis del flujo de actividades.....	51
CAPÍTULO 3. APLICACIÓN DEL PROCEDIMIENTO EN EL MÓDULO EMERGENCIAS DEL SISTEMA XAVIA HIS	53
3.1 Módulo emergencias del sistema xavia HIS.....	53
3.2 Aplicación del procedimiento para la detección de eventualidades	53
3.2.1 Fase 1. Planificación y justificación	53
3.2.2 Fase 2. Caracterización del proceso de negocio	54
3.2.3 Fase 3. Preparación del registro de eventos.....	56
3.2.4 Fase 4. Inspección del registro de eventos	60
3.2.5 Fase 5. Inspección del flujo de actividades	66
3.2.6 Detección de eventualidades en el módulo emergencias del sistema xavia HIS	70
CONCLUSIONES.....	73
RECOMENDACIONES.....	74
BIBLIOGRAFÍA.....	75
GLOSARIO DE TÉRMINOS.....	81
ANEXOS.....	82
Anexo 1. Formas de manifestación del ruido	82
Anexo 2. Modelo de van Giessel.....	82

Anexo 3. Modelo de Rozinat	83
Anexo 4. Modelo de Bozkaya	83
Anexo 5. Modelo de van der Aalst.....	84
Anexo 6. Caso de estudio de Jans.....	85
Anexo 7. Entrevista a Ing. Dunia Santos Curbelo	85

INTRODUCCIÓN

El significativo avance de las Tecnologías de la Información y la Comunicación (TIC) a nivel mundial, ha propiciado el desarrollo informático de las organizaciones. En la actualidad, la mayoría de las tareas de oficina, en las empresas o incluso en lugares como bancos, hospitales, centros educativos y agencias de viajes, son ejecutadas y controladas por sistemas de información con el objetivo de recolectar, procesar, almacenar y distribuir datos en apoyo al control y la toma de decisiones.

Las organizaciones actuales reconocen la ventaja que supone incorporar los nuevos avances tecnológicos a sus procesos de negocio. Con el fin de obtener resultados eficientes, constituye una prioridad, mejorar la administración y desempeño organizacional a través de la incorporación de la gestión o enfoque basado en procesos.

Deming plantea que la mejora de la calidad de los procesos implica mayor productividad (Deming, 1989), por tanto se hace necesario romper viejos paradigmas e incorporar nuevos para optimizar los procesos en las organizaciones. Se convierte en el aspecto de mayor importancia el análisis de los datos que se generan durante la ejecución de los procesos.

Mediante el enfoque basado en procesos se pueden identificar y gestionar un gran número de procesos relacionados entre sí y analizar y seguir de forma racional su desarrollo. Permite además, obtener la mejora continua de los resultados por medio de la eliminación de errores y procesos redundantes en las diferentes funciones de la organización. El aumento de la complejidad de este enfoque, enmarcado por el crecimiento del volumen de datos que se almacena continuamente, ha favorecido el surgimiento y desarrollo de tecnologías inteligentes, convirtiendo estos datos en información relevante para la toma de decisiones.

La minería de proceso se propone como una alternativa novedosa para mejorar los procesos en una variedad de dominios de aplicación, desde software de aplicaciones, sistemas de inteligencia artificial hasta aplicaciones web. Es una disciplina de investigación relativamente joven que se ubica entre la inteligencia computacional y la minería de datos, por una parte; y la modelación y análisis de procesos, por otra (van der Aalst, 2011). Surge con el objetivo de analizar los registros de eventos

destinados a recopilar las trazas de los procesos que ocurren en los sistemas de información actuales, en busca de errores, inconsistencias y vulnerabilidades.

Aunque la aplicación de esta tecnología es reciente, (Reijers, van der Aalst, Weijters et al., 2007; Rozinat, Mans, Song & van der Aalst, 2009; van der Aalst, 2011; Hernández, 2012); las empresas, centros e instituciones a nivel mundial la están incorporando a sus aplicaciones con el objetivo de descubrir, monitorear y mejorar sus procesos de negocio. De igual forma, se observa una vinculación a la mayoría de los sectores sociales, principalmente al sector de la salud, que tiene una influencia alta en la calidad de vida de los ciudadanos.

Cuba, desde el año 2003, en el marco del proceso de informatización de la sociedad, ha destinado numerosos recursos humanos y materiales a la creación de centros para el desarrollo de soluciones informáticas, con el propósito de facilitar la prestación de servicios médicos a la población. El Ministerio de Salud Pública (MINSAP) como organismo rector del Sistema Nacional de Salud (SNS) es el encargado de garantizar el acceso a servicios integrales de salud que se prestan en unidades ambulatorias y hospitalarias.

El SNS se divide en tres niveles de atención médica. La Atención Primaria de Salud (APS) (Gérvás, 1987; Martín y Jodar, 2011), donde se solucionan aproximadamente el 80% de los problemas de salud y se ofrecen las acciones de promoción y prevención; la Atención Secundaria que cubre cerca del 15% y la Atención Terciaria que solo abarca un 5% (Domínguez, 2011).

Con el objetivo de informatizar los procesos del nivel secundario se encuentra en proceso de construcción el Sistema de Información Hospitalaria xavia HIS, desarrollándose en la Universidad de las Ciencias Informáticas (UCI) por el Centro de Informática Médica (CESIM). El sistema xavia HIS provee la informatización de los procesos fundamentales de la capa clínica del nivel secundario de atención y está orientado a satisfacer las necesidades de almacenamiento, procesamiento, recopilación e interpretación de los datos médico-administrativos generados. Está compuesto por diversos módulos que responden a procesos, destacándose principalmente admisión, bloque quirúrgico, enfermería, configuración y emergencias.

El módulo emergencias interpreta el proceso del área de Emergencias de los centros asistenciales que reciben a los pacientes para atención médica de forma inmediata. El proceso en el área de

Emergencias comienza cuando llega el paciente y se decide si es una emergencia. Si la respuesta es positiva se procede a recibir al paciente y crear la Hoja de Emergencia, donde se registra la información asociada al mismo. Luego se clasifica el paciente según su estado y se define si necesita interconsulta, ingreso o cirugía.

Para el caso de la interconsulta se realiza la solicitud y se crea la Orden Médica. Si la interconsulta no es necesaria, el proceso concluye con el egreso del paciente, de igual forma que cuando es admitido por ingreso o remitido al quirófano. Durante el proceso, en el módulo emergencias se tiene la posibilidad de actualizar la Hoja de Emergencia.

En el proceso de desarrollo de software del módulo emergencias se describen todos los artefactos generados en el flujo de trabajo análisis y diseño (Modelo del negocio, Descripción de los requisitos y Modelos de análisis y diseño). En estos documentos queda descrito el flujo del proceso a informatizar.

El sistema xavia HIS, como la mayoría de los sistemas de información actuales, tiene implementado una bitácora, registro de trazas o log (en lo adelante, registro de trazas), como vía para recopilar información de la actividad de los usuarios y procesos que ocurren en el sistema. Actualmente no se realiza un procedimiento de análisis y aprovechamiento de los datos almacenados en el registro de trazas, por tanto se desconoce el comportamiento adecuado de los procesos implementados, así como los errores de programación que comprometen el desempeño del sistema.

En el registro de trazas se recoge directamente la actividad de los usuarios que interactúan con el sistema, esto no permite que los procesos queden plasmados de manera explícita y solo se presentan como un resultado en un segundo plano dentro de la actividad del usuario. La información de los procesos se encuentra desestructurada dentro de la información de los usuarios, esto trae como consecuencia poca legibilidad de los mismos y que resulte difícil apreciar con claridad el flujo del proceso implícito en la traza del usuario y realizar comparaciones con el modelo preestablecido. Gestionar las instancias de los procesos se torna una tarea engorrosa, ya que no se pueden adicionar o crear instancias que puedan servir de referencia, en caso de ser necesarias, sin modificar los datos de actividad de los usuarios; para observar la ejecución de alternativas dentro del proceso que permita a los directivos adaptar sus modelos de negocio ante las nuevas exigencias del cliente.

A partir de lo expuesto anteriormente se concluye que la estructura de la registro de trazas del sistema xavia HIS presenta un enfoque basado en sesiones, lo que imposibilita:

1. Apreciar con claridad el flujo de los procesos implementados con respecto a los modelos pre-establecidos.
2. Gestionar las instancias correspondientes a los procesos, de manera que los directivos puedan tomar decisiones.
3. Predecir la ejecución de las alternativas de los procesos.

Después de haber realizado un análisis de la **situación problemática**, se establece el siguiente **problema a resolver**: ¿Cómo interpretar los datos almacenados en el registro de trazas del Sistema de Información Hospitalaria xavia HIS?

Se define como **objeto de estudio** gestión basada en procesos en los sistemas de información hospitalarios y como **campo de acción** técnicas de minería de proceso para la detección de eventualidades en los sistemas de información hospitalarios.

El **objetivo general** consiste en diseñar un procedimiento para la detección de eventualidades aplicando técnicas de minería de proceso, que garantice la adecuada interpretación de los datos almacenados en el registro de trazas del Sistema de Información Hospitalaria xavia HIS.

Para alcanzar el objetivo planteado se proponen las siguientes **tareas de la investigación**:

1. Establecimiento de los fundamentos teórico-metodológicos sobre la gestión basada en procesos y la aplicación de técnicas de minería de proceso.
2. Selección de las herramientas, técnicas y modelos de minería de proceso para la propuesta de solución.
3. Análisis de la arquitectura del Sistema de Información Hospitalaria xavia HIS y de las herramientas actuales para el desarrollo del procedimiento.
4. Implementación de una herramienta que garantice el proceso transformación del registro de trazas.
5. Diseño de un procedimiento para la detección de eventualidades aplicando técnicas de minería de proceso en el Sistema de Información Hospitalaria xavia HIS.

6. Validación de la aplicación de técnicas de minería de proceso a partir de la puesta en práctica del procedimiento en el Sistema de Información Hospitalaria xavia HIS.

Como métodos científicos para el desarrollo de la investigación se utilizan los métodos teóricos histórico-lógico y análisis-síntesis para el estudio bibliográfico sobre la problemática existente y el análisis de las diferentes técnicas de minería de proceso. El método modelación permite la creación de la propuesta de solución. Los métodos empíricos generales se utilizan a través de la observación y los particulares a través de las entrevistas como guía de orientación. Se aplica el análisis documental y bibliográfico para la asimilación y estudio de la documentación del proyecto.

El documento está compuesto por tres capítulos, que incluyen elementos teóricos sobre minería de proceso, aplicación y validación de las técnicas de minería de proceso en el sistema, propuestas en el procedimiento. Además cuenta con Introducción, Conclusiones, Recomendaciones, Referencias Bibliográficas, Glosario de Términos y Anexos.

El **Capítulo 1. Fundamentos Teórico-Methodológicos sobre la minería de proceso**, aborda los conceptos fundamentales sobre minería de proceso: enfoques, tipos, técnicas, perspectivas, modelos y antecedentes; así como las herramientas a utilizar. Además se analizan varias propuestas de diferentes autores sobre minería de proceso aplicada al área de salud.

En el **Capítulo 2. Procedimiento para la aplicación de técnicas de minería de proceso en la detección de eventualidades**, a partir del análisis de los modelos consultados, se describe un conjunto de fases y pasos que guían la utilización de las técnicas y algoritmos de minería de proceso para la detección de eventualidades en los sistemas de información de salud.

Por último, en el **Capítulo 3. Aplicación del procedimiento en el módulo emergencias del sistema xavia HIS** se realiza la validación de la aplicación de las técnicas de minería de proceso, a través de la puesta en práctica del procedimiento en el módulo emergencias del sistema xavia HIS y se muestran los resultados obtenidos.



CAPÍTULO 1. FUNDAMENTOS TEÓRICO-METODOLÓGICOS SOBRE MINERÍA DE PROCESO

En el presente capítulo se abordan conceptos fundamentales sobre minería de proceso: enfoques, tipos, técnicas, perspectivas, modelos y una visión general de las herramientas a utilizar para cumplir las metas de la investigación. Además se analizan varias propuestas de diferentes autores sobre minería de proceso aplicada al área de salud.

1.1 Gestión basada en procesos

El modelo de la Fundación Europea para la Gestión de la Calidad (*European Foundation for Quality Management*, EFQM por sus siglas en inglés) (2003), define un proceso como “sucesión de actividades en el tiempo con un fin definido; organización lógica de personas, materiales, energía, equipos y procedimientos en actividades de trabajo diseñadas para generar un resultado específico (...)”.

La norma ISO 9000 (2005), lo define como “conjunto de actividades mutuamente relacionadas o que interactúan, las cuales transforman elementos de entrada en resultados”; y el Soporte a Procesos Empresariales (*Enterprise Process Support*, EPS por sus siglas en inglés) (n.d.) lo plantea como “conjunto de tareas de trabajo interrelacionadas que se inician en respuesta a un evento y logra un resultado específico para el cliente del proceso”. Los autores de este trabajo toman como definición de proceso, la presentada por la norma ISO 9000.

Para que las organizaciones operen de manera eficaz, tienen que identificar y gestionar numerosos procesos interrelacionados y que interactúan. La identificación y gestión sistemática de los procesos empleados en la organización y en particular las interacciones entre tales procesos, se conocen como enfoque basado en procesos (NC/ISO-9000, 2005).

El enfoque basado en procesos en las organizaciones de información es la forma más eficaz para desarrollar acciones que satisfagan las necesidades de los usuarios internos y externos con información relevante, oportuna y precisa que facilite la toma de decisiones estratégicas y operativas. Además pueden identificarse y gestionarse numerosos procesos interrelacionados, analizar y seguir coherentemente el desarrollo de los procesos en su conjunto, así como obtener la

mejora continua de los resultados por medio de la erradicación de errores y procesos redundantes en las diferentes funciones de la organización.

En un análisis de (Mallar, 2010) se concluye que existen diversos motivos que mueven este enfoque dentro de una organización, entre los cuales se encuentran: extensión del programa institucional de calidad, cumplimiento de legislaciones vigentes, mejoramiento continuo, entender qué se está haciendo bien o mal a través de la comprensión de los procesos, así como la automatización y organización de los mismos.

Durante mucho tiempo, las organizaciones han puesto en práctica la tradicional división de funciones por departamentos; por eso, este nuevo enfoque significa un cambio paradigmático (Moreira, 2006). Actualmente las TIC tienen una influencia vital para la adopción de cualquiera de estos cambios tecnológicos, que las convierten en una herramienta ágil y operativa que permite la mejora en la gestión de los procesos. Por tanto, no se puede prescindir de las tecnologías de la información disponibles para dar soporte a un sistema de gestión con un enfoque basado en procesos.

1.2 Breve reseña de la minería de proceso

Los autores consultados en la literatura proponen la minería de proceso como una alternativa novedosa para mejorar los procesos en una organización. Según (van der Aalst, Adriansyah, Medeiros et al., 2011), la minería de proceso es una tecnología que surge a partir de la minería de datos con el objetivo de analizar los registros de eventos destinados a recopilar las trazas de los procesos que ocurren en los sistemas de información actuales.

Para comprender esta nueva tecnología es necesario conocer su punto de partida: la Gestión de Procesos de Negocio (*Business Process Management*, BPM por sus siglas en inglés), que tiene como objetivo utilizar métodos, técnicas y software para diseñar, ejecutar, controlar y analizar procesos operacionales que involucran personas, organizaciones, aplicaciones, documentos y otras fuentes de información (van der Aalst, ter Hofstede, Arthur & Weske, 2003).

Al realizar un estudio de diversos trabajos (Oracle, 2008; Nolasco, 2008; Rico, 2011) se obtiene que desde el punto de vista de la gestión, BPM es el renacer de la orientación a los procesos de los años 90's, con un fuerte impulso en la orientación al cliente y la mejora continua apoyado por tecnologías orientadas a los procesos. Luego surgen los Sistemas de Administración de Procesos

de Negocios (*Business Process Management Systems*, BPMS por sus siglas en inglés), como plataformas de software que soportan la definición, ejecución y rastreo de procesos de negocios. Un correcto análisis de los *logs* de ejecución de los BPMS puede mostrar importantes conocimientos y ayudar a organizaciones a mejorar la calidad de sus procesos de negocio y servicios. La aparición del BPMS representa quizás la más importante revolución en el dominio del software empresarial, comparable en su entorno a la revolución que ha supuesto la aparición de Internet.

El conjunto de herramientas integradas que dan soporte a negocios y a usuarios informáticos se denomina Inteligencia de Procesos de Negocio (*Business Process Intelligence*, BPI por sus siglas en inglés). Según (Grigori, Casati, Castellanos et al., 2004) BPI se refiere a la “aplicación de varias técnicas de medición y análisis en el área de BPM. En la práctica, BPI está envuelta en herramientas para gestionar la calidad de la ejecución de los procesos al ofrecer varias características tales como análisis, predicción, monitoreo, control y optimización”.

La minería de proceso es una herramienta BPI (Seguel, 2008). Las técnicas de minería de proceso son "inteligentes", ya que el análisis que ellas proveen está basado en registros de eventos y modelos y requieren poca intervención humana para efectuarlo. Desde la fundación del Grupo de Trabajo del Instituto de Ingenieros Eléctricos y Electrónicos (*Institute of Electrical and Electronic Engineers*, IEEE por sus siglas en inglés) en el año 2009, se han realizado talleres y sesiones especiales sobre Inteligencia de Procesos de Negocio (*Business Process Intelligence*, BPI por sus siglas en inglés), tutoriales (*World Congress on Computational Intelligence*, WCCI por sus siglas en inglés), escuelas de verano (*Estonian Summer School in Computer and Systems Science*, ESSCa por sus siglas en inglés), videos y varias publicaciones.

En el año 2010 se denominó XES como formato estándar de registro de eventos. En el 2011 se escribe el primer libro sobre minería de proceso¹ y se realiza el primer Desafío de Inteligencia de Procesos de Negocio (*Business Process Intelligence Challenge*, BPIC'11), donde los participantes extrajeron conocimiento relevante de un complejo registro de eventos. De esta manera se

¹ van der Aalst, W.M.P. *Process Mining Discovery, Conformance and Enhancement of Business Processes*, Berlín Springer-Verlag, 2011.

desarrolla la minería de proceso, la cual sustenta sus objetivos en base a la gestión, donde el análisis de información constituye el punto de partida y el elemento fundamental en la mejora de los procesos.

1.3 Minería de proceso

La minería de proceso es una tecnología relativamente joven, según (van der Aalst, 2011) provee un puente importante entre la minería de datos y la gestión de procesos de negocio (BPM).

El concepto de minería de proceso surgió hace más de una década (Agrawal, Gunopulos, Leymann et al., 1998). La disciplina de minería de proceso también tiene sus raíces en el trabajo de Cook y Wolf, quien propuso el descubrimiento de modelos de procesos a partir de los datos contenidos en los registros de eventos (Cook & Wolf, 1999). Desde entonces ha sido objeto de numerosas investigaciones, y por tanto, aplicada a la mayoría de las ramas de la sociedad, desde ingeniería, ciencias computacionales, educación y salud hasta el desarrollo de software.

El crecimiento de un universo digital que está bien alineado con los procesos en las organizaciones hace posible registrar y analizar eventos (van der Aalst, 2011). Los eventos podrían variar desde el envío de información a través del correo electrónico, un técnico ajustando un equipo de radiología, un ciudadano solicitando un préstamo, hasta la recepción del boleto de un viajero.

El mayor desafío consiste en utilizar al máximo estos datos de eventos de una forma significativa. La minería de proceso se refiere a la extracción de modelos de procesos a partir de un registro de eventos; los modelos obtenidos según van der Aalst pueden ser representados a través de una red de Petri o la Notación de Gestión de Procesos de Negocio (*Business Process Management Notation*, BPMN por sus siglas en inglés), (Rozinat & van der Aalst, 2006). Además proporciona soporte operacional, lo que le permite detectar, predecir y recomendar eventualidades que se presenten en el proceso. Por ejemplo en el momento en que un caso se desvía del proceso predefinido, es detectado y el sistema genera una alerta y sugiere algún método para su solución o prevención (van der Aalst, 2011).

Conceptos fundamentales

A continuación se explican varios términos comprendidos en el área de la minería de proceso. En la **Figura 1.1** se muestra la relación entre ellos, al ser elementos fundamentales de un proceso.

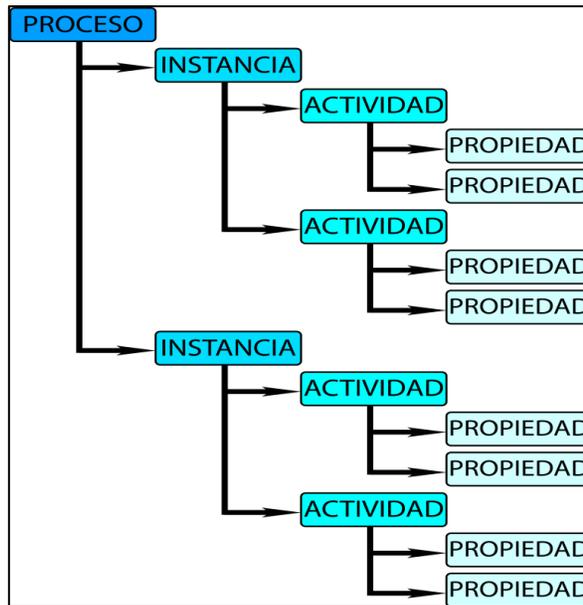


Figura 1.1. Relación entre términos de minería de proceso

Fuente: Elaboración propia

La mayoría de los sistemas de información actuales almacenan los datos de los eventos, o sea poseen un tipo de **registro de trazas**; este contiene la información de los sucesos que ocurren en el sistema. Al consultar la literatura (Jans, 2010), define **instancia** de proceso como caso o sujeto que puede ser sistemáticamente seguido a través del proceso, asociando un identificador único a las actividades que son ejecutadas. Por ejemplo, un paciente en un hospital, una solicitud de trabajo o un préstamo bancario.

Las acciones almacenadas en el registro, por ejemplo, el inicio, conclusión o cancelación de una tarea para una instancia particular de un proceso, se denominan **actividades**. Will van der Aalst destaca que una actividad es un paso bien definido en el proceso.

El registro de eventos contiene una secuencia de entradas en la forma identificación de instancia ($id_instancia$) e identificación de actividad ($id_actividad$); donde $id_instancia$ identifica la instancia de proceso e $id_actividad$ específica la actividad que se ha ejecutado. Cada una de estas actividades está determinada por diversas características o atributos (recursos, costo, tiempo, etc.), denominadas **propiedades**, las cuales se manifiestan en dependencia de las necesidades de almacenamiento en cada organización.

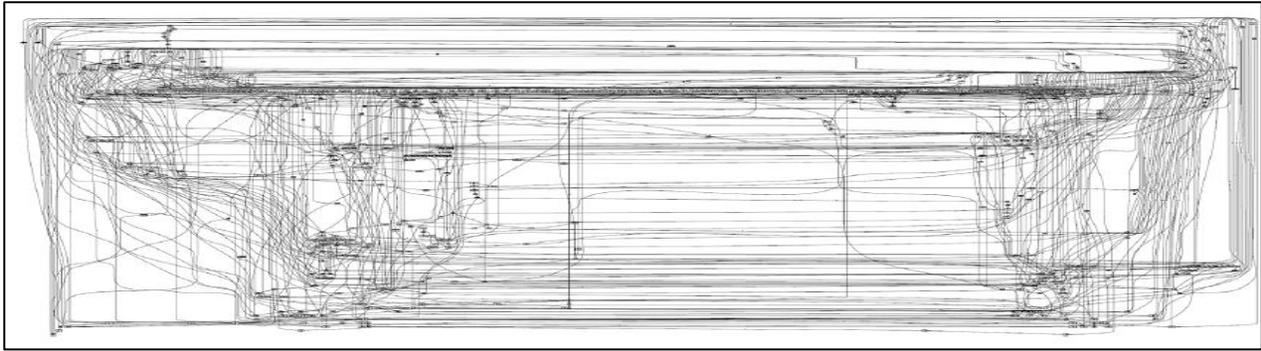


Figura 1.3. Proceso Espaguetis

Fuente: van der Aalst, 2011

En resumen, un Proceso Lasaña o estructurado está determinado por un comportamiento lógico de los eventos, lo que permite una adecuada estructura y por tanto, validez. Un Proceso Espaguetis o desestructurado, define un comportamiento sin un orden definido, sin una estructura clara, ya que de un posible conjunto de opciones se actúa de acuerdo a las necesidades.

Tipos de minería de proceso

Según (van der Aalst, 2011) existen en la actualidad tres tipos de minería de proceso, que se ilustran en la **Figura 1.4**.

Al análisis de un registro de eventos y la producción de un modelo, sin usar ninguna información previa para descubrir los procesos reales, meramente basado en las muestras de ejecución de los registros de eventos se le denomina **descubrimiento**.

El chequeo de **conformidad** relaciona eventos del registro de eventos con actividades del modelo del proceso y las compara (Rozinat & van der Aalst, 2008). La verificación de conformidad puede ser usada para chequear si el modelo es equivalente a la información que está almacenada en el registro de eventos. Se está en presencia de **extensión o mejora** cuando se mejora un modelo de procesos existente utilizando la información acerca del proceso real almacenada en el registro de eventos (Rozinat & van der Aalst, 2006).

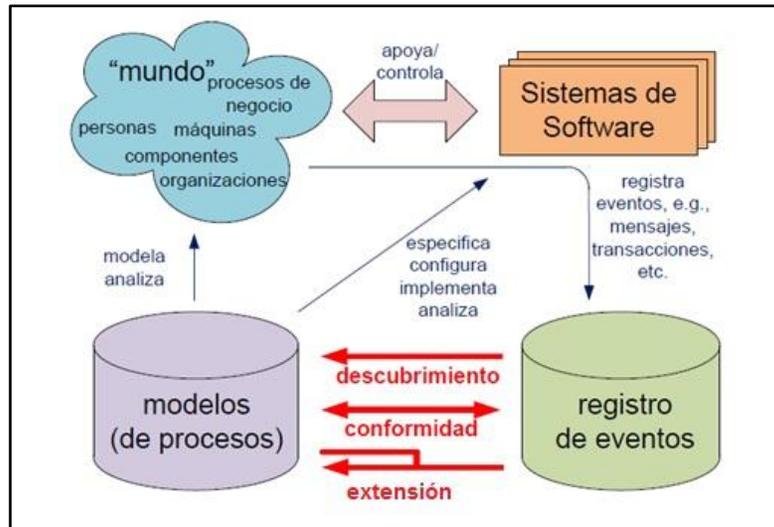


Figura 1.4. Tipos de minería de proceso

Fuente: van der Aalst, 2011

En el mundo, los procesos de negocio, personas, máquinas y organizaciones, están estrechamente relacionados con los sistemas de software. Estos sistemas registran todo tipo de eventos que se almacenan en un registro que es utilizado para aplicar minería de procesos, con la finalidad de descubrir, monitorear y mejorar los procesos a través de modelos, que a su vez contribuyen al análisis de los sistemas de software.

La aplicación del tipo de minería depende de las necesidades particulares de análisis de datos en cada organismo, entidad o empresa. Este trabajo se centra en el descubrimiento, aunque da algunos pasos iniciales en la conformidad.

Perspectivas de la minería de proceso

A partir del análisis de la literatura (Homayounfar, 2012; van der Aalst, 2011; Bratosin, 2011; van der Aalst et al., 2007), se identifican varias perspectivas de análisis con la minería de proceso. La perspectiva de control de flujo (¿Cómo se realiza el proceso?), la perspectiva organizacional (¿Quién realiza el proceso?), la perspectiva de caso (¿Qué propiedades tiene el proceso?) y la perspectiva de tiempo (¿Cuánto demora un proceso?).

- ✓ **Perspectiva de control de flujo:** se centra en el orden de las actividades. Su objetivo es encontrar correctas caracterizaciones de todas las vías posibles, por ejemplo, una red de Petri u otra notación (BPMN y UML).
- ✓ **Perspectiva organizacional:** se enfoca en la información sobre los recursos ocultos en el registro, es decir, qué actores (personas, sistemas, funciones y departamentos) están involucrados y cómo se relacionan. El objetivo es clasificar a las personas en términos de roles y unidad organizacional o bien mostrar la red social.
- ✓ **Perspectiva del caso:** se centra en las propiedades del caso. Los casos se pueden caracterizar por la ruta de ejecución que muestran o por quiénes intervienen y además por los valores de atributos.
- ✓ **Perspectiva del tiempo:** tiene que ver con el tiempo y la frecuencia de los eventos. Cuando los eventos tienen marcas de tiempo, es posible descubrir cuellos de botella, niveles de servicio, medir, monitorear la utilización de los recursos, y predecir el tiempo de procesamiento de los casos restantes que se ejecutan.

Estas perspectivas están dirigidas a orientar al investigador en distintas direcciones en busca de resultados más completos, todo depende del alcance deseado. Se puede descubrir un modelo del proceso a partir del registro de eventos, comprobar su validez y expresividad. En este trabajo se utiliza la perspectiva de control de flujo, pues su objetivo es encontrar irregularidades en cuanto al flujo de actividades específicamente, o sea actividades completas, incompletas, fuera de lugar o de orden. Como resultado se obtiene una caracterización del flujo del proceso.

1.4 Minería de proceso en el sector de la salud

Los nuevos avances tecnológicos están destinados a resolver problemas en los diversos sectores de la sociedad, principalmente en el sector de la salud, el cual tiene como objetivo elevar la calidad de vida de los ciudadanos. El aumento considerable de la visita de pacientes a las áreas de salud en busca de diagnósticos y tratamientos eficientes a sus problemas, ha dado como resultado que las organizaciones en este sector busquen alternativas para mejorar el servicio en términos de calidad, funcionalidad y rapidez (Gupta, 2007).

Los procesos en el área de salud incluyen tareas no solo clínicas sino administrativas, las cuales generan grandes volúmenes de datos acerca de los pacientes y el personal en general. Además están vinculados a disciplinas como la gestión, economía y nuevas tecnologías, o sea, no solo se

generan datos sobre la salud de los pacientes. Por ejemplo, cuando llega un paciente al hospital refiriendo un problema, puede requerir además atención sobre otro problema, lo que trae como consecuencia su interacción con diversas áreas.

La aplicación de minería de proceso se presenta como una tecnología de importancia para su puesta en práctica en el sector de la salud y está determinada por la premisa de que el sistema de información posea un registro de trazas donde se almacene la actividad de los usuarios y procesos. Estos registros se analizan con herramientas inteligentes, como el ProM, y se obtienen gráficas, secuencias y modelos que identifican y caracterizan los procesos ejecutados (Homayounfar, 2012).

Con estos resultados es posible observar los pasos seguidos por los pacientes, conocer las excepciones e identificar cuellos de botella en los procesos, así como, cuáles son las personas o departamentos médicos que trabajan juntos. Esta información permite tomar acciones que traen como consecuencia la mejora de la organización en cuanto a reducción de costos, del tiempo de atención o tratamiento a un paciente, creación de nuevos estándares sobre la forma de trabajo, disminución del nivel de complejidad de los procesos y el fortalecimiento de la colaboración entre las distintas disciplinas médicas (Gupta, 2007).

1.5 Técnicas de minería de proceso

Las técnicas son procedimientos y recursos que se emplean para lograr un resultado específico; son la vía para analizar, diagnosticar, descubrir y mejorar los procesos en las organizaciones. A continuación se presentan las técnicas de minería de proceso de interés y alcance para esta investigación.

- ✓ Inspección del registro de eventos.

Para realizar la inspección es necesario visualizar el registro de eventos antes de aplicar alguna técnica de minería, con el fin de conocer la información almacenada en el registro. Solo se puede responder a determinadas preguntas si los datos están almacenados en dicho registro (Verbeek, 2010).

- ✓ Análisis de Gráfico de Puntos (*Dotted Chart Analysis*, DCA por sus siglas en inglés).

Esta técnica, semejante a los diagramas de Gantt, permite la visualización de las instancias contenidas en el registro de eventos, que son descritas como puntos en un plano en el que una

dimensión hace referencia a casos y la otra al tiempo de ejecución de eventos (Song & van der Aalst, 2007). Tiene como objetivo ayudar en el análisis de rendimiento de procesos al representar los eventos gráficamente y calcular indicadores de desempeño. Es difícil determinar patrones comunes entre casos distintos, dado que la forma en que se grafican determina puntos no alineados, siendo su mayor desventaja.

Además permite identificar problemas de ruido existentes en el registro de eventos. Según (van Arendonk, 2011) se pueden detectar tres formas en las que se manifiesta el ruido (Ver **Anexo 1**):

1. Eventos del sistema que no tienen nada que ver con el proceso y que se incluyen en el registro de eventos. La gran cantidad y los cortos intervalos de tiempo en el que estos eventos suceden crean líneas verdes, en lugar de un número apreciable de puntos.
2. Una sola actividad que tuvo lugar mucho antes que el resto de las actividades en el registro de eventos.
3. Una brecha en el registro donde en un determinado período no hay eventos registrados.

Dado que estas tres formas de ruido son en realidad ocurrencias distintas en los registros de eventos, van Arendonk define términos para tratarlos: a) los casos incompletos son rastros en un registro de eventos (ausencia de eventos de inicio y fin), b) el comportamiento excepcional es comportamiento de baja frecuencia que difiere de la mayoría de las trazas y c) el ruido del sistema como acontecimientos que se registran incorrectamente debido a un error del sistema o que no tienen nada que ver con el proceso realizado.

Al estudiar estos tipos de ruido, van Arendonk plantea que el segundo debe estar presente en el registro de eventos, pues estas desviaciones son un hecho relevante en la vida real, y por tanto, deben tenerse en cuenta en la construcción del modelo del proceso.

✓ Alineación de Trazas (*Trace Alignment*).

La alineación de trazas facilita el problema de determinar patrones interesantes en medianos y grandes *log*, los que se determinan automáticamente y se muestran al usuario. Las trazas alineadas constituyen una representación de las tareas de acuerdo a un orden relativo y su estructuración en casos. El orden establecido entre las tareas permite identificar los patrones de flujo de control que se manifiestan en el proceso (Bose et al., 2012).

✓ Minería Difusa (*Fuzzy Miner*).

Con la aplicación de esta técnica se obtiene un modelo basado en gráficos, capaz de proporcionar una vista de alto nivel de un proceso, con la abstracción de los detalles no deseados. Según (Günther, 2009) está caracterizada por la presencia de dos tipos de nodos; nodos primitivos (se refieren a una tarea) y nodos que se refieren a un conjunto de tareas o clúster.

Además permite agrupar tareas, aunque considera que cada tarea pertenece a un único nodo; y es empleada en el diagnóstico del registro de eventos, donde posibilita realizar análisis preliminares al descubrimiento, basados en la correlación entre las tareas y la importancia de una secuencia.

✓ Filtros.

La aplicación de filtros permite eliminar información redundante del registro de eventos antes del proceso de minado, con el objetivo de analizar solo los datos más importantes. Existen diversos filtros: el filtrado por ocurrencia permite filtrar el registro por fecha de ocurrencia y el filtrado simple, permite marcar directamente los eventos indeseados para ser descartados, tanto por tipos de eventos, como de inicio y fin. Se consideran de gran utilidad, pero según (Verbeek, 2010) presenta como limitación que no se puede cambiar el nombre de las tareas.

En **Tabla 1.1** se muestra un conjunto de características que se identifican a través de las técnicas de minería antes mencionadas.

Tabla 1.1. Características detectadas por las técnicas de minería

Características	Inspección del registro	Gráfico de puntos	Minería difusa	Alineación de trazas
Detección de irregularidades		x	x	x
Ruido		x		x
Paralelismo			x	x
Lazos			x	x
Dependencias				x
Tipo de proceso	x	x		
Tareas duplicadas		x		x

Fuente: Elaboración propia

Algoritmos de descubrimiento

Los algoritmos se clasifican en las siguientes categorías: (van Dongen, Medeiros & Wen, 2009)

- ✓ Algoritmos basados en abstracciones (*Abstraction-based algorithms*): utilizan abstracciones del *log* para derivar modelos de procesos. La abstracción consiste en extraer las relaciones de orden básico entre los eventos. Ejemplo de estos algoritmos son el algoritmo *Alpha* (α *algorithm*) (van der Aalst, Weijters & Maruster, 2004) y el α ++ *algorithm* (Wen, Wang & Sun, 2006), que es la segunda extensión del algoritmo *alpha*, y puede minar redes de Petri. Son algoritmos muy rápidos, poco robustos ante el ruido y la falta de información.
- ✓ Algoritmos basados en heurística (*Heuristic-based algorithms*): toman en cuenta las frecuencias de los eventos y de las secuencias, para la construcción del modelo del proceso. Los caminos infrecuentes no deben ser incorporados al modelo. Estos algoritmos descubren casi todas las estructuras comunes de control (decisión, secuencia, lazos, paralelismo, etc.) y son robustos ante ruidos en el *log* de eventos. Ejemplo: *Minero Heurístico* (Weijters & van der Aalst, 2003) y *Fuzzy Miner* (Günther, 2009).
- ✓ Algoritmos basados en búsquedas (*Search-based algorithms*): utilizan técnicas de búsqueda en el espacio de todas las posibles soluciones para encontrar el modelo del proceso que represente el registro de eventos. Ejemplo: *Genetic Miner* (Medeiros, 2006). Estos algoritmos manejan todos los patrones comunes de los flujos de trabajo y son robustos ante el ruido.
- ✓ Algoritmos basados en lenguaje (*Language-based algorithms*): aplican la teoría de regiones basada en lenguajes para construir una red de Petri compuesta solo de transiciones directamente de un registro de eventos, donde cada transición corresponde a un evento. Entonces, los lugares son agregados para restringir el comportamiento. Ejemplo: *ILP Miner* (van der Werf & van Hee, 2008) utiliza programación lineal entera para encontrar estos lugares que refuerzan las dependencias casuales presentes en el registro de eventos.
- ✓ Algoritmos de descubrimiento de estados (*State-discovery algorithms*): traducen cada traza en una secuencia de estados con transiciones correspondientes a los eventos, pero como los sistemas de transiciones proveen un modelo de bajo nivel, se necesitan regiones basadas en estados para sintetizar a un modelo más compacto (van der Aalst, 2011). La idea central es descubrir regiones que correspondan a lugares de una red de Petri.

Las siguientes características se tienen en cuenta para la realización de la **Tabla 1.2** donde se comparan los algoritmos de descubrimiento y la “x” determina una respuesta positiva en los siguientes aspectos:

Ruido: comportamiento anormal e infrecuente que se registra en el registro de eventos y que no representa comportamiento característico o usual del proceso.

Tareas duplicadas: tareas que se encuentran de forma repetida en un registro de eventos.

Tareas ocultas: representan ejecuciones normales de un proceso que por sus condiciones excepcionales ocurren de forma aislada.

Alternativas no libres: constructores de control de flujo donde las alternativas y la concurrencia coinciden, por tanto una tarea A depende indirectamente de una tarea B; o sea en un punto de división o unión, la elección de un camino u otro puede estar en dependencia de las alternativas escogidas en otra parte de dicho proceso.

Lazos: tarea o conjunto de tareas que se repiten varias veces dentro de un proceso.

Tabla 1.2 Comparación de los algoritmos de descubrimiento

Algoritmo	Ruido	Tareas ocultas	Tareas duplicadas	Alternativas no libres	Lazos
Fuzzy	x	-	-	-	x
Genético	x	x	x	x	x
Heurístico	x	x	-	x	x
Minero ILP	-	-	-	-	-
α	-	-	-	x	x

Fuente: Elaboración propia

Luego del análisis de la comparación mostrada en la tabla anterior, se define:

- ✓ Algoritmo α : no posee una adecuada respuesta ante la falta de información y el ruido, solo funciona con procesos estructurados, no tolera todas las construcciones estructurales y únicamente comprueba si la relación existe o no, o sea, no tiene en cuenta la frecuencia de

las tareas. Además presenta incapacidad ante los lazos, aunque en una nueva versión del ProM responde adecuadamente.

- ✓ Algoritmo Heurístico: uno de los más completos, pues tolera la falta de información de los registros de eventos. Como desventaja presenta que permite más “comportamientos” que los que se encuentran en el registro de eventos, o sea genera modelos de proceso generales que poseen rutas de ejecución que no aparecen dentro del registro de eventos. Además no tiene un comportamiento positivo si encuentra en las trazas tareas duplicadas, pero presenta buenos resultados en cuanto a registros de eventos reales. Se obtiene como resultado modelos precisos y comprensibles.
- ✓ Algoritmo Genético: necesita muchas iteraciones para converger a buenas soluciones, lo cual aumenta su tiempo computacional. Además responde de forma positiva al ruido, tareas ocultas y lazos.
- ✓ Algoritmo *Fuzzy*: permite una configuración de alto detalle; encontrar el ajuste correcto de los parámetros puede ser engorroso y necesita de usuarios avanzados (Günther & van der Aalst, 2007). Tiene gran utilidad en entornos de procesos desestructurados, pero no soporta todas las construcciones estructurales (de Weerdt et al., 2012). El modelo difuso no puede ser traducido a una red de Petri, lo cual limita una evaluación comparativa con otras técnicas de descubrimiento y la aplicación de análisis de ajuste y precisión (de Weerdt et al., 2012). Como resultado se obtienen modelos de procesos potentes, fáciles de entender y precisos. Además consume muy poco tiempo analizando las trazas.
- ✓ Algoritmo *ILP*: no es robusto ante el ruido, debido a la gran cantidad de soluciones que puede tener un sistema de inecuaciones lineal. Posee un alto tiempo computacional (van der Aalst, 2011).

1.6 Modelos para aplicar técnicas de minería de proceso

Un modelo es un patrón o pauta que se sigue como representación para la realización de algo, se conforma de pasos, fases, tareas y proporciona una guía para llevar a cabo un objetivo.

A continuación se describen los modelos más representativos, diseñados para aplicar técnicas de minería de proceso sobre los registros de eventos de los sistemas de información actuales, así como la propuesta de un caso de estudio. De forma general se presentan las etapas, fases o pasos dentro de los tipos de minería de proceso descubrimiento, conformidad y extensión o mejora.

1.6.1 Modelo de van Giessel

Este modelo (van Giessel, 2004) propone aplicar minería de proceso en un sistema SAP². Está compuesto por dos fases fundamentales y cinco pasos. (Ver **Anexo 2**)

Fase 1: Se determina cuáles tablas son relevantes a través de los objetos de negocio. Se establece un enlace entre los objetos de negocio y el proceso de negocio, mediante componentes de la aplicación. Esta fase está parcialmente automatizada por la herramienta *Table Finder*.

Fase 2: Se recupera el flujo del documento. Primeramente se localizan los datos de las tablas y se extraen. La recuperación se realiza mediante la agrupación de todos los datos que tienen el mismo número de documento o que están relacionadas unas con otras. Cuando todos los datos relacionados se agrupan, se recupera el flujo de documentos que al final se exporta en un archivo XML.

Una vez terminadas estas dos fases se procede a aplicar técnicas de minería a través de los siguientes pasos:

Paso 1: Determinar el modelo del proceso: Se descubre el modelo del proceso que representa las actividades ejecutadas por los usuarios. Este modelo del proceso puede ser comparado con el proceso de negocios definido, utilizando el análisis-delta. Todas las actividades ejecutadas se monitorean y la información se utiliza para determinar el modelo del proceso. Existen herramientas como *EMiT* y *Little Thumb* para generar estos modelos, la primera centrada en el tiempo y la segunda en el ruido.

Paso 2: Análisis de la red social: Se centra en las relaciones entre los individuos o grupos de personas; mediante la herramienta MISON se puede descubrir las relaciones y realizar el análisis. Las relaciones se construyen por métricas predefinidas y los resultados se pueden mostrar en la representación de una matriz o en un gráfico.

² Sistemas, Aplicaciones y Productos en Procesamiento de datos.

Paso 3: Determinar los parámetros básicos de rendimiento: La herramienta *ARIS Process Performance Manager*, calcula los indicadores pre-definidos de rendimiento de procesos, pero también se pueden establecer indicadores de rendimiento de procesos a sí mismo. La minería de proceso calcula los indicadores básicos de rendimiento del proceso a partir del registro de eventos.

Paso 4: Determinar el modelo de la organización: Las empresas están organizadas de diversas formas tanto funcional, lineal como multi-funcional, donde cada integrante es responsable de actividades específicas. Si se desea obtener el modelo de organización sin previo conocimiento acerca de la estructura real, se puede utilizar minería de proceso, ya que en el registro de eventos se guarda el originador de cada instancia de un proceso, a través del cual se puede determinar qué actividades son ejecutadas por quiénes y por tanto, obtener el modelo de la organización.

Paso 5: Análisis de las características de rendimiento: Este paso se basa en el "si-entonces", ya que cada instancia de proceso tiene características específicas y a través de las técnicas de minería se pueden encontrar relaciones significativas entre diferentes características de rendimiento.

1.6.2 Modelo de Rozinat

Se desarrolla con el objetivo de extraer información clave (de los datos, el rendimiento y la organización), que puede ser usada para la creación de un modelo de simulación, que permite visualizar las dependencias entre las técnicas de minería de proceso a utilizar (Rozinat et al., 2009). (Ver **Anexo 3**). Está compuesto por seis pasos que se describen a continuación:

Paso 1: Creación del registro de eventos: La mayoría de los sistemas de información poseen algún tipo de registro de eventos, que contiene el flujo de actividades ejecutadas durante el proceso de negocio. Cada evento se refiere a una instancia de proceso y a una actividad, cuya actividad está determinada por propiedades. El registro de eventos es transformado al formato MXML.

Paso 2: Descubrimiento del flujo de actividades: A partir de la información recopilada del registro de eventos, es empleado un algoritmo de descubrimiento que refleje las relaciones entre las actividades, por ejemplo el algoritmo *alpha* (α) que automáticamente construye el modelo en una red de Petri.

Paso 3: Análisis de puntos de decisión: Después del descubrimiento se pretende profundizar en la perspectiva de los datos del proceso para descubrir dependencias de datos que influyen en la

ruta de un caso. Para analizar las opciones en un proceso, primero se debe identificar las partes del modelo en las que el proceso se divide en varias ramas alternativas, también llamadas puntos de decisión. Luego, se desea encontrar las reglas para el seguimiento de una ruta o la otra, sobre la base de los atributos asociados a los eventos en el registro.

A partir de identificar un punto de decisión en un proceso de negocio y la clasificación de las decisiones de todas las instancias de proceso en el registro, el siguiente paso es determinar si las decisiones pudieran ser influidas por los datos de los casos, es decir, si los casos con ciertas características, por lo general siguen una ruta específica.

Paso 4: Análisis de rendimiento: Se realiza con el objetivo de profundizar en la perspectiva del rendimiento del proceso, para mejorar el modelo del proceso con información sobre los tiempos de ejecución y tiempos de espera para las actividades. También se desea mejorar el modelo del proceso con las probabilidades para tomar caminos alternativos, y con información sobre el esquema de generación de caso.

En la red de Petri descubierta se puede reproducir cada instancia del proceso para que la información sobre la ejecución y los tiempos de espera sea recogida para las actividades en el proceso. Además, para cada punto de decisión se pueden derivar las probabilidades de caminos alternativos basados en la frecuencia con que cada ruta fue seguida durante la reproducción del registro. Por último, la tasa de llegada de los casos puede ser fácilmente derivada de las horas de inicio de la actividad, por primera vez en cada instancia del proceso.

Paso 5: Descubrimiento de roles: La minería organizacional pretende descubrir el modelo de la organización (las relaciones entre los recursos y sus roles o unidades funcionales) y las reglas de asignación (las relaciones entre los roles o unidades funcionales y actividades). Con sólo utilizar un registro de eventos, es difícil descubrir las diferencias entre estos conceptos. Por lo tanto, es necesario formar grupos de recursos en los que las personas realicen actividades similares, por ejemplo a partir del perfil.

Paso 6: Modelo integrado: En los pasos anteriores se evidencia cómo pueden ser extraídas de un registro de eventos diferentes características de un proceso. Estas pueden utilizarse para construir un modelo de simulación. Para obtener una mejor visión sobre el proceso en su conjunto, es útil integrar las perspectivas descubiertas en un modelo holístico. Esto es bastante fácil, siempre y

cuando las características del proceso descubiertas sean ortogonales entre sí (es decir, no hay información contradictoria). Si hay características conflictivas, entonces esto se convierte principalmente en un reto técnico.

1.6.3 Modelo de Bozkaya

Este modelo se desarrolla con el objetivo de ofrecer una visión general de los procesos dentro de los sistemas de información actuales en un corto período de tiempo (Bozkaya, 2009). (Ver **Anexo 4**). Está compuesto por seis fases que se describen a continuación:

Fase 1: Preparación del registro de eventos: La mayoría de los sistemas de información poseen un formato personal de registro, el cual es necesario extraer y transformar. Primeramente se selecciona la mejor idea de un caso, luego se identifican las actividades y sus eventos, todo esto con el fin de obtener un registro de eventos adecuado para proceder a la siguiente fase.

Fase 2: Inspección del registro de eventos: El objetivo fundamental es entender con mayor claridad el registro de eventos y obtener una primera vista del proceso. Primeramente se hace un resumen estadístico del registro que da una idea de la cantidad de eventos almacenados y de los posibles algoritmos de minería a utilizar para obtener buenos resultados en las fases siguientes. Teniendo en cuenta las estadísticas, se filtra el registro de eventos con el objetivo de eliminar los casos incompletos; constituyendo este registro de eventos filtrado la entrada para las siguientes fases. Para obtener una primera vista, se usa el *Fuzzy Miner*, (*plugin* de la herramienta ProM), con un evento de inicio y fin artificial, agregado a cada caso en el registro.

Fase 3: Análisis del flujo de actividades: Esta fase tiene como objetivo dar respuesta a la pregunta: "¿Cómo se ve el proceso actual?". Si se tiene una descripción del proceso, se ejecuta una verificación de conformidad para comprobar si el proceso se ajusta a la especificación, o sea, que cada caso en el registro de eventos puede ser reproducido en el proceso definido. Si no existe una descripción del proceso, el flujo de actividades necesita ser descubierto. Actualmente hay disponibles numerosos algoritmos para descubrir los procesos que derivan en modelos. Una buena comprobación para el modelo del proceso consiste en ejecutar una verificación de la conformidad sobre el mismo.

Para encontrar las rutas más frecuentes en el registro de eventos, se emplea el *plugin* Análisis de Secuencias de Ejecución (*Performance Sequence Analysis*) de ProM, el cual muestra los diferentes

patrones, conjuntamente con alguna métrica de rendimiento de cada patrón, como el tiempo promedio de producción.

Fase 4: Análisis de rendimiento: A partir del descubrimiento del flujo de actividades del proceso, se pueden utilizar los modelos para analizar el rendimiento del proceso. Una de las preguntas que da respuesta a esta fase es: "¿Hay cuellos de botella en el proceso?".

Se utiliza el análisis de gráficos de puntos (*Dotted Chart*) para comparar los casos y sus tiempos de producción. Luego, se calculan los tiempos de ejecución de las actividades individuales y el proceso mismo, mediante la reproducción del registro en el modelo del proceso y los cuellos de botella. Después de la reproducción, las secuencias diferentes deben ser inspeccionadas por separado, dando una valiosa información sobre los tiempos de producción en los casos con excepciones u otros comportamientos poco frecuentes.

Fase 5: Análisis de los roles: En caso de que el registro de eventos contenga información sobre quién ejecutó determinado evento, los roles en dicho proceso se pueden analizar. Durante el desarrollo de esta fase se responden preguntas como: "¿Quién ejecuta qué actividades?" y "¿Quiénes están trabajando juntos?".

Primeramente se crea una matriz de rol-actividad, luego cada función de un perfil se hace a partir de esta matriz. Para formar un grupo es necesario que los roles presenten perfiles similares. Se puede comprobar el modelo descubierto con el organizativo usado, lo que permite obtener información sobre el desempeño de los trabajadores, cómo es la comunicación entre ellos, etc. Para calcular el perfil de cada uno de los roles presentes en el registro de eventos se puede utilizar el *plugin* Minería Organizacional.

Fase 6: Mostrar los resultados al cliente: Como resultado de las fases anteriores se obtiene el comportamiento real del sistema, que generalmente se desvía del proceso diseñado, puede ser tanto por conductas inadecuadas como por aportes al flujo de actividades para facilitar el trabajo. Este resultado debe ser tratado directamente con el jefe de la organización, quien determina cuáles fueron los comportamientos deseados y no deseados en el sistema. Luego utiliza todo este conocimiento para rediseñar su sistema de información para hacerlo más eficiente y competitivo.

1.6.4 Modelo de van der Aalst

Este modelo hace una descripción del ciclo de vida (L*) de un proyecto de minería de proceso (van der Aalst, 2011). (Ver **Anexo 5**) y está compuesto por cinco etapas que se describen a continuación:

Etapas 0 - Planificar y justificar. Los proyectos de minería de proceso comienzan por planificar y justificar las actividades a desarrollar. Para obtener adecuados resultados es necesario valorar los beneficios, antes de gastar esfuerzos en las actividades de minería. Existen tres tipos básicos de proyectos de minería de proceso:

- ✓ Basado en datos: impulsado por la disponibilidad de los datos de un evento; no posee preguntas concretas u objetivas, sino que espera surjan ideas valiosas al analizar los datos de eventos.
- ✓ Impulsado por preguntas: con el objetivo de responder a preguntas específicas, por ejemplo: ¿Por qué los casos atendidos por el equipo determinado tardan más que los casos atendidos por otro equipo?
- ✓ Impulsado por objetivos: con el objetivo de mejorar los procesos en relación a los indicadores claves de rendimiento (reducción de costos o tiempos de respuesta).

Etapas 1. Extracción. Luego de iniciar el proceso del proyecto se deben extraer del sistema los datos de eventos, modelos, objetivos y las preguntas. Para ello, es necesario entender los datos almacenados para analizar cuáles pueden ser utilizados en el análisis y qué preguntas son de mayor importancia.

En esta etapa se describe un método para la extracción de datos de una variedad de sistemas, lo cual representa un proceso iterativo; se describe la técnica de análisis de gráficos de puntos como vía factible para explorar el registro de eventos y una guía para el proceso de filtrado del registro.

Según van der Aalst es posible que se encuentren modelos realizados manualmente, que aunque tienden a tener poca calidad son el punto de partida de los modelos actuales, para aprovechar todo el conocimiento posible. Los modelos existentes pueden ayudar a establecer el ámbito del proceso y juzgar la integridad de los registros de eventos.

Etapla 2. Crear el modelo de flujo de actividades y conectar con el registro de eventos. Tiene como objetivo fundamental obtener el modelo de flujo de actividades del proceso que se analiza. Dicho modelo puede ser descubierto usando las técnicas de descubrimiento de procesos; sin embargo, si es un buen modelo, se puede verificar a través del análisis de conformidad, comparar los modelos o combinar el modelo hecho a mano y el modelo descubierto.

Como resultado se obtiene un modelo de flujo de actividades conectado con el registro de eventos, es decir, los eventos en el registro se refieren a actividades en el modelo. Esta conexión es fundamental para los pasos posteriores, siendo este modelo la entrada del proceso en la **Etapla 3**.

Etapla 3. Crear un modelo de proceso integrado. La relación entre el registro de eventos y el modelo resultante de la **Etapla 2** es utilizada para extender el modelo en esta etapa, el cual ha sufrido cambios factibles a través de la incorporación de nuevas perspectivas. A continuación se describen algunos pasos en el proceso de integración de las diferentes perspectivas:

Paso 3a: agregar la perspectiva organizacional. Es posible analizar la red social y posteriormente identificar las entidades organizativas que conectan las actividades de los grupos de recursos.

Paso 3b: agregar la perspectiva del tiempo. Las marcas de tiempo y las frecuencias se pueden utilizar para conocer las distribuciones de probabilidad que describen adecuadamente los tiempos de espera y servicio y las probabilidades de enrutamiento.

Paso 3c: agregar la perspectiva del caso. Los atributos incorporados en el registro son utilizados para la minería de decisión. Esto demuestra que los datos son relevantes y deben ser incluidos en el modelo.

Paso 3d: agregar otras perspectivas. Dependiendo de la información en el registro se pueden añadir otras perspectivas en el modelo como por ejemplo, la información sobre los riesgos y los costos.

Como resultado se obtiene un modelo integrado con el fin de analizarlo para comprender mejor el proceso e identificar cuellos de botella. Además se puede utilizar para responder determinadas preguntas y tomar decisiones adecuadas. El modelo integrado constituye la entrada a la **Etapla 4**.

Etapla 4. Apoyo a las operaciones. Esta etapa cumple con los objetivos de la minería de proceso, que son las actividades de apoyo operativo: detectar, predecir y recomendar. Se debe tener en

cuenta que el apoyo operativo es la forma más ambiciosa de la minería del proceso. Este sólo es posible para los procesos Lasaña. Por otra parte, es necesario que haya una infraestructura avanzada de las TIC que proporcione una alta calidad de los registros de eventos y permita la incorporación de un sistema de soporte operativo. Para poder alcanzar las **Etapas 3 y 4** es necesario tener un proceso suficientemente estable y estructurado.

1.6.5 Caso de estudio de Jans

Este caso de estudio no se encuentra estandarizado como modelo, pero constituye un ejemplo para guiar esta investigación (Jans, 2011). (Ver **Anexo 6**). A continuación se realiza un resumen de los pasos fundamentales como resultado de la aplicación del caso de estudio:

Paso 1: Caracterización del proceso: Para poder obtener buenos resultados en el proceso de minería es necesario realizar un exhaustivo análisis del proceso, en el que se recogen todos los datos relevantes y se crea un mapa de dicho proceso. Este mapa consta de cuatro componentes: objetivos del proceso, actividades, flujos de información y contabilidad del impacto, siendo el segundo componente el más importante por constituir la entrada principal al **Paso 2**, o sea, la creación del registro de eventos.

Además se realizan consultas a los ejecutivos de negocios y especialistas del sistema de información, entrevistas y observaciones a los empleados de distintos departamentos durante horario laboral; así como la revisión de los manuales de usuarios internos del sistema.

Paso 2: Creación del registro de eventos: Luego de realizar toda la caracterización del proceso se procede a construir el registro de eventos adecuado, a partir de los datos capturados por el sistema de información y almacenados en el registro. Estos datos se encuentran en grandes cantidades y dispersos por numerosas tablas del sistema, las cuales son específicas para cada sistema y cada proceso de negocio. Para minar los datos es necesario configurarlos dentro del registro de eventos con unos requisitos de formato que facilite la minería de proceso.

Este formato constituye la estructura de los datos relevantes alrededor de las actividades que conforman el proceso. Para crear el registro de eventos es necesario conocer cuáles son las actividades que constituyen el proceso y cuál es el caso o instancia de proceso que es el foco de estas actividades, a través de lo que se define el contenido y la estructura de este. Luego es necesario convertir el registro de eventos en el formato MXML.

Paso 2.1: Refinamiento del registro de eventos: Una actividad, un caso, un usuario y el tiempo son las características para minar un proceso. El momento en que las actividades y el originador son ejecutados, son creadas *meta data* por el sistema de información y no se basan en información de entrada creada por el usuario. Esta información de entrada que se contrapone a la *meta data* es llamada atributos.

Paso 3: Análisis con minería de proceso: Después de crear el registro de eventos, se analiza el proceso a través de las siguientes tareas:

Tarea 1. Descubrimiento de proceso: se analiza el orden de las actividades para descubrir cómo se efectúa el proceso en la vida real.

Tarea 2. Análisis de roles: se analiza la asociación de las actividades con las personas a través de la prueba de controles de segregación de tareas.

Tarea 3. Verificación por análisis de atributos: se analiza además la salida de los pasos previos. Se utilizan los atributos de los casos con el fin de verificar la eficacia de los controles internos de la empresa a lo largo del proceso en análisis.

Tarea 4. Análisis de redes sociales: se realiza una búsqueda de las redes sociales entre los empleados que participan en el proceso, con el objetivo de comprender mejor la interacción entre las funciones y actividades para las transacciones anómalas.

En resumen, se puede determinar el grado de concordancia entre lo detectado por los auditores internos en su análisis estándar y la información descubierta al aplicar el proceso de minado.

1.6.6 Valoración de los modelos anteriores

Después del estudio de los modelos que se describen con anterioridad, se puede concluir que son modelos muy generales. En la **Tabla 1.3** se resumen las principales características definidas en cada modelo, donde se destaca el modelo de van der Aalst como el más completo, seguido por el modelo de Bozkaya.

Tabla 1.3. Comparación de modelos sobre minería de proceso

Característica	Jans	Giessel	Rozinat	Bozkaya	van der Aalst
Enfoque de análisis					x
Diagnóstico preliminar			x	x	
Análisis a partir de tipo de proceso				x	x
Caracterización del proceso de negocio	x		x	x	x
Análisis con el cliente				x	x
Método de extracción del registro		x		x	x
Técnicas a utilizar	x	x	x	x	x
Análisis de la perspectiva organizacional	x	x	x	x	x

Fuente: Elaboración propia

A partir de los modelos consultados, se considera por los autores de la investigación, utilizar de las propuestas realizadas por van der Aalst, Bozkaya y Jans, las características que se ajustan a las necesidades de la presente investigación, lo cual se fundamentará en el siguiente capítulo.

1.7 Investigaciones sobre minería de proceso en la salud

Con el descubrimiento y desarrollo de la minería de proceso se ha dado un gran paso de avance para el mejoramiento de los procesos en las organizaciones, aun cuando esta se manifiesta como una tecnología incipiente. A continuación se describen investigaciones de utilización de minería de proceso en el campo de la salud, consideradas relevantes para la realización de este trabajo.

1.7.1 Aplicación de la minería de proceso en el área de salud

En el artículo *Application of Process Mining in Healthcare* (Mans et al., 2009), se aplican técnicas de minería de proceso para la obtención de conocimientos sobre los procesos que se realizan en los hospitales, específicamente su uso en casos reales de procesos de ginecología y oncología en un hospital holandés, demostrándose la necesidad e importancia de su aplicación.

A través del empleo de una variedad de técnicas de minería se analiza el proceso de atención de la salud desde tres enfoques: la perspectiva de control del flujo de procesos, la perspectiva

organizacional y la perspectiva de desarrollo. Utilizando para ello la herramienta ProM que demuestra la viabilidad de las técnicas de minería para optimizar los procesos existentes.

Se parte del enfoque orientado al trabajo con hospitales para poder tener un idea global de qué buscar en el momento de enfrentarse a una actividad de minería de proceso en esta esfera, la variedad de técnicas utilizadas da muestras de los diferentes resultados estadísticos que se pueden obtener a través de las mismas y de este modo saber seleccionar la técnica adecuada para obtener un resultado determinado. Se refuerza el uso de la herramienta ProM como la mejor selección para este tipo de análisis de procesos y el uso del algoritmo alpha (α) por sus características como la base de la minería de proceso.

1.7.2 Descubrimiento automatizado de modelos de flujo de trabajo a partir de datos hospitalarios

Automated Discovery of Workflow Models from Hospital Data (Maruster et al., 2001), es una investigación que presenta las Redes de Flujo (Workflow Nets, WF por sus siglas en inglés) como una subclase de las redes de Petri. Además se propone un método que parte del descubrimiento automático de las WF en los logs de procesos, siendo esta una ventaja relevante. Se realiza una prueba del método en logs de procesos hospitalarios simulados, los cuales contienen información médica acumulada durante determinado período de tiempo.

Los resultados arrojan que el método es capaz de descubrir modelos en procesos acíclicos, incluyendo paralelismo, condiciones y construcciones secuenciales. El estudio no es concluyente en WF cíclicas y de decisiones del tipo libre y no libre.

El uso de las redes de Petri ofrece una visión global de los procesos que se analizan, pero si se desea detallar en estos, es recomendable utilizar una WF. El método propuesto es provechoso debido a su demostrada aplicación en el sector de la salud, automaticidad y sencillez, pero su ineficiencia en procesos cíclicos y de cierta complejidad supone un obstáculo para su uso en esta investigación.

1.7.3 BPM EHR: de minería de proceso a mejora de procesos a usabilidad de procesos

Se publica el artículo sobre gestión de procesos de negocios en Hoja Clínica Electrónica, *EHR BPM: From Process Mining to Process Improvement to Process Usability* (Webster, 2011), donde

se enmarca la importancia de la minería de proceso en las Hojas Clínicas Electrónicas (*Electronic Health Record*, EHR por sus siglas en inglés), como un nuevo método para mejorar los procesos de cuidado a pacientes. La creciente demanda de mapas de procesos mejor evidenciados en los sistemas de salud genera necesidad de mejores *logs* de eventos en las EHR.

La utilización de minería de proceso en las EHR aumentaría su usabilidad y adopción. Las EHR son importantes para los principales asuntos médicos, como son la coordinación entre el personal médico en cuanto a cuidados, uso eficiente de los recursos médicos y el registro de los procesos en los que ha intervenido el paciente.

En las EHR se recogen los datos de los pacientes y su diagnóstico, pero esto, a su vez es un proceso que se instancia por cada EHR obtenida. El análisis de uno de estos ficheros simplifica el problema que muchas veces genera realizar el proceso de Extracción, Transformación y Carga (*Extraction, Transformation and Load*, ETL por sus siglas en inglés), que por lo general, debe sufrir la fuente de datos para quedar conformada como registro de eventos, donde los procesos quedan descritos de modo explícito. Este artículo aporta una alternativa para obtener los datos cuando la fuente resulte inaccesible o su modificación esté dada por una alta complejidad.

1.7.4 Desafíos de la minería de proceso en los sistemas de información hospitalarios

En los últimos años los sistemas de información hospitalarios se están convirtiendo en sistemas adecuados a utilizar minería proceso para el descubrimiento de conocimiento y optimización de procesos.

Por esta razón, se presenta el artículo *Process mining challenges in hospital information systems* (Homayoungfar, 2012) en la conferencia *The Proceedings of the Federated Conference on Computer Science and Information Systems*, donde su aplicación es un enfoque moderno y recomendable en el cuidado de la salud.

Sin embargo, las técnicas de minería de proceso sólo pueden proporcionar un resultado de alta calidad si se conoce la estructura de los datos y si la estructura de los registros de eventos son mantenidos adecuadamente. En este trabajo se describe la minería de proceso y los sistemas de información hospitalaria, y se muestra dónde están los retos si las dos áreas se combinan.

Luego del estudio de las investigaciones que muestran la aplicación de las técnicas de minería de proceso en el campo de la salud, se puede concluir que los resultados obtenidos por estos autores no se ajustan a la investigación actual, ya que analizan problemáticas específicas y no se pueden utilizar como base para lograr los objetivos que persiguen los autores de este trabajo.

1.8 Herramientas seleccionadas

Durante la realización de este trabajo se utilizan las herramientas descritas a continuación:

Pentaho Data Integration en su versión 4.2 para el proceso de ETL, el servidor de bases de datos PostgreSQL en su versión 8.4, MonoDevelop en su versión 2.0 para el desarrollo de la herramienta de apoyo JyX que realiza la conversión al formato estándar de registro de eventos y el ProM en la versión 6.0 para hacer el proceso automatizado de extracción de información relevante. Además para guiar el trabajo se utiliza el modelo propuesto por van der Aalst y Bozkaya, así como el caso de estudio de Jans, con adaptaciones para aplicar en esta investigación.

1.8.1 Pentaho Data Integration

Las organizaciones se enfrentan a una avalancha de grandes volúmenes y diversos tipos de datos que siguen creciendo y crean un desafío para el análisis de negocios. *Pentaho Data Integration* permite a las organizaciones extraer datos de fuentes complejas y heterogéneas y diferentes tipos de datos para proporcionar alta calidad. Con una rica interfaz gráfica de usuario y un motor de procesamiento paralelo, ofrece alto rendimiento para realizar los procesos de ETL que soporta todas las necesidades de integración de datos (Pentaho Data Integration, n.d.).

Incluye:

- ✓ Inteligencia de negocios y almacenamiento de datos.
- ✓ Migración de datos.
- ✓ Consolidación de aplicaciones.
- ✓ Sincronización de datos.

Proporciona:

- ✓ Rico entorno visual de diseño gráfico.
- ✓ Amplia conectividad a cualquier tipo de fuente de datos.

- ✓ Escalabilidad y rendimiento empresarial.
- ✓ Información integrada, visualización, análisis de datos y análisis predictivo.

Pentaho Data Integration es altamente utilizado para procesos de ETL gracias a su compatibilidad con diferentes formatos y su multiplataforma. Es una herramienta de software libre que posee, entre otras, funcionalidades de entrada y salida de diversos tipos de archivos y una gama de herramientas para someterlos a modificaciones durante el proceso. Todo esto en un entorno completamente gráfico sin necesidad de conocimientos de programación por parte de los usuarios.

1.8.2 PostgreSQL

Sistema gestor de bases de datos objeto-relacional (*Object-relational Database Management System*, ORDBMS por sus siglas en inglés) basado en Postgres. Fue pionera en varios conceptos que sólo estuvieron disponibles en algunos sistemas de bases de datos comerciales mucho después. PostgreSQL es un descendiente de código abierto del original de Berkeley.

Soporta una gran parte del estándar SQL y ofrece varias características modernas:

- ✓ Consultas complejas.
- ✓ Claves foráneas.
- ✓ Desencadenantes o disparadores.
- ✓ Vistas.
- ✓ Integridad transaccional.
- ✓ Control de concurrencia multiversión.

El tamaño máximo de la base de datos es ilimitado; el de una tabla asciende a 32 TB, el de una fila a 1.6 TB y el de un campo de datos a 1 GB. De igual forma el número de filas en una tabla es ilimitado, pero no el de columnas, que oscila entre 250 y 1 600 columnas por tabla. El número de índices por tabla es también ilimitado (PostgreSQL, n.d.). Además, PostgreSQL puede ser ampliado por el usuario en muchos aspectos, por ejemplo, la adición de nuevos:

- ✓ Tipos de datos.
- ✓ Funciones.
- ✓ Operadores.
- ✓ Funciones de indexado.
- ✓ Métodos de índice.

- ✓ Lenguajes procedurales.

Debido a la licencia libre, PostgreSQL puede ser utilizado, modificado y distribuido por todo el mundo de forma gratuita para cualquier propósito, sea comercial privado o académico.

1.8.3 ProM

Desarrollado por la Universidad Técnica de Eindhoven, es la herramienta más potente en el campo de la minería de proceso. Es de software libre, multiplataforma y está diseñada para que en ella se puedan desarrollar y ejecutar algoritmos de minería. ProM es un marco extensible que es compatible con una amplia variedad de técnicas minería de proceso en forma de *plugins*. Es independiente de la plataforma, ya que es implementado en Java y puede ser descargado sin ningún costo. Está publicado bajo una licencia de código abierto. Presenta una gran variedad de filtros para la limpieza de registros no deseados o sin importancia (ProM, n.d.).

En la actualidad, ya hay más de 90 complementos disponibles que son compatibles con la importación y la conversión de varios lenguajes de modelado de los procesos, tales como: EPKs/EPCs (formato gráfico ARIS, EPML), TPN, PNML, YAWL y más. Posee *plugins* de minería, tales como:

- ✓ Compatibilidad con las técnicas de minería de control de flujo (como el algoritmo Alpha, Genético, Múltiples fases, etc.).
- ✓ Análisis de la perspectiva organizacional.
- ✓ Análisis de la perspectiva de datos.

Además cuenta con *plugins* de análisis que permiten:

- ✓ La verificación de los modelos de procesos.
- ✓ La verificación de las fórmulas de lógica temporal lineal en un registro.
- ✓ La comprobación de la conformidad entre un modelo de proceso dado y un registro.
- ✓ El análisis de rendimiento (análisis estadístico básico y análisis de rendimiento con un modelo de proceso dado).

1.8.4 MonoDevelop

MonoDevelop es un IDE multiplataforma diseñado principalmente para C # y otros lenguajes .NET. Permite a los desarrolladores escribir rápidamente aplicaciones web o de escritorio en sistemas operativos como Linux, Windows y Mac OSX. Las aplicaciones creadas mantienen una base de código único para todas las plataformas (MonoDevelop, n.d.).

Entre sus características destacan la capacidad de multi-plataforma (Linux, Windows y Mac OS X), edición de texto avanzada, banco de trabajo configurable (diseños de ventana personalizables, totalmente definidos por el usuario, atajos de teclado y herramientas externas), soporte para múltiples idiomas (C #, Visual Basic.Net y C / C + +), depurador integrado, diseñador visual y soporte de completamiento de código. Además proporciona gestión de bases de datos con funcionalidades como:

- ✓ Exploración de bases de datos.
- ✓ Ejecución de consultas y vista de resultados en su interior.
- ✓ Creación de nuevas bases de datos, tablas, procedimientos y vistas.
- ✓ Modificación del nombre de las bases de datos, tablas, procedimientos y vistas.

Conclusiones del capítulo

1. Se plantearon conceptos fundamentales para lograr una mejor comprensión de la problemática.
2. Se compararon los modelos de aplicación de minería de proceso, para determinar el más adecuado y además las herramientas para el desarrollo del trabajo.
3. El análisis de la literatura evidenció la carencia de publicaciones sobre minería de proceso aplicados al sector de la salud en el ámbito nacional, mientras que a nivel internacional se consultaron diversos artículos y publicaciones.
4. La utilización de un algoritmo de descubrimiento específico, está determinado por las características especiales de cada organización o proyecto.
5. El estudio de la minería de proceso ha demostrado que esta constituye una herramienta de mejora que puede ser aplicada a la mayoría de los sectores sociales.



CAPÍTULO 2. PROCEDIMIENTO PARA LA DETECCIÓN DE EVENTUALIDADES APLICANDO TÉCNICAS DE MINERÍA DE PROCESO

En este capítulo se propone el procedimiento para la aplicación de técnicas de minería de proceso, a partir del análisis de los modelos consultados. Se describe un conjunto de fases y pasos que guían la utilización de las técnicas y algoritmos de minería de proceso para la detección de eventualidades en los sistemas de información de salud.

2.1 Características del procedimiento

El procedimiento para la detección de eventualidades aplicando técnicas de minería de proceso en los sistemas de información de salud, tiene como objetivo proveer al proceso de desarrollo de software del sistema xavia HIS de una herramienta que contribuya a la mejora de la implementación del modelo de negocio.

Esta guía es aplicable a cualquier proceso de desarrollo de software que gestione los procesos de salud, donde los líderes de proyectos estén comprometidos con su puesta en práctica y dispuestos a ofrecer la información necesaria y a mejorar sus procesos.

Para su aplicación se establecen las siguientes premisas: la información de los procesos debe estar recogida en sistemas de información automatizados, que registren como información mínima asociada al evento un identificador, el originador y la marca de tiempo; y haga posible la vinculación de cada evento a la instancia correspondiente.

El procedimiento propone 5 Fases para su aplicación: “Planificación y justificación”, “Caracterización del proceso de negocio”, “Preparación de registro de eventos”, “Inspección del registro de eventos” y “Análisis del flujo de actividades”. En la **Figura 2.1** se muestra el esquema del procedimiento diseñado.

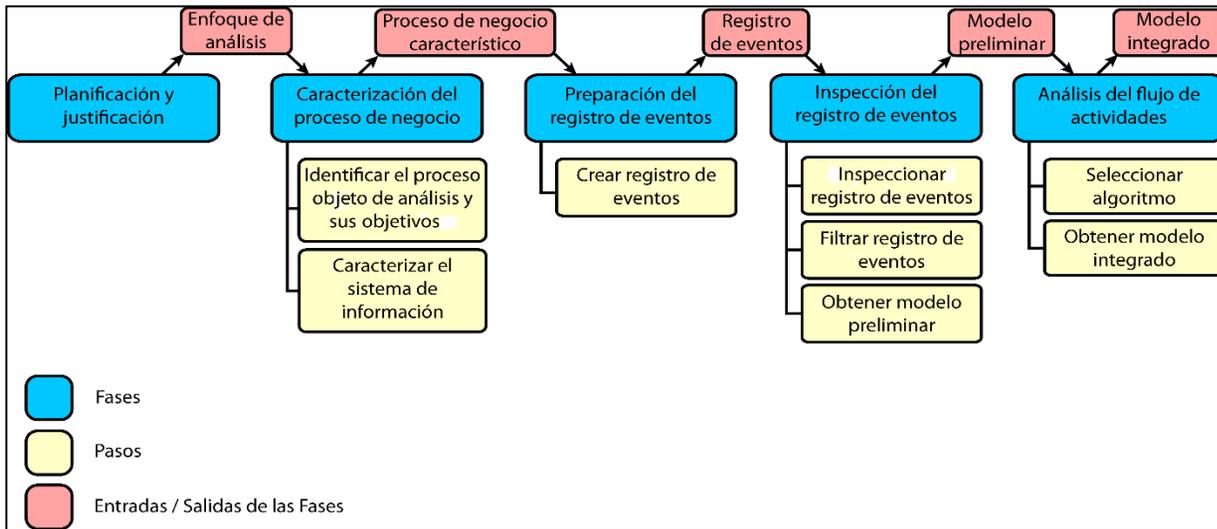


Figura 2.1. Procedimiento para la detección de eventualidades aplicando técnicas de minería de proceso

Fuente: Elaboración propia

Fase 1. Planificación y justificación, se realiza con el objetivo de seleccionar el enfoque de análisis para la aplicación de la minería de proceso, puede ser orientado a datos, a preguntas o a objetivos.

Fase 2. Caracterización del proceso de negocio, tiene como objetivo caracterizar la ejecución del proceso de negocio objeto de análisis, con el fin de obtener un modelo preliminar del proceso. Se divide en dos pasos: **Describir el proceso objeto de análisis y sus objetivos** y **Caracterizar el sistema de información**.

Fase 3. Preparación del registro de eventos, con el objetivo de obtener un registro de eventos aplicable a las técnicas de minería de proceso. Contiene el paso **Crear registro de eventos**.

Fase 4. Inspección del registro de eventos, se realiza con el objetivo de inspeccionar el registro de eventos y obtener una representación preliminar de la secuencia de actividades del proceso, haciendo uso de las técnicas de diagnóstico de la minería de proceso. Se divide en tres pasos: **Inspeccionar registro de eventos**, **Filtrar registro de eventos** y **Obtener modelo preliminar**.

Fase 5. Análisis del flujo de actividades, tiene como objetivo obtener el modelo del proceso a partir del registro de eventos y exponer los resultados. Se divide en dos pasos: **Seleccionar algoritmo y Obtener modelo integrado**.

2.2 Procedimiento para la detección de eventualidades aplicando minería de proceso

2.2.1 Fase 1. Planificación y justificación

Objetivo: Seleccionar el enfoque de análisis para la aplicación de minería de proceso a partir de las necesidades de los líderes de proyecto.

A continuación se describen los tres enfoques de análisis para la aplicación de minería de proceso, propuestos en la etapa 0 del modelo de van der Aalst:

- ✓ Enfoque orientado a datos: Tiene como objetivo realizar un análisis general del proceso, es impulsado por la disponibilidad de los datos. No hay preguntas u objetivos concretos, sino que las partes interesadas esperan que del análisis de los datos surja información relevante. Tiene carácter exploratorio.
- ✓ Enfoque orientado a preguntas: Su objetivo es dar respuesta a preguntas como: el tiempo que demora determinado trabajador o equipo con respecto a otro, irregularidades de una fecha con respecto a otra, etc. Tiene aplicación especial en auditorías.
- ✓ Enfoque orientado a objetivos: Pretende la mejora de un proceso en particular con relación a los Indicadores Claves de Rendimiento (*Key Performance Indicators*, KPI por sus siglas en inglés), como por ejemplo: reducción de costos o tiempos de respuesta.

2.2.2 Fase 2. Caracterización del proceso de negocio

Objetivo: Caracterizar la ejecución del proceso de negocio a analizar, con el fin de obtener el proceso de negocio característico.

Esta fase del proceso toma como base la propuesta realizada por Jans en el paso inicial de su caso de estudio, y los autores de este trabajo proponen una adaptación específica para la problemática abordada.

La información necesaria del proceso de negocio a analizar es la siguiente:

- ✓ Nombre del proceso objeto de análisis.
- ✓ Objetivos del proceso.
- ✓ Características del sistema de información que automatiza el proceso.

A continuación se describen los dos pasos a ejecutar en esta fase:

Paso 1. Describir el proceso objeto de análisis y sus objetivos

Objetivo: Seleccionar el proceso a analizar, delimitar el alcance para la aplicación de la minería de proceso y por último identificar los objetivos que ayudan a determinar las tareas de inicio y fin del proceso. Este paso se propone a partir de la adaptación del caso de estudio de Jans.

Paso 2. Caracterizar el sistema de información

Objetivo: Realizar un estudio de la documentación del sistema de información, que incluye revisión de las tablas de la base de datos, o sea las trazas y la relación entre los módulos, con el fin de obtener información básica del sistema.

2.2.3 Fase 3. Preparación del registro de eventos

Objetivo: Obtener el registro de eventos con las características necesarias para aplicar técnicas de minería de proceso.

Esta fase se corresponde con la fase 1 “Preparación del registro de eventos”, propuesta en el modelo de Bozkaya, ya que el alcance de la presente investigación se basa fundamentalmente en el descubrimiento.

En la **Figura 2.2** se muestra la estructura de un registro de eventos, propuesta en el modelo de van der Aalst, en la cual se observa que:

- ✓ un proceso está formado por casos (instancias),
- ✓ un caso consiste de uno o más eventos (actividades) y
- ✓ los eventos poseen atributos (propiedades).

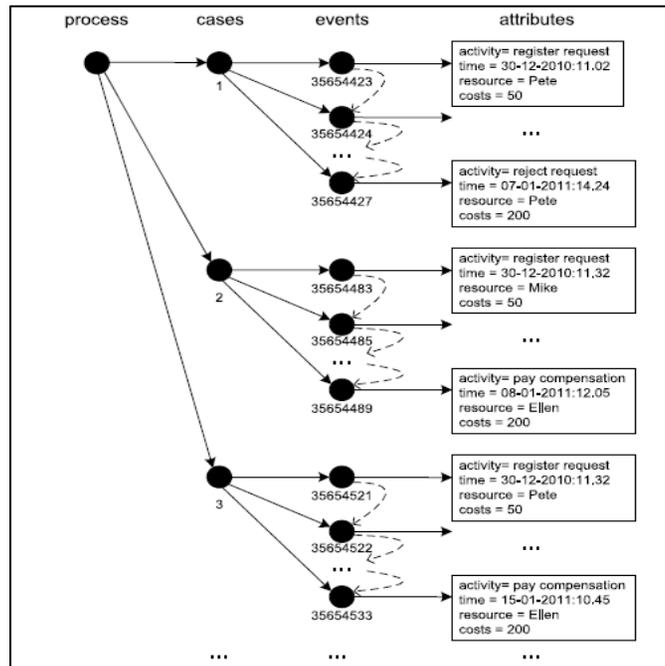


Figura 2.2. Fragmento de un registro de eventos

Fuente: van der Aalst, 2011

A continuación se describen los pasos para realizar esta fase:

Paso 1. Crear registro de eventos

Objetivo: Extraer los datos relevantes de los registro de trazas y realizar la conversión al formato estándar, lo que permite obtener el registro de evento.

Para poder aplicar las técnicas de minería de proceso es esencial extraer los registros de eventos a partir de fuentes de datos (bases de datos, registros de transacciones, pistas de auditoría, etc.). Los formatos soportados por la herramienta ProM son MXML (*Mining eXtensible Markup Language*) y XES (*eXtensible Event Stream*), donde XES fue seleccionado por el Grupo de Trabajo de la IEEE sobre minería de proceso como el formato estándar para registro de eventos.

La **Figura 2.3** muestra el *meta* modelo de un fichero XES en términos de diagrama de clases UML. Contiene un registro compuesto por cantidades de instancias. Cada instancia describe una lista secuencial de tareas correspondiente a un caso particular. El registro, sus instancias y sus tareas

pueden tener cualquier número de atributos, que pueden ser anidados. Existen cinco tipos: *string*, *data*, *int*, *float* y *boolean*. Por ejemplo: 2013-01-17T21:00:00.000+02:00 es un valor de tipo *xs:datetime* representando las nueve en punto de la noche del 17 de enero de 2013 en la zona horaria GMT+2.

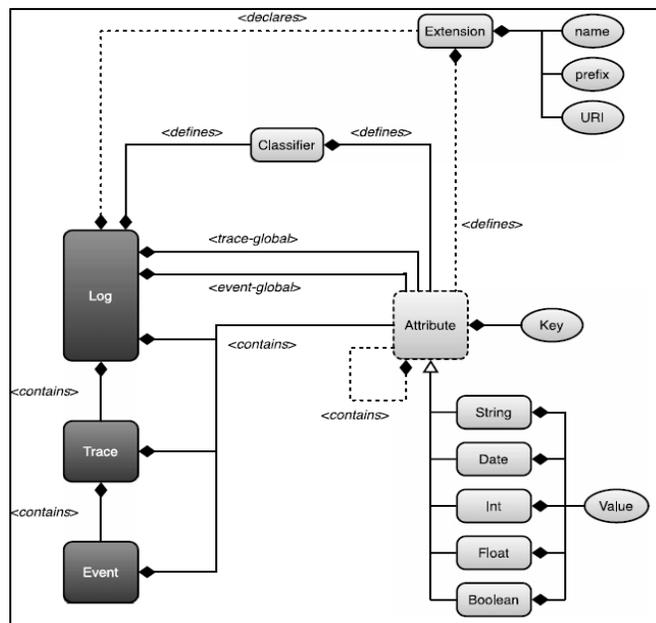


Figura 2.3. Meta modelo de un fichero .XES

Fuente: van der Aalst, 2011

Existen varias herramientas para extraer los registros de MXML o XES de las fuentes de datos, por ejemplo XESame y Nitro. Aunque son eficientes en su campo, poseen alta complejidad de trabajo, lo que resulta inconveniente para usuarios inexpertos. Por esta razón, se propone como parte del procedimiento en esta fase, la utilización de la herramienta JyX, desarrollada por los autores de esta investigación.

Herramienta JyX v1.0

Permite facilitar el trabajo de conversión del registro de eventos en un archivo de extensión XES, el cual es utilizado como entrada en la herramienta ProM.

JyX v1.0 fue desarrollado con MonoDevelop para utilizar la propiedad de multiplataforma, lo que evita la necesidad de utilizar la máquina virtual de Java para su ejecución. Tiene como objetivo fundamental recoger los datos del registro de eventos y estructurarlos jerárquicamente para simplificar su posterior escritura en un archivo.

Se proponen dos versiones de JyX, una para sistemas Linux con un entorno de consola que permite la ejecución rápida, a partir de una línea de comando para automatizar el proceso de minería y la otra versión para sistemas Windows, como se muestra en la **Figura 2.4**, y cuenta con una interfaz visual donde se realiza el proceso de control de las trazas almacenadas en el fichero *.xes.

Para el uso de JyX se presenta el manual de trabajo con una explicación de las funciones y configuraciones de la herramienta, de forma tal que apoye en el proceso de desarrollo de software para la salud, en cuanto a la aplicación de las técnicas de minería para detectar eventualidades (González, 2013).

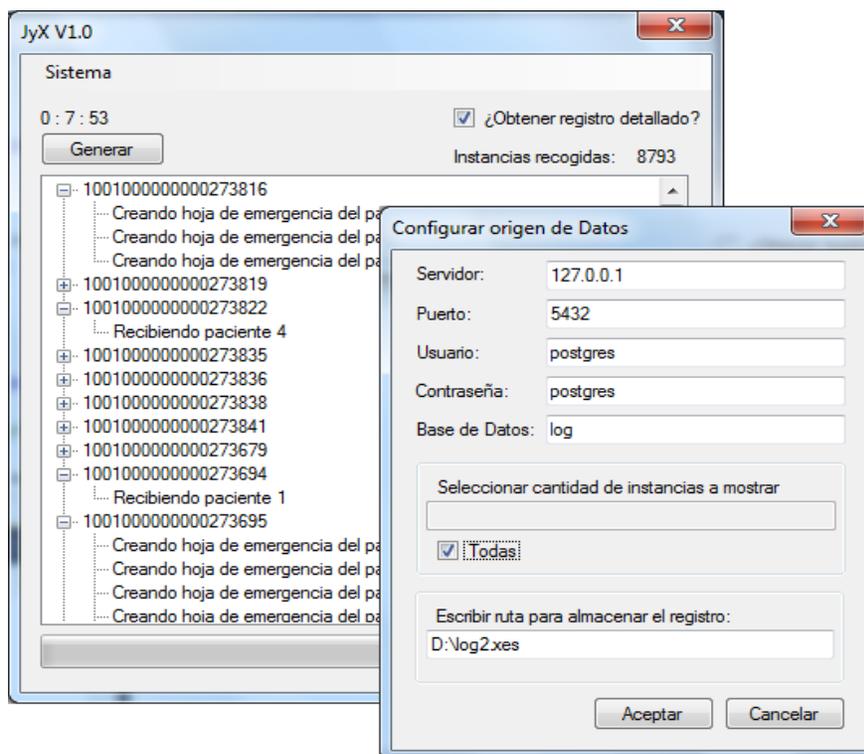


Figura 2.4. Interfaz para Windows de la herramienta JyX v1.0

Fuente: Elaboración propia

Funcionamiento de JyX v1.0

Para una correcta extracción de los datos almacenados en el registro de trazas, es necesario configurar ciertos parámetros de JyX. Estas configuraciones se realizan en el menú "Sistema", opción "Configurar origen de Datos". Los campos "Servidor" y "Puerto" recogen la dirección IP del registro de trazas. La autenticación requerida por el registro de trazas se recibe en los campos "Usuario" y "Contraseña" y posteriormente en el campo "Base de Datos" se deja almacenado el nombre de la base de datos que contiene el registro de trazas.

La herramienta permite decidir cuántas instancias se desean mostrar en el registro de eventos; para ello en el campo "Seleccionar cantidad de trazas a mostrar" se puede escribir el número de instancias deseado o marcar la casilla "Todas" para que la herramienta recoja todos los datos almacenados. Finalmente, en el campo "Escribir ruta para almacenar el registro" se escribe la dirección donde se desea guardar el nuevo registro de eventos.

Una vez configurado el origen de datos se puede decidir qué tipo de registro de eventos obtener, el registro general o el detallado, para lo cual debe marcar la casilla "¿Obtener registro detallado?". La diferencia entre ambos se basa en que el registro general ofrece una visión panorámica del sistema, lo que resulta útil para la obtención del modelo preliminar. El registro detallado convierte cada tarea en única para cada paciente, lo que posibilita observar el comportamiento del modelo de negocio en pacientes específicos.

El botón "Generar" de la interfaz principal de JyX comienza el proceso de extracción de los datos del registro de trazas, un cronómetro ubicado sobre este botón registra el tiempo de ejecución del proceso, valor de tiempo que puede ser variable tomando en cuenta el tamaño de la fuente de datos y la capacidad del hardware donde se emplea la herramienta. Un contador de instancias en la parte derecha de la interfaz muestra la cantidad de instancias que se han procesado hasta el momento "Instancias recogidas".

Después de obtenidas las instancias, estas se muestran en la interfaz principal, para una evaluación preliminar por parte del usuario. Si los datos mostrados son satisfactorios se procede a guardar el registro de eventos al accionar el botón "Guardar" ubicado en la esquina inferior derecha de la

interfaz. De esta forma se toman las instancias obtenidas durante la extracción y se guardan con el formato correspondiente al estándar *.xes.

2.2.4 Fase 4. Inspección del registro de eventos

Objetivo: Realizar una inspección del registro de eventos para obtener una representación preliminar de la secuencia de actividades del proceso. Como resultado se obtiene un modelo preliminar del proceso.

Esta fase se corresponde con la fase 2 “Inspección del registro de eventos” propuesta en el modelo de Bozkaya. A continuación se describen los tres pasos a desarrollar para cumplir el objetivo:

Paso 1. Inspeccionar registro de eventos

Objetivo: Analizar el registro de eventos, para conocer la cantidad de casos, eventos y recursos que intervienen en el proceso, así como una descripción de todos los eventos por instancias y un resumen general del registro de eventos.

A partir de este paso se comienzan a aplicar las técnicas de minería de proceso utilizando la herramienta ProM 6. Un ejemplo que ilustra la primera visualización del registro de eventos (opción “View”) se muestra en la **Figura 2.5**, en la cual se recogen las estadísticas del registro, lo que incluye información sobre:

- ✓ el número total de casos (instancias),
- ✓ los eventos,
- ✓ los recursos,
- ✓ las clases de eventos (actividades),
- ✓ los tipos de eventos,
- ✓ el número mínimo, máximo y el promedio de eventos por caso,
- ✓ las clases de eventos por caso y
- ✓ las fechas de inicio y fin del registro.

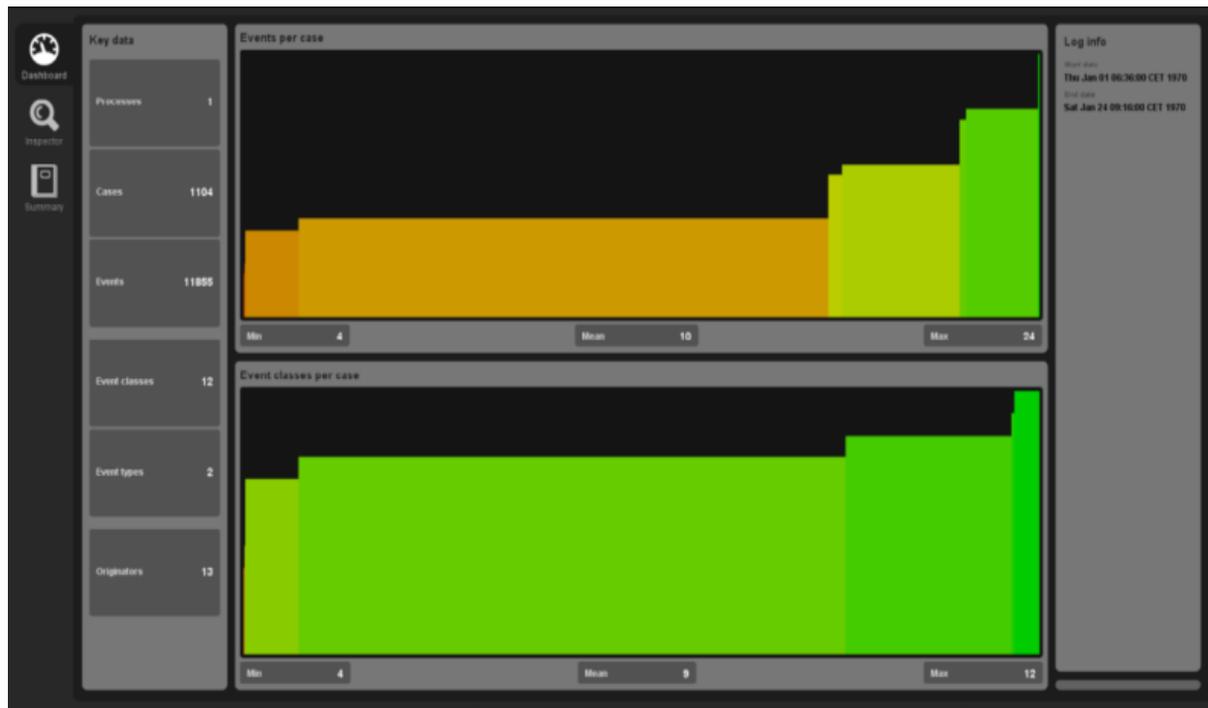


Figura 2.5. Visualización del log de eventos

Fuente: Herramienta ProM 6.0

Por otra parte, la **Figura 2.6**, es un ejemplo donde se puede apreciar una descripción de todos los eventos por instancias del proceso y a su vez, de forma gráfica se muestra el comportamiento de las instancias mediante la visualización de la secuencia de tipos de eventos (opción “Inspector”).

Los tipos de eventos de mayor frecuencia de ocurrencia se representan con color verde y los de menor frecuencia con color rojo.

A partir de esta vista se pueden identificar posibles eventos aislados e instancias que no presenten tareas de inicio y fin definidas.

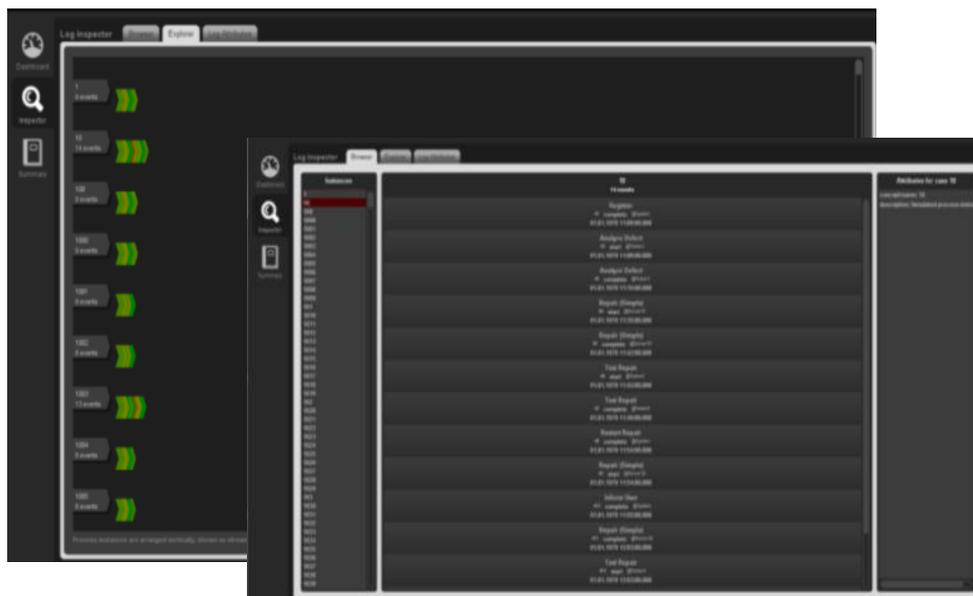


Figura 2.6. Inspección del log de eventos

Fuente: Herramienta ProM 6.0

En la **Figura 2.7** se muestra un resumen del *log* de eventos (opción “*Summary*”), donde se ejemplifican todas las clases de eventos con su ocurrencia absoluta y relativa, así como los eventos de inicio y fin con mayor y menor frecuencia de ocurrencia, y todos los recursos con su frecuencia de trabajo.

Esto permite asociar la clase de evento con el recurso que inicia o cierra la mayoría de los casos y valorar posibles irregularidades.

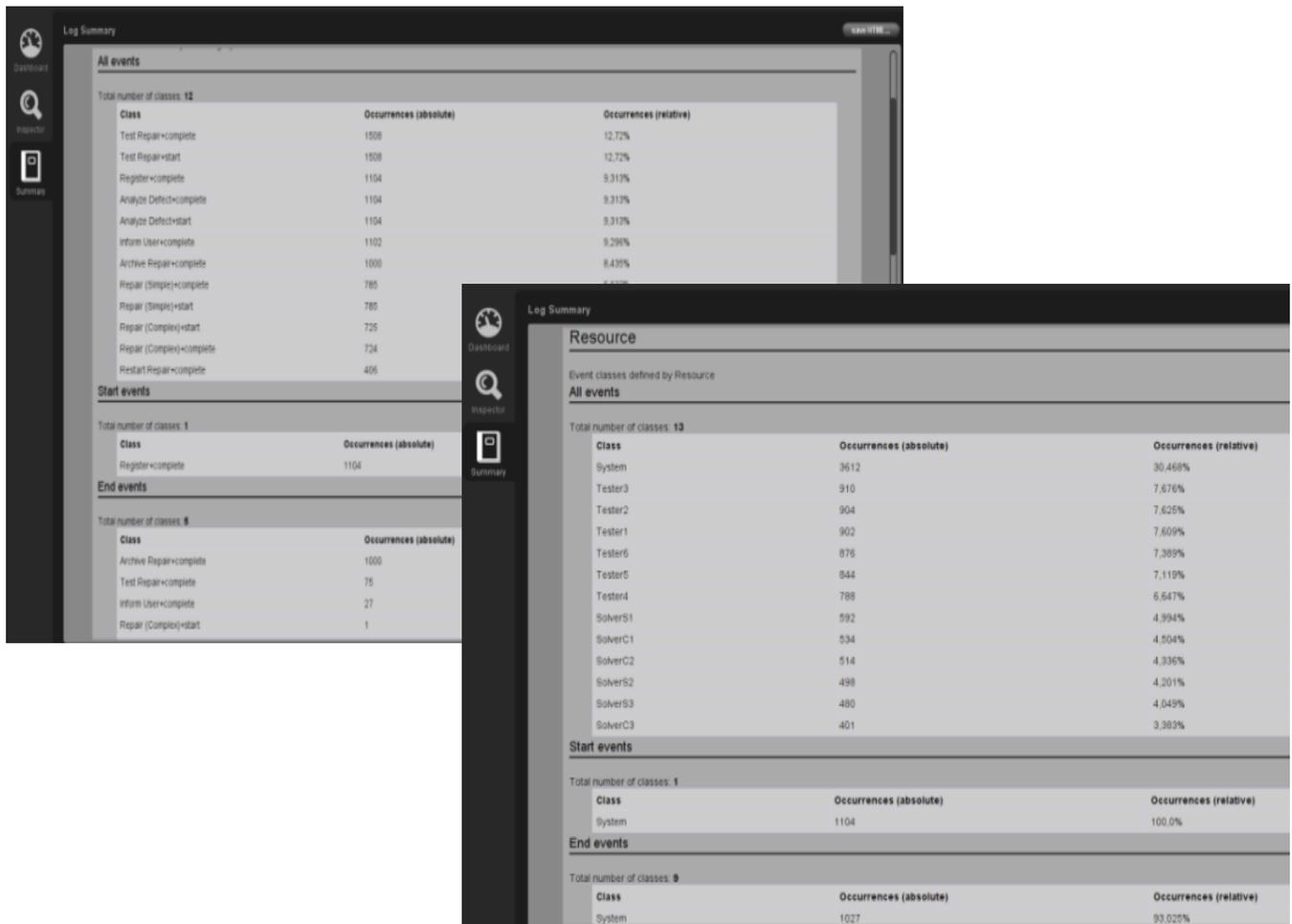


Figura 2.7. Resumen del log de eventos

Fuente: Herramienta ProM 6.0

Como resultado de esta inspección preliminar se obtienen estadísticas sobre el registro de eventos, que brindan información acerca de:

- ✓ Posibles irregularidades detectadas, por ejemplo: tareas aisladas con muy baja frecuencia de ocurrencia que dan inicio y fin a las instancias.
- ✓ Instancias incompletas identificadas, como son las instancias que fueron iniciadas antes del comienzo del registro y las instancias que están aún en ejecución al final del registro de eventos.

Paso 2. Filtrar registro de eventos

Objetivo: Eliminar las instancias no significativas para el análisis, o sea, tareas de muy baja frecuencia que dan inicio y fin a las instancias. Se pueden mencionar, entre otras, las siguientes: excesiva extensión del registro de eventos, posibles irregularidades e instancias incompletas detectadas.

Con vistas a la modelación de la secuencia de actividades del proceso para obtener una mejor representación del flujo, se realiza el proceso de filtrado. Para esto se utiliza el *plugin* "Simple Log Filter" de la herramienta ProM 6. Como primer paso se filtran las instancias con un bajo valor de frecuencia, evitando que puedan dar como resultado un modelo erróneo.

En la categoría "filtro de evento" del *plugin* se marcan los eventos indeseados para ser descartados del registro. Las secciones "eventos de inicio" y "eventos de fin" muestran las tareas que han sido observadas directamente al inicio o al final de una instancia de un proceso, respectivamente. Si se desmarcan ciertos finales de eventos, conlleva a ignorar estas instancias de proceso. Esto es útil para limpiar el registro de casos que no han finalizado aún. La sección "tipos de eventos" permite mantener o retirar algunos tipos de eventos, o descartar toda la instancia que contenga un cierto tipo de evento (que es útil para limpiar el registro de instancias de proceso que contienen errores).

En caso de detectar instancias que no presenten tareas de inicio y fin definidas, es necesario aplicar el *plugin* "Add Artificial Events", con el objetivo de introducir en el registro de eventos tareas artificiales de inicio y fin.

Paso 3. Obtener modelo preliminar

Objetivo: Aplicar técnicas de análisis para obtener una representación del flujo de actividades y realizar un diagnóstico del proceso, que posibilite identificar la secuencia principal del proceso.

Para obtener el modelo preliminar se utiliza el *plugin* "Fuzzy Miner" que puede ser configurado mediante el ajuste de los controles representados en la **Figura 2.8**, para visualizar el modelo con menor o mayor detalle, así como el comportamiento del proceso reflejado en el registro de eventos.

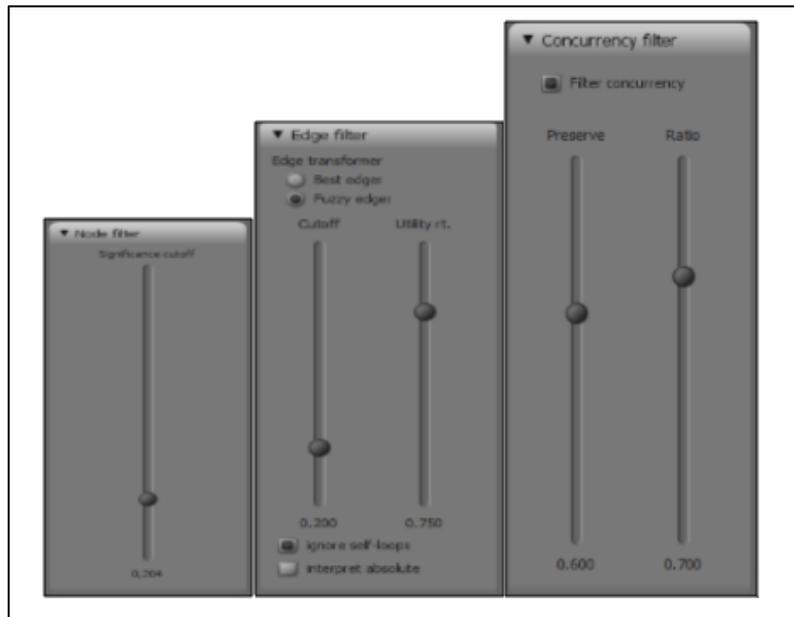


Figura 2.8. Controles para ajustar el modelo preliminar

Fuente: Herramienta ProM

Con los controles se puede ajustar el modelo a través de:

- ✓ Filtro de nodos.

Significance cutoff: Un valor bajo especificado para la tasa de corte representa mayor cantidad de actividades en el modelo como nodos primitivos, mientras que un valor alto provoca mayor cantidad de clúster.

- ✓ Filtro de arcos.

Cutoff: Este parámetro determina cuán abarcador puede ser el algoritmo, es decir, cuanto mayor sea su valor, el algoritmo permitirá la representación de un mayor número de arcos.

Utility rt: Un alto valor para la tasa de utilidad preserva los arcos más significativos, mientras que un valor bajo favorece la correlación entre arcos.

En procesos muy desestructurados, donde es probable que las relaciones de precedencia tengan un significado balanceado, es útil emplear una tasa de utilidad con un valor bajo, de tal manera que

la correlación se tenga en cuenta con mayor énfasis y se resuelvan las situaciones ambiguas. Además un valor de *cutoff* alto puede actuar como un amplificador para distinguir las relaciones más importantes.

- ✓ Filtro de concurrencia.

Preserve: Mientras más bajo sea el valor especificado del umbral a preservar mayor cantidad de relaciones conflictivas permitirá el modelo.

Ratio: Un alto valor especificado para el umbral de la tasa permite mostrar en el modelo mayor cantidad de relaciones conflictivas concurrentes. Mientras que, para un valor bajo de esta tasa las relaciones son consideradas una excepción y se eliminan del modelo, lo que posibilita mostrar menor cantidad de relaciones conflictivas.

2.2.5 Fase 5. Análisis del flujo de actividades

Objetivo: Obtener el modelo integrado del proceso a partir del registro de eventos y validar resultados.

Esta fase se corresponde con la fase 3 “Análisis del flujo de actividades” propuesta en el modelo de Bozkaya y da respuesta a la pregunta: “¿Cómo se ve el proceso actual?”. Si se tiene una descripción del proceso, se ejecuta una verificación de conformidad para comprobar si el proceso se ajusta a la especificación, o sea, que cada caso en el registro de eventos puede ser reproducido en el proceso definido. Si no existe una descripción del proceso, entonces se procede al descubrimiento del flujo de actividades. Actualmente, hay disponibles numerosos algoritmos para descubrir los procesos que derivan en modelos.

Paso 1. Seleccionar el algoritmo

Objetivo: Seleccionar el algoritmo de descubrimiento que mejor se ajuste a las características del proceso.

En la selección del algoritmo adecuado es necesario tener en cuenta su respuesta ante diversos problemas almacenados en el registro de eventos como: el comportamiento anormal e infrecuente (ruido), las actividades repetidas (tareas duplicadas), las tareas ocultas y los lazos.

A partir del análisis de los algoritmos de descubrimiento propuestos y que aparecen en la **Tabla 1.2** se selecciona el algoritmo más conveniente.

Paso 2. Obtener modelo integrado

Objetivo: Descubrir el modelo del proceso contenido en el registro de eventos aplicando los algoritmos propuestos y validar este modelo.

Como resultado de la aplicación de varios algoritmos se obtienen diversos modelos. Luego se elige el adecuado teniendo en cuenta su valor de métrica de ajuste (*fitness*), que indica cuánto del comportamiento observado en un registro de eventos es capturado por el modelo original del proceso (los valores están entre 0 y 1). Si el valor es mayor de 0.8 se considera aceptable. Para el cálculo del *fitness* se utiliza el *plugin* “*Conformance Checker*” de ProM 5 (Rozinat & van der Aalst, 2006).

Conclusiones del capítulo

1. Se diseñó un procedimiento para la aplicación de técnicas de minería de proceso, tomando como base la propuesta de los modelos de van der Aalst y Bozkaya, y del caso de estudio de Jans.
2. Se describieron las fases para la aplicación del procedimiento para la detección de eventualidades, así como sus pasos correspondientes.

3

CAPÍTULO 3. APLICACIÓN DEL PROCEDIMIENTO EN EL MÓDULO EMERGENCIAS DEL SISTEMA XAVIA HIS

En este capítulo se realiza la validación de la aplicación de las técnicas de minería de proceso, a través de la puesta en práctica del procedimiento para la detección de eventualidades en el módulo emergencias del sistema xavia HIS y se muestran los resultados obtenidos.

3.1 Módulo emergencias del sistema xavia HIS

El sistema xavia HIS provee la informatización de los procesos fundamentales de la capa clínica del nivel secundario de atención médica y está orientado a satisfacer las necesidades de almacenamiento, procesamiento, recopilación e interpretación de los datos médico-administrativos generados. Está compuesto por diversos módulos que responden a los procesos de admisión, bloque quirúrgico, enfermería, emergencias, entre otros.

El módulo emergencias interpreta el proceso que se realiza en el área de Emergencias de los centros asistenciales que reciben a los pacientes que necesitan atención médica de forma inmediata. El proceso en el área de Emergencias comienza cuando llega el paciente y se decide si se trata de una emergencia. Si la respuesta es positiva se procede a recibir al paciente y crear la Hoja de Emergencia, donde se registra la información asociada al paciente.

Luego se clasifica el paciente según su estado y se define si necesita interconsulta, ingreso o cirugía. Para el caso de la interconsulta se realiza la solicitud de interconsulta y se crea la orden médica. Si la interconsulta no es necesaria, el proceso concluye con el egreso del paciente, de igual forma que cuando es admitido por ingreso (Ingresar Paciente) o remitido al quirófano (Paciente Cirugía). Durante el proceso, en el módulo emergencias se tiene la posibilidad de actualizar la Hoja de Emergencia.

3.2 Aplicación del procedimiento para la detección de eventualidades

3.2.1 Fase 1. Planificación y justificación

El objetivo fundamental de esta investigación es detectar irregularidades en el módulo emergencias del sistema xavia HIS. Por este motivo, se tiene en cuenta realizar el análisis con un enfoque basado en datos, o sea, analizar el proceso en general, en espera de que surjan preguntas sobre el desarrollo de una actividad o de encontrar problemas.

3.2.2 Fase 2. Caracterización del proceso de negocio

Paso 1. Describir el proceso objeto de análisis y sus objetivos

Como objeto de análisis se tiene el proceso “Emergencias”, correspondiente al módulo emergencias del sistema xavia HIS. Este módulo tiene como objetivo guardar la información de los pacientes que llegan al área de urgencias y necesitan atención médica inmediata. A través de entrevista (Santos, 2013) se concluye que el proceso tiene como actividad de inicio Recibir paciente y como actividad de fin Egresar paciente, Ingresar paciente o Paciente a cirugía. (Ver **Anexo 7**) En la **Figura 3.1** se muestra el proceso de negocio objeto de estudio.

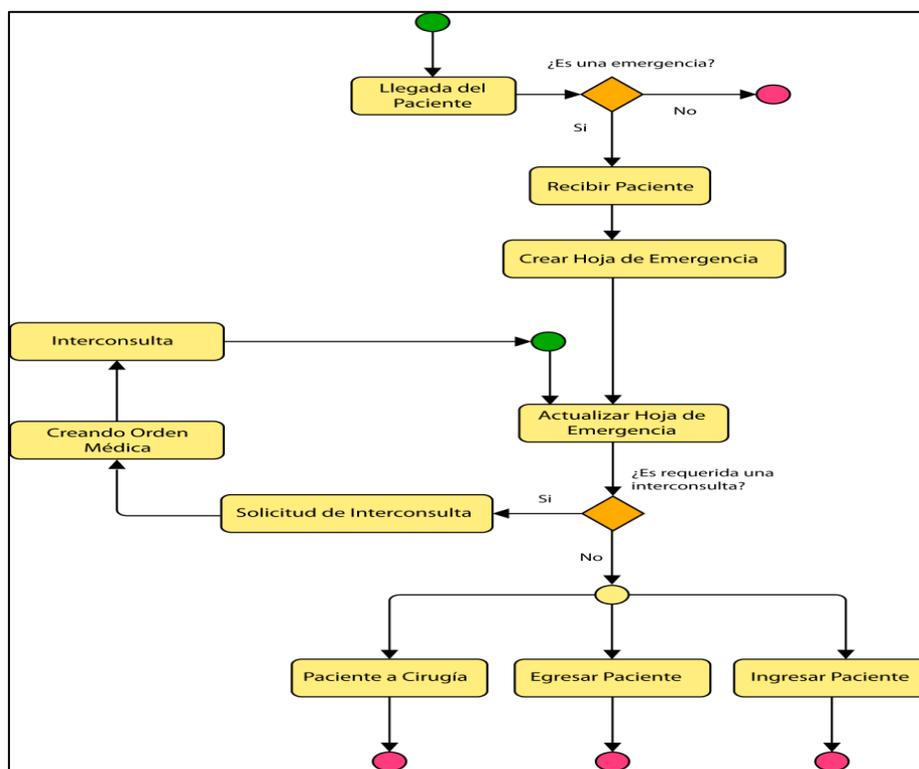


Figura 3.1. Negocio del proceso “Emergencias”

Fuente: Elaboración propia

Paso 2. Caracterizar el sistema de información

La base de datos está compuesta por tablas que responden a cada uno de los procesos o relaciones que ocurren en el sistema y tiene como objetivo registrar las acciones del usuario en cada sesión. El registro de trazas está compuesto por cuatro tablas, en ellas se guarda información sobre la actividad de los usuarios durante su intercambio con la aplicación.

En la **Figura 3.2** se muestra la base de datos que corresponde al registro de trazas, con sus tablas, atributos y las relaciones entre tablas. La tabla *traza_session*, al igual que las demás, poseen una llave primaria *id*. Además presenta atributos como: *usuario*, *direccion_ip*, *fecha_inicio*, *hora_inicio*, *anno*, *mes* y *dia*, y tiene relación de uno a muchos con la tabla *traza_modulo_accedido*. Dicha tabla presenta una llave foránea *session* y los atributos *modulo* y *hora* y, tiene el mismo tipo de relación con la tabla *traza_accion*, por lo que esta última hereda la llave foránea *modulo_accedido*, además *traza_accion* contiene los atributos: *hora_fin*, *accion_realizada* y *hora_inicio*. Por último, la tabla *traza_atributo_modificado* contiene la llave foránea *accion_realizada*, resultado de una relación uno a muchos con la tabla de ese nombre, y los atributos: *entidad*, *id_entidad*, *atributo*, *valor_antes* y *valor_despues*.

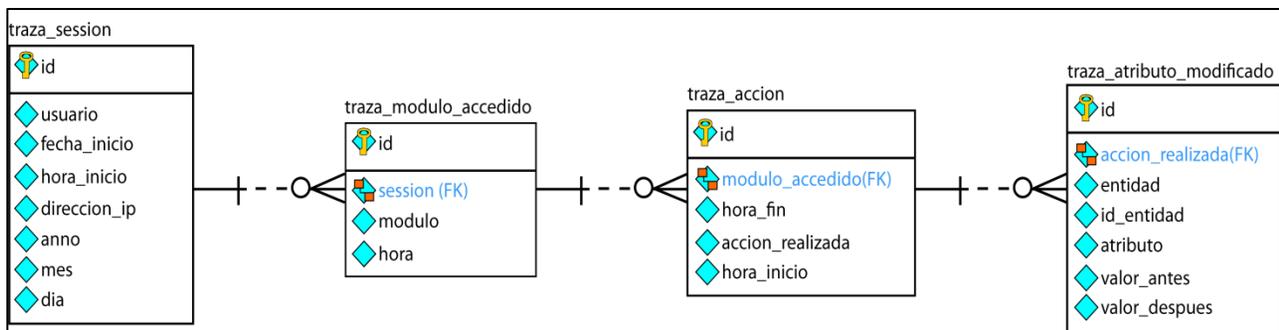


Figura 3.2. Registro de trazas

Fuente: Base de Datos del sistema xavia HIS

Luego de realizar el análisis, se concluye que:

- ✓ El registro de trazas posee un enfoque basado en sesión, lo que imposibilita una correcta manipulación de la información contenida, por parte de los directivos del proyecto.

- ✓ En el registro de trazas se encuentra información poco relevante para su uso en técnicas de minería de proceso, ya que tiene poca organización y estructura.

A partir del resultado obtenido se hace necesario realizar una transformación de las tablas de modo que solucione el problema anterior, o sea, adaptar el registro de trazas con un enfoque basado en procesos, ya que este constituye la base para el registro de eventos, como se muestra en la **Figura 3.3**.

La nueva propuesta para el registro de trazas contiene los datos necesarios para analizar con minería; a partir de las tablas *traza_session* y *traza_modulo_accedido* se propone la tabla *instancia* que contiene el **id** del proceso analizado (*id_proceso*), el id del módulo accedido (*id_modulo*), la fecha de inicio (*fecha_inicio*) y uuna llave primaria (*id_instancia*), presenta además una relación de uno a muchos con la tabla *tarea*. La tabla *tarea* posee una llave primaria *id_tarea*, se define a partir de de las tablas *traza_session* y *traza_accion* y posee una llave foránea de la tabla instancia (*id_instancia*) y los atributos: marca de tiempo (*timestamp*), descripción (*descripcion*), el estado (*estado*) y el nombre de usuario (*username*), también una relación de uno a muchos, con la tabla *atributo*. Esta última tabla surge a partir de la tabla *traza_atributo_modificado*, está compuesta por la llave primaria (*id_atributo*), una llave foránea (*id_tarea*) y los atributos entidad (*entidad*) y propiedad (*propiedad*).

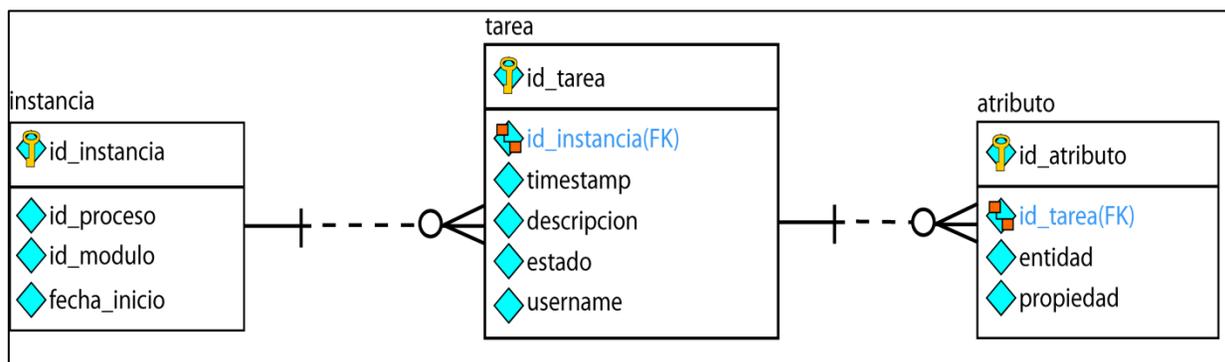


Figura 3.3. Adaptación del registro de trazas

Fuente: Elaboración propia

3.2.3 Fase 3. Preparación del registro de eventos

Paso 1. Crear registro de eventos

El sistema que se analiza en esta investigación no construye de forma automática el registro de eventos; es necesario reconstruirlo a partir de la información contenida en el registro de trazas. Para ello, se realizan varias transformaciones durante el proceso de Extracción, Transformación y Carga (ETL) sobre la información almacenada en la base de datos, a través de la herramienta *Pentaho Data Integration*.

A partir de la modificación realizada al registro de trazas original (Fase 2, Paso 2), se procede al llenado del registro de trazas modificado con las características necesarias para el registro de eventos, o sea, se escogen los datos imprescindibles para realizar el análisis correspondiente. Se define el identificador de la instancia y se asocia el *id_instancia* a cada una de las tareas que le correspondan, así como asociar el identificador de tarea, *id_tarea*, a los atributos pertenecientes a ellas. Se renombran los siguientes valores: en la tabla *tarea* se modifican, *hora_inicio* e *id_usuario* y en la tabla *instancia* se renombran *id* y *modulo*. En las **Figuras 3.4 – 3.6** se muestra el proceso descrito anteriormente.

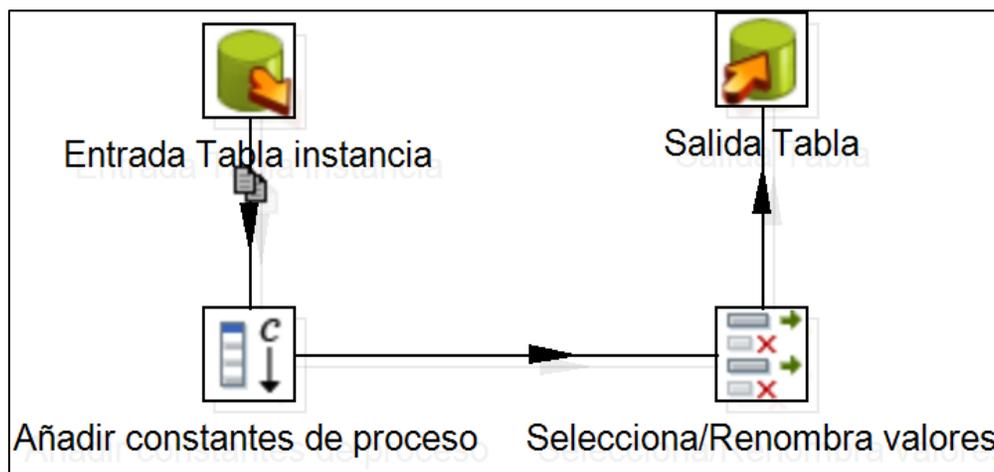


Figura 3.4. Cambios tabla *instancia*

Fuente: Elaboración propia



Figura 3.6. Cambios tabla tarea

Fuente: Elaboración propia

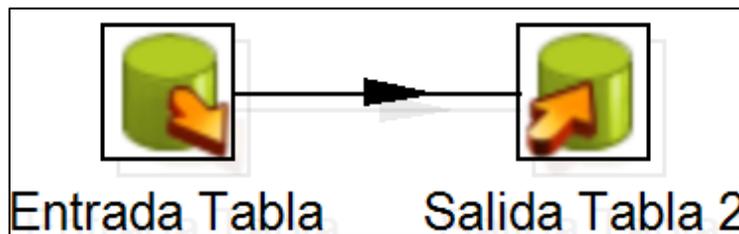


Figura 3.5. Cambios tabla atributos

Fuente: Elaboración propia

Este proceso se realiza a partir de consultas a la base de datos, en la **Figura 3.7** se muestra un ejemplo. Luego de los cambios aplicados se obtiene la base de datos con la información adecuada.

```

SELECT
bitacora.traza_accion.id AS id_tarea,
bitacora.traza_accion.accion_realizada AS descripcion,
bitacora.traza_accion.hora_inicio,
bitacora.traza_modulo_accedido.id AS id_instancia,
comun.usuario.id AS id_usuario,
comun.usuario.username
FROM
bitacora.traza_session
INNER JOIN comun.usuario ON (bitacora.traza_session.usuario = comun.usuario.id)
INNER JOIN bitacora.traza_modulo_accedido ON (bitacora.traza_session.id = bitacora.traza_modulo_accedido.session)
INNER JOIN bitacora.traza_accion ON (bitacora.traza_modulo_accedido.id = bitacora.traza_accion.modulo_accedido)
INNER JOIN seguridad.funcionalidad ON (bitacora.traza_modulo_accedido.modulo = seguridad.funcionalidad.id)
WHERE
seguridad.funcionalidad.label = 'Emergencias'
    
```

Figura 3.7. Consulta a la base de datos

Fuente: Elaboración propia

Después de realizar el proceso de ETL es necesario estructurar los datos en el registro para aplicar técnicas de minería de proceso, por lo que se ejecuta su transformación al estándar XES. Este proceso se lleva a cabo a través de la herramienta JyX v1.0, aplicando las siguientes configuraciones: dirección del servidor y puerto: 127.0.0.1:5432, usuario: postgres, contraseña: postgres, base de datos: log. Se seleccionaron todas las instancias contenidas en el registro de trazas y como ruta para almacenar el registro se utilizó: D:\log2.xes. En la **Figura 3.8** se muestra una vista parcial del archivo resultante.

```

<trace>
  <string key="concept:name" value="100110441"/>
  <string key="description" value="Simulated process instance"/>
  <event>
    <string key="org:resource" value="marquezsv@pdvsa.com"/>
    <date key="time:timestamp" value="2012-11-08T12:42:03.000+01:00"/>
    <string key="concept:name" value="Recibiendo paciente"/>
    <string key="lifecycle:transition" value="complete"/>
  </event>
</trace>
<trace>
  <string key="concept:name" value="100110451"/>
  <string key="description" value="Simulated process instance"/>
  <event>
    <string key="org:resource" value="marquezsv@pdvsa.com"/>
    <date key="time:timestamp" value="2012-11-08T9:34:50.000+01:00"/>
    <string key="concept:name" value="Recibiendo paciente"/>
    <string key="lifecycle:transition" value="complete"/>
  </event>
</trace>
<trace>
  <string key="concept:name" value="100110461"/>
  <string key="description" value="Simulated process instance"/>
  <event>
    <string key="org:resource" value="marquezsv@pdvsa.com"/>
    <date key="time:timestamp" value="2012-11-08T8:03:25.000+01:00"/>
    <string key="concept:name" value="Recibiendo paciente"/>
    <string key="lifecycle:transition" value="complete"/>
  </event>
</trace>

```

Figura 3.8. Vista parcial del fichero *log2.xes*

Fuente: Elaboración propia

En la **Figura 3.8** se observa un ejemplo de tres instancias diferentes con sus respectivos atributos, las tres corresponden a la tarea “Recibiendo Paciente” y fueron ejecutadas por el mismo usuario, en el mismo día.

3.2.4 Fase 4. Inspección del registro de eventos

Paso 1. Inspeccionar el registro de eventos

El registro de eventos visualizado en la **Figura 3.9**, posee 8817 casos con un total de 22786 eventos y se realizan 8 actividades diferentes con 63 originadores involucrados en la realización de las mismas. Se considera un *log* extenso debido a la gran cantidad de casos registrados y existe solo el tipo de evento “completado”, o sea, compuesto por tareas atómicas.

Para analizar la variabilidad en la ejecución de la secuencia de actividades se observa que como mínimo se ejecutaron 1 evento por cada caso y como máximo 69 eventos, así como 1 clase de eventos diferentes por cada caso como mínimo y un máximo de 6 actividades o clases de eventos.



Figura 3.9. Visualización del registro de eventos

Fuente: Elaboración propia

En la vista Resumen del *log* de eventos, que se muestra en la **Figura 3.10**, se visualizan todos los eventos contenidos en el registro (8 eventos. Se puede concluir a partir de esta inspección que el registro de eventos necesita ser filtrado, ya que se aprecian todas las tareas aparecen como eventos de inicio y fin del proceso.

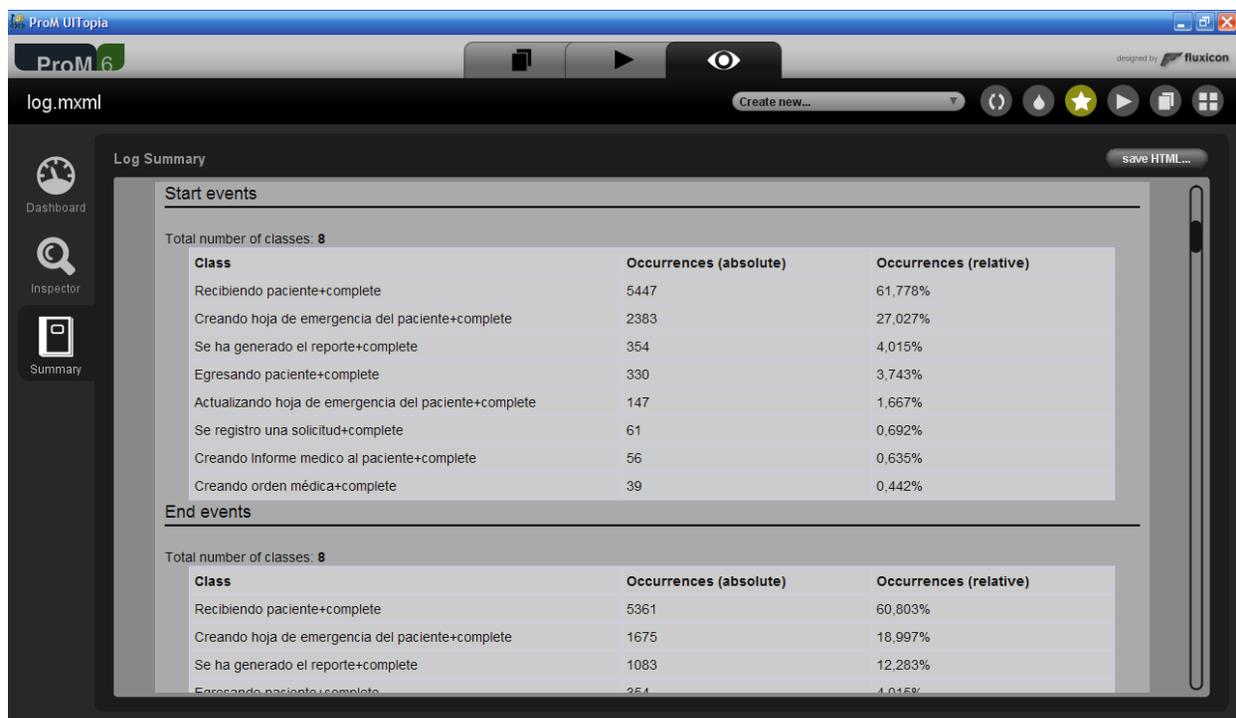


Figura 3.10. Visualización del registro de eventos

Fuente: Elaboración propia

Como resultado de esta inspección preliminar se observan las tareas “Se registró una solicitud”, “Creando informe médico al paciente” y “Creando orden médica” con muy baja frecuencia de ocurrencia, que son actividades candidatas a ser posibles tareas aisladas. Esta definición no es concluyente en este momento del procedimiento, ya que la mayoría de las instancias no presentan más de una tarea.

Paso 2. Filtrar registro de eventos

En la **Figura 3.11** se muestra un ejemplo de la visualización de un registro de eventos filtrado, con los eventos de inicio y fin artificiales. Se observan 8817 instancias, 40420 actividades y 10 clases de actividades. Como mínimo se ejecutaron 3 actividades por cada instancia y como máximo 71, el mínimo de actividades por clases de instancia es de 3 y el máximo de 8; además la existencia de 64 originadores. Durante el filtrado, el ProM reconoce todas las tareas como tareas de inicio y tareas de fin.

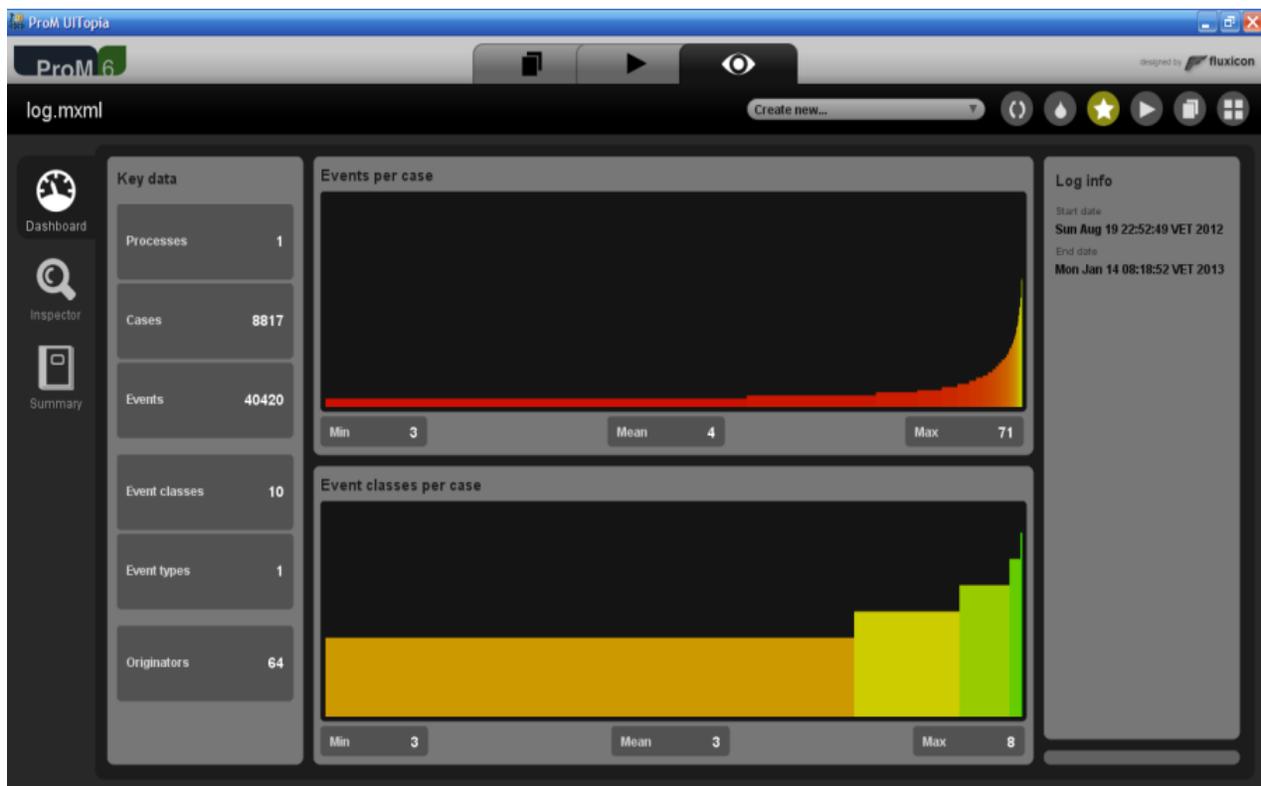


Figura 3.11. Visualización del registro de eventos filtrado

Fuente: Elaboración propia

Paso 3. Obtener modelo preliminar

A partir del registro de eventos filtrado, se aplican las técnicas de minería de proceso Análisis de Gráfico de Puntos y *Fuzzy Miner* para obtener una representación del flujo de actividades y realizar un diagnóstico del proceso, que posibilite identificar la secuencia principal del mismo.

En la **Figura 3.12** se muestra el Análisis de Gráfico de Puntos, donde se evidencian patrones de llegada irregulares, gran variación en el tiempo de ejecución de las instancias e inexistencia de patrones recurrentes, factores que determinan que el proceso analizado es de tipo espaguetis o desestructurado.



Figura 3.12. Análisis de Gráficos de Puntos

Fuente: Elaboración propia

Luego se aplica la técnica *Fuzzy Miner*, manteniendo la configuración de los controles por defecto. El resultado es un modelo (**Figura 3.13**), donde se evidencia poca consistencia de los datos almacenados. Por ejemplo, la actividad “Egresando Paciente” debe estar representada como una tarea final del proceso, sin embargo en el modelo se observa que es un nodo intermedio que además presenta un lazo con la actividad “Creando hoja de emergencia”. También se evidencia que actividades como “Se registró una solicitud” y “Recibiendo paciente” no conectan con otras actividades, sino que aparecen aisladas. Una característica notable en este modelo es que todas las actividades presentan lazos con ellas mismas (**Figura 3.13a**). De manera general, el modelo preliminar del proceso obtenido difiere del modelo de negocio establecido.

Cada nodo del modelo representado en la figura anterior, muestra la significación que posee su correspondiente actividad. A partir de estos valores se identifica como la actividad más importante para la estructuración del modelo del proceso “Creando hoja de emergencia”.

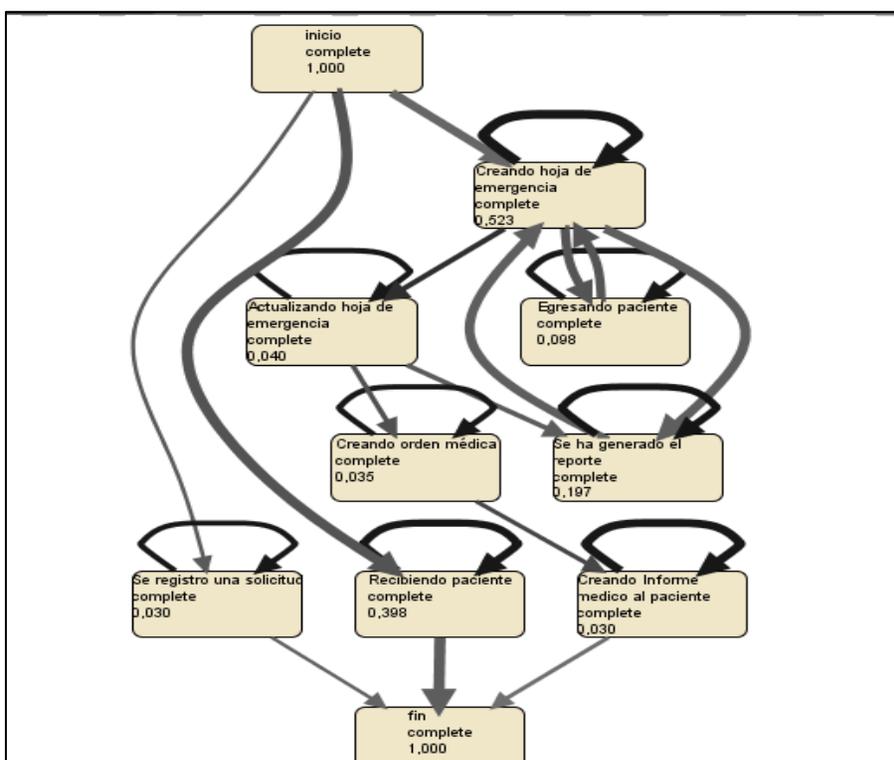


Figura 3.13. Modelo Fuzzy

Fuente: Elaboración propia

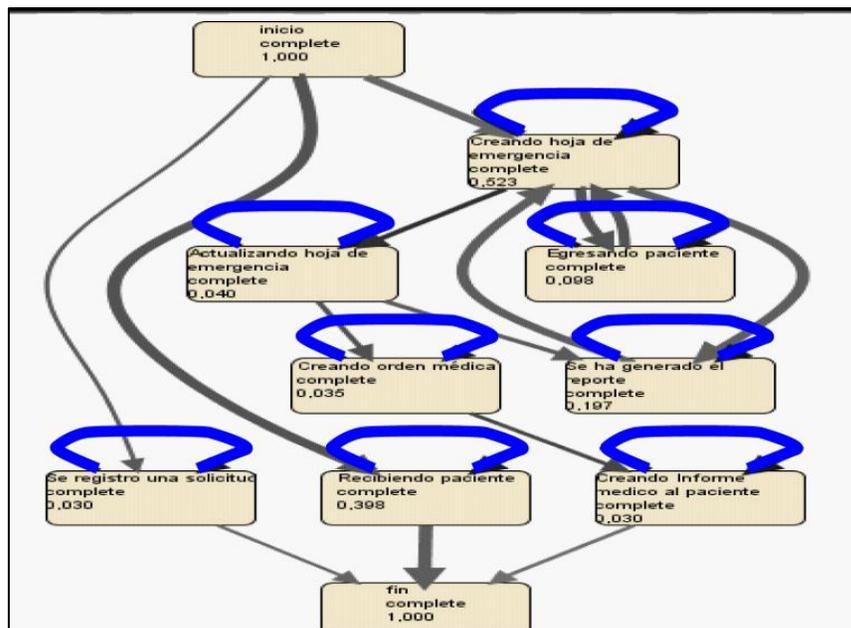


Figura 3.13a. Modelo Fuzzy (Lazos)

Fuente: Elaboración propia

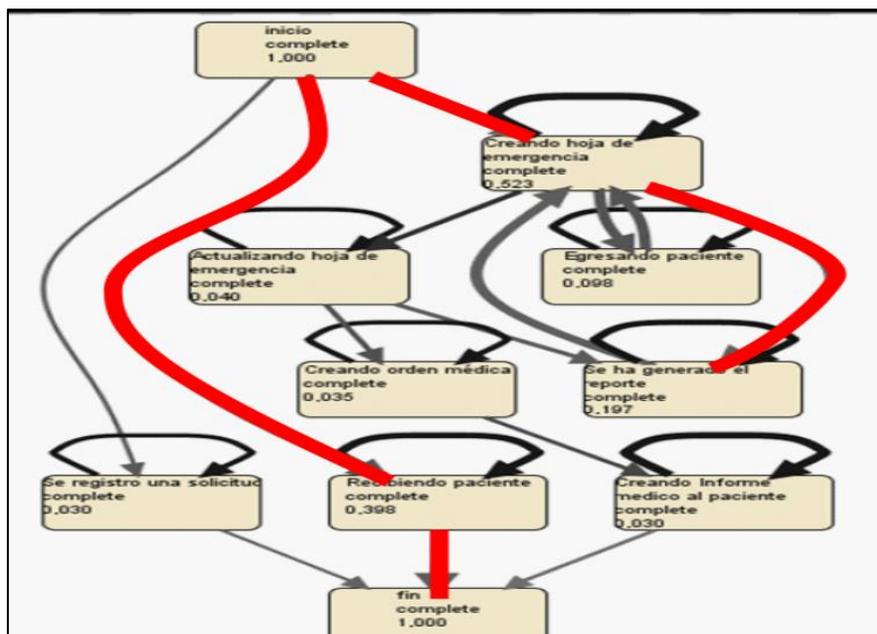


Figura 3.13. Modelo Fuzzy (Secuencias principales)

Fuente: Elaboración propia

Según el modelo preliminar obtenido, se identifican como secuencias principales del proceso (**Figura 3.13b**): “Inicio” – “Creando hoja de emergencia” – “Se ha generado el reporte” – “Fin” e “Inicio” – “Recibiendo paciente” – “Fin”, se utiliza como criterio el color y el grosor de los arcos del modelo.

3.2.5 Fase 5. Inspección del flujo de actividades

Paso 1. Seleccionar el algoritmo

En la **Tabla 1.2** se propone la comparación de diversos algoritmos donde se destaca el Heurístico, aun así se decide obtener el modelo de todos los algoritmos comparados y escoger el adecuado, teniendo en cuenta el valor de *fitness*.

Por otra parte, para seleccionar el algoritmo debe tenerse en cuenta su capacidad para representar todas las características estructurales existentes en un flujo de actividades, con el objetivo de mostrar lo que realmente ha sido almacenado en el registro de eventos. Como característica fundamental debe tener robustez ante el ruido y la incompletitud, ya que estos se comportan regularmente en registros de eventos que contienen procesos de la vida real.

En el paso siguiente se muestran los modelos obtenidos por los algoritmos: ILP, Genético y Heurístico.

Paso 2. Obtener modelo integrado

En la **Figura 3.14** se muestra el modelo del proceso “Emergencias”, a partir de la aplicación del algoritmo ILP. Las actividades identificadas son: “Inicio” y “Fin” (en lo adelante, representadas en color rojo) y “Egresando paciente”, “Recibiendo paciente”, “Se registró una solicitud”, “Creando hoja de emergencia”, “Creando orden médica”, “Se ha generado el reporte”, “Creando informe médico” y “Actualizando hoja de emergencia” (en lo adelante, representadas en color verde).

Se evidencia que el proceso descubierto es de tipo espaguetis demostrado por la falta de estructura que presenta, ya que está constituido por un gran número de actividades interrelacionadas y sin un orden establecido para ejecutarlas, lo que hace difícil entender el proceso real.

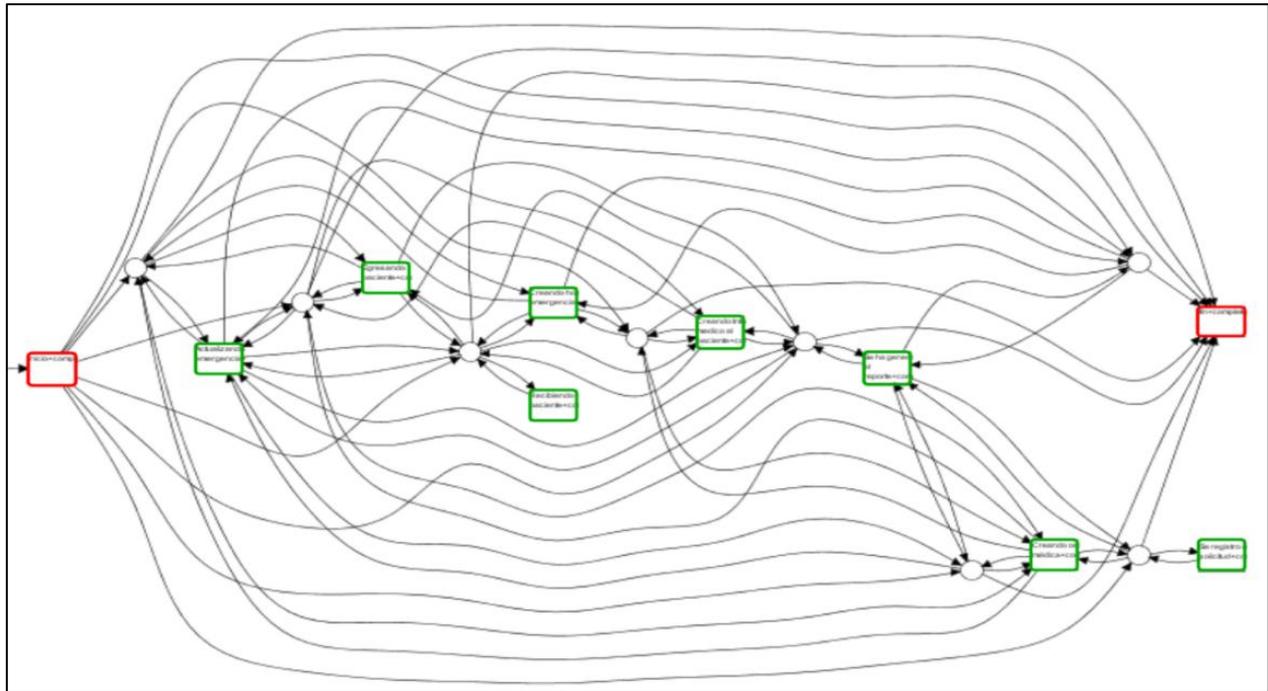


Figura 3.14. Modelo a partir del algoritmo ILP

Fuente: Elaboración propia

A continuación se aplica el algoritmo Genético que retornó 10 modelos diferentes, 9 de ellos poseen un bajo valor de *fitness*, lo que determina poca consistencia de los datos almacenados.

En la **Figura 3.15** se muestra el modelo obtenido con un *fitness* de 0.83, o sea, el 83% de los eventos pueden ser reproducidos correctamente en este modelo.

Con la aplicación del algoritmo Heurístico se representa el modelo mostrado en la **Figura 3.16**; este manifiesta un valor de *fitness* de 0.90, o sea el 90% de las instancias del proceso emergencias se comportan de la manera representada.

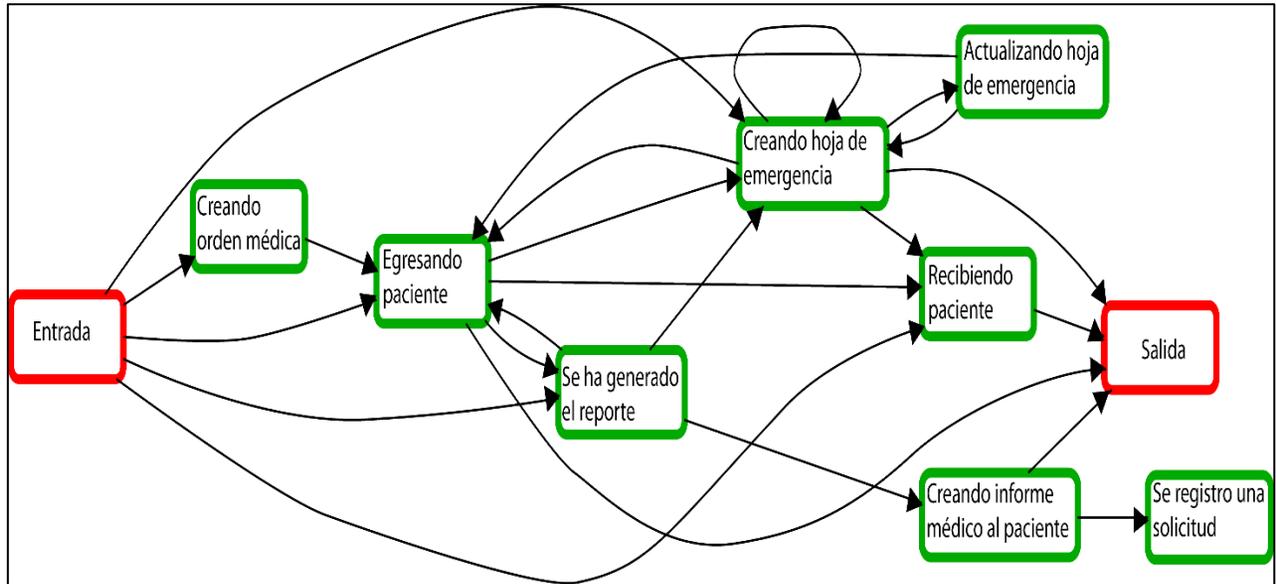


Figura 3.15. Modelo a partir del algoritmo Genético

Fuente: Elaboración propia (ProM)

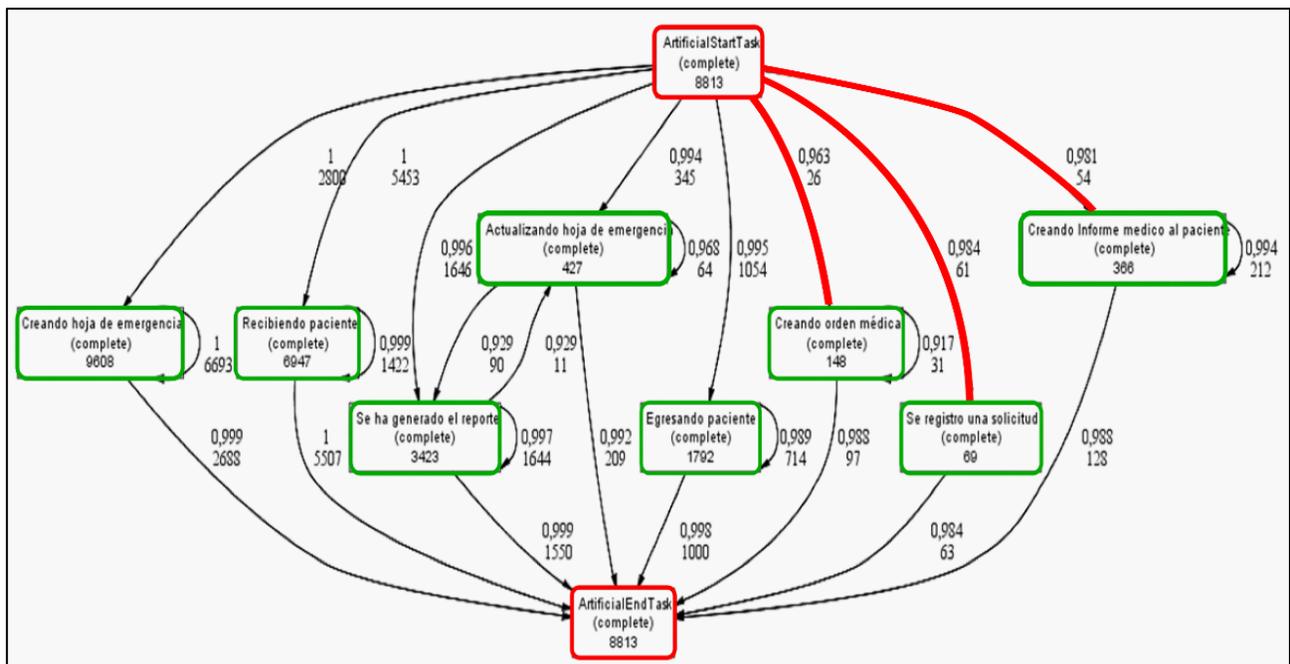


Figura 3.16. Modelo a partir del algoritmo Heurístico

Fuente: Elaboración propia

Se puede concluir que de los modelos obtenidos con la aplicación de los algoritmos de descubrimiento seleccionados, el más representativo según el análisis de la métrica de *fitness* es el modelo Heurístico. A partir de este resultado se realiza la transformación del modelo a una red de Petri para ejecutar el *plugin* “*Conformance Checker*” con el objetivo de analizar el comportamiento real de las trazas almacenadas con respecto al modelo obtenido. **(Figura 3.17)**

El valor de *fitness* obtenido es de 0,914; este es un valor alto para la métrica, por lo que se considera que el modelo descubierto tiene un buen ajuste, o sea, el 91,4% de los eventos pueden ser reproducidos correctamente en el modelo.

Las marcas de color verde que se observan en la figura anterior, representan tareas invisibles, algunas asociadas a decisiones no explícitas en el modelo.

De igual manera, se expone en cada arco del modelo, la cantidad de punteros (*tokens*) que transitan de una actividad a la otra y se destacan con círculos rojos los *tokens* restantes y *tokens* ausentes que existen en el modelo.

La aparición de estos *tokens* penaliza el valor de la métrica de ajuste. A partir de estos punteros se pueden analizar las posibles desviaciones existentes en el flujo de actividades del proceso.

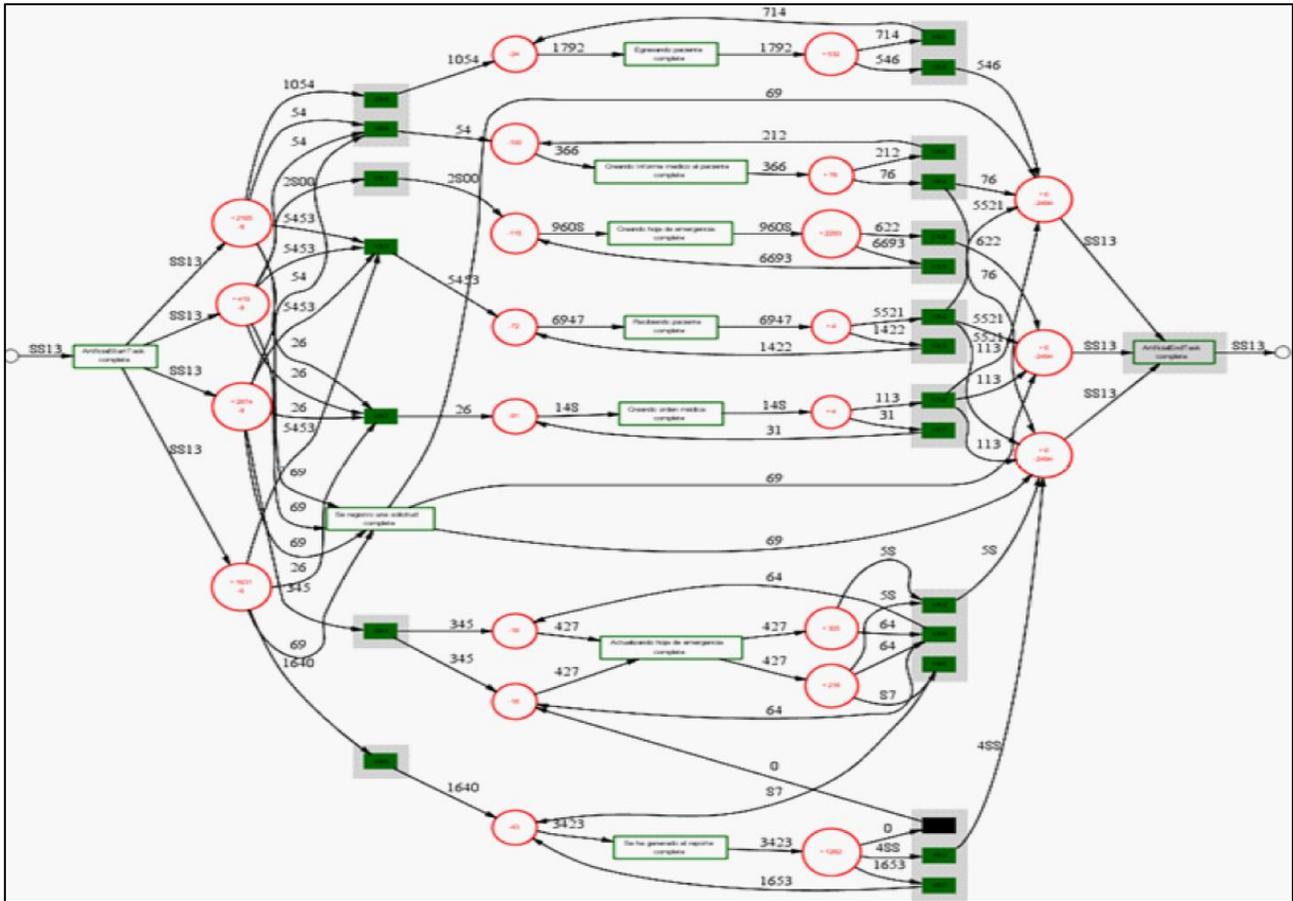


Figura 3.17. Modelo Heurístico transformado a red de Petri y ejecutado con el *Conformance Checker*

Fuente: Elaboración propia

3.2.6 Detección de eventualidades en el módulo emergencias del sistema xavia HIS

Las actividades **Entrar al sistema** y **Salir del sistema** no están definidas en el proceso de negocio. Ninguna de las actividades presentes en el modelo obtenido, se comportan como actividades de Inicio y Fin del proceso. Por tanto, se añaden eventos artificiales de inicio y fin del proceso para aplicar los algoritmos de descubrimiento.

Esta situación evidencia el resultado obtenido en el Paso 1 Inspección del registro de eventos de la Fase 4, donde se observó que el ProM reconoce todas las tareas como inicio y fin, lo que significa que el sistema no las tiene definidas.

En la **secuencia de actividades**, a partir de la lógica del proceso de negocio, la primera actividad a realizar debe ser “Recibiendo paciente” y la próxima “Creando hoja de emergencia”.

En el modelo Heurístico se muestra que estas dos actividades no se conectan y no forman una secuencia de actividades, por tanto, se observa la actividad “Recibiendo paciente” con un lazo en la actividad (quiere decir que se realiza repetida varias veces y después el usuario se desconecta del sistema). Después de un periodo de tiempo se comienza a realizar la actividad “Creando hoja de emergencia” de manera cíclica (o sea, se observa un lazo en esta actividad) y luego se sale del sistema. Ambas tareas manifiestan el mismo comportamiento y evidencian irregularidad porque no cumplen con el modelo de negocio.

Para el caso del modelo Heurístico transformado a red de Petri, se observa la existencia de 23 actividades ausentes, que según el modelo debieron haber ocurrido, o sea que según la interpretación del algoritmo, se asumen como actividades que están faltando y las completa con actividades ausentes. Cuando se va a crear la actividad “Creando orden médica” el modelo reconoce que debe existir una actividad “Se registró una solicitud” y como no está, entonces la completa con una actividades ausente. En este modelo se evidencia poca consistencia entre la secuencia de actividades del modelo obtenido con el modelo de negocio.

Para las actividades “Se registró una solicitud” y “Creando orden médica” el comportamiento es similar, pues se observa en el modelo Heurístico que la actividad “Creando orden médica” no tiene una actividad previa de “Se registró una solicitud”. Ambas actividades no están conectadas, pues cuando accede el usuario al sistema se crea la actividad “Se registró una solicitud” de forma consecutiva y posteriormente se abandona el sistema, sin completar la secuencia prevista en el modelo de negocio, relacionada con la actividad “Creando orden médica”. Por lo tanto, se demuestra en el Heurístico que las actividades se realizan desconectadas.

En el modelo Genético se observan más relaciones entre las actividades, aunque se muestran los mismos patrones de comportamiento que en el modelo Heurístico, o sea, no se conectan directamente las secuencias de actividades que se analizan para las actividades “Recibiendo paciente” - “Creando hoja de emergencia” y “Se registró una solicitud” - “Creando orden médica”.

Las **actividades aisladas** son: “Se ha generado el reporte” y “Creando informe médico al paciente”. Para la tarea “Se ha generado el reporte” aunque no presenta baja frecuencia, aparece aleatoriamente durante el proceso, por lo que no se puede vincular a otras actividades. En el caso de la actividad “Creando informe médico al paciente” tampoco se observa vinculada directamente a ninguna actividad y si presenta una baja frecuencia de ocurrencia.

Los ajustes (*fitness*) obtenidos para el proceso objeto de estudio (“Emergencias”) son elevados, lo que demuestra que los modelos obtenidos con las técnicas de minería están reproduciendo bastante bien los comportamientos que observa en el registro de eventos analizado. Esto evidencia que en la realidad, el proceso real no se corresponde con el modelo de negocio diseñado.

Conclusiones del capítulo

1. En los modelos obtenidos se visualiza un proceso de tipo espagueti debido a la poca estructuración del mismo.
2. Se detectaron eventualidades en la definición de las actividades de Inicio y Fin, en la secuencia de las actividades del proceso, se descubrieron actividades ocultas y se encontraron actividades aisladas.
3. Se observa poca consistencia de los procesos programados, con los artefactos de ingeniería, ya que los modelos obtenidos presentan un alto nivel de ajuste (*fitness*) y muestran poca correspondencia con el modelo de negocio.
4. Se analizaron 15894 instancias almacenadas en el registro de eventos, de ellas 8813 fueron utilizables, ya que las restantes, 7081, se encontraron vacías.

CONCLUSIONES

Se realizó un estudio de las funcionalidades de las técnicas y algoritmos de minería de proceso más referenciadas en la literatura, lo que permitió seleccionar las características de los modelos que más se ajustaban a las particularidades de la investigación. La utilización de los algoritmos de descubrimiento estuvo determinada por las características especiales de este trabajo. Para facilitar la conversión de los registros de trazas al formato estándar XES, se implementó una herramienta, con mejor facilidad de configuración que las analizadas durante la investigación.

Se realizó el diseño de un procedimiento compuesto por cinco fases, que posibilitó la detección de eventualidades mediante la aplicación de técnicas de minería de proceso, tomando como base la propuesta de los modelos de van der Aalst y Bozkaya, y del caso de estudio de Jans. El resultado de este procedimiento quedó expresado en los modelos obtenidos en los cuales se visualiza un proceso de tipo espagueti debido a la poca estructuración del mismo; se observó poca consistencia de los procesos programados, con respecto a los artefactos de ingeniería, ya que los modelos obtenidos presentan un alto nivel de ajuste (*fitness*), lo que evidencia poca correspondencia con el modelo de negocio. Finalmente se detectaron eventualidades en la definición de las actividades “Inicio” y “Fin” y en la secuencia de las actividades del proceso, además se encontraron actividades ocultas y aisladas.

RECOMENDACIONES

Para dar continuidad a este trabajo, los autores proponen las siguientes recomendaciones:

1. Realizar la reestructuración del registro de trazas del sistema xavia HIS a partir del enfoque basado en procesos.
2. Continuar la aplicación de las restantes técnicas de minería de proceso (conformidad y mejora).

BIBLIOGRAFÍA

Agrawal, R., Gunopulos, D. & Leymann, F. (1998). Mining Process Models from Workflow Logs, Sixth International Conference on Extending Database Technology, p. 469-483.

Anyanwu, K. (2003). Healthcare Enterprise Process Development and Integration. Trabajo presentado en el Journal of Research and Practice in Information Technology, p.83-98.

Bratosin, C. (2011). Grid Architecture for Distributed Process Mining. Tesis Doctoral, Eindhoven University of Technology, The Netherlands.

Jagadeesh, Ch. B. R. P. & van der Aalst, W.M.P. (2010). Trace Alignment in Process Mining: Opportunities for Process Diagnostics, Business Process Management, Vol. 6336, p 227-242.

Bozkaya, M., Gabriels, J. & van der Werf, J. M. (2009, Febrero). Process Diagnostics: a Method Based on Process Mining. Trabajo presentado en The International Conference on Information, Process, and Knowledge Management, Cancun, México.

Cook, J.E. & Wolf, A.L. (1999). Software process validation: Quantitatively measuring the correspondence of a process to a model, ACM Transactions on Software Engineering and Methodology, p. 147-176.

Deming, W. E. (1989). Calidad, productividad y competitividad: La salida de la crisis. España: Díaz de Santos S.A.

De Weerd, J et al. (2012). A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real life event *logs*. Information Systems, Vol.37, p. 654-676.

Domínguez, E Alonso y Zacca, E. (2011, Abril 24). Sistema de salud de Cuba. Salud Pública México, Vol. 53, p. 168-176.

EFQM, (2003). Modelo EFQM de Excelencia. Adaptación a la Administración Pública. Ministerio de Administraciones Públicas, España: Ministerio de la Presidencia.

Gérvas, J. J. (1987). Los sistemas de registro en la atención primaria. España: Díaz Santos, p. 3-15.

- Grigori, D., Casati, F., Castellanos, M. et al. (2004).** Computers in Industry. Process / Workflow Mining. Vol. 53, p. 321-343.
- Günther, C.W. & van der Aalst, W.M.P. (2007).** Fuzzy Mining: Adaptive Process Simplification Based on Multi-Perspective Metrics. Trabajo publicado en The International Conference on Business Process Management.
- Günther, C.W. (2009).** Process Mining in Flexible Environments. Tesis Doctoral, Eindhoven University of Technology, The Netherlands.
- Gupta, S. (2007).** Workflow and Process Mining in Healthcare. Tesis de Maestría. Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands.
- Hernández, J.O. (2009).** Introducción a la Minería de Datos. España: Pearson, p. 3-39.
- Hernández, P.M. (2012).** Aplicación de técnicas de minería de proceso para el control y mejoramiento del proceso de compras nacionales e internacionales de bienes y servicios para proyectos de investigación de la Pontificia Universidad Javeriana. Tesis de Maestría. Pontificia Universidad Javeriana, Facultad de Ingeniería Industrial, Bogotá.
- Homayounfar, P. (2012).** Process mining challenges in hospital information systems. Trabajo presentado en la Proceedings of the Federated Conference on Computer Science and Information Systems, Wroclaw, Poland, p. 1135–1140.
- Jans, M. (2010).** Process Mining of Event Logs in Auditing: Opportunities and Challenges. Trabajo presentado en el International Symposium on Accounting Information Systems, Orlando.
- Jans, M. (2011, Abril).** Process Mining of Event Logs in Internal Auditing: A Case Study. Trabajo presentado en The 2nd International Symposium on Accounting Information Systems, Italy.
- González, J.G. (2013).** Manual de usuario de la herramienta JyX. Universidad de las Ciencias Informáticas. La Habana, Cuba.
- Mallar, M.A. (2010).** La gestión por procesos: un enfoque de gestión eficiente: Visión de futuro, 13(1).

- Mans, R.S. et al. (2009).** Application of Process Mining in Healthcare. A Case Study in a Dutch Hospital, Eindhoven University of Technology, The Netherlands, p. 425-438.
- Maruster, L. (2001).** Automated Discovery of Workflow Models from Hospital Data. Trabajo publicado en The 13th Belgium-Netherlands Conference on Artificial Intelligence, p. 183-190.
- Martín, A. Zurro & Jodar, G. Solá. (2011).** Atención primaria de salud y atención familiar y comunitaria, España: Elsevier.
- Medeiros, A. K. A. d. (2006).** Genetic Process Mining. Tesis. Eindhoven University of Technology, The Netherlands.
- Moreira, M.C. (2006).** La gestión por procesos en las instituciones de información: Acimed, 14(5).
- NC/ISO. (2005).** Sistemas de Gestión de la Calidad-Fundamentos y Vocabularios, NC ISO 9000:2005.
- Nolasco, P.B. (2008).** Metodología de Gestión de los Procesos de Negocio Sustentada en el uso de Patrones, Tesis Doctoral, Universidad Central de Venezuela, Venezuela.
- Oracle. (2008).** Gestión de Procesos de Negocio, Arquitectura Orientada a Servicios y Web 2.0: ¿Transformación de Negocios o Problemática Global? , Informe Ejecutivo de Oracle.
- Reijers, A, van der Aalst, W.M.P, Weijters, A.J. et al. (2007).** Business Process Mining: An Industrial Application. Information Systems, p. 713-732.
- Rico, C.P.A. (2011).** Metodología para gestión de proyectos de administración de Procesos de negocio - BPM - en empresas de servicios en Latinoamérica. Trabajo de Grado para optar por el título de Magíster en Administración, Facultad de Ciencias Administrativas y Económicas – AB Freeman School, Universidad ICESI – Tulane University.
- Rozinat, A. & van der Aalst, W.M.P. (2006).** Conformance testing: Measuring the fit and appropriateness of event logs and process models. Business Process Management Workshops, p. 163-176.

Rozinat, A. & van der Aalst, W.M.P. (2008). Conformance Checking of Processes Based on Monitoring Real Behavior.

Rozinat, A., Mans, R. S., Song, M., & van der Aalst, W.M.P. (2009). Discovering Simulation Models. *Information Systems*, Vol. 34(3), p. 305-327.

Santos, D.C. (2013). Entrevista a Ing. Dunia Santos Curbelo. Analista del módulo Emergencias del sistemas xavia HIS.

Song, M. & van der Aalst, W.M.P. (2007). Supporting Process Mining by Showing Events at a Glance. Trabajo presentado en la 17th Annual Workshop on Information Technologies and Systems, Montreal, Canadá.

Van Arendonk, R.P.J.M. (2011). A Benchmark Set for Process Discovery Algorithms. Tesis de Maestría, Eindhoven University of Technology, The Netherlands.

van der Aalst, W.M.P., ter Hofstede, Arthur, H. M. & Weske, M. (2003). Business Process Management: International Conference. Eindhoven University of Technology, The Netherlands: Springer, Vol. 1, p. 6-22.

van der Aalst, W.M.P., Weijters, T & Maruster, L. (2004). Workflow Mining: Discovering Process Models from Event Logs. Trabajo presentado en IEEE Transactions on Knowledge and Data Engineering, Vol. 16, p. 1128-1142.

van der Aalst, W.M.P. et al. (2007). Business Process Mining: An Industrial Application. *Information Systems*, p. 713-732.

van der Aalst, W.M.P. (2011). Process Mining. Discovery, Conformance and Enhancement of Business Processes. London New York: Springer.

van der Aalst, W.M.P., Adriansyah, A., Medeiros, A.K.A. et al. (2011). Process Mining Manifesto. IEEE.

van der Aalst, W.M.P. (n.d.). Making Work Flow: On the Application of Petri nets to Business Process Management.

van der Werf, J.M.E.M. et al. (2008). Process Discovery using Integer Linear Programming. Trabajo publicado en The 29th International Conference, Petri nets Xi'an, China.

van Dongen, B. F., Medeiros, A. K. A., & Wen, L. (2009). Process Mining: Overview and Outlook of Petri Net Discovery Algorithms. Transactions on Petri nets and Other Models of Concurrency, Berlin: Springer, Vol. 2, p. 225-242.

van Giessel, M. (2004). Process mining in SAP R/3. Tesis de Maestría, Eindhoven University of Technology, The Netherlands.

Webster, Ch. (2011). EHR BPM: From Process Mining to Process Improvement to Process Usability. Trabajo presentado en EHR Workflow Inc.

Weijters, A. J. M. M. & van der Aalst, W.M.P. (2003, Abril). Rediscovering Workflow Models from Event-Based Data using Little Thumb. Trabajo publicado en Journal Integrated Computer-Aided Engineering, Vol. 10, p. 151-162.

Wen, L., Wang, J. & Sun, J.G. (2006). Detecting Implicit Dependencies Between Tasks from Event Logs. Trabajo publicado en la Asia-Pacific Web Conference on Frontiers of WWW Research and Development, p. 591-603.

Yzquierdo, R. (2012). Modelo para la estimación de información ausente en las trazas usadas en la minería de proceso. Tesis de Doctorado, Universidad de las Ciencias Informáticas, Cuba.

WEBGRAFÍA

Enterprise Process Support (EPS). (n.d.). Information Technology and Services. Syracuse University Center for Science & Technology. Extraído el 17 de febrero de 2013 desde http://its.syr.edu/eps/services/process/what_is.html

MonoDevelop. (n.d.). Extraído el 8 de abril de 2013 desde <http://www.monodevelop.com/>

Pentaho Data Integration. (n.d.). Extraído el 15 de noviembre de 2012 desde <http://www.pentaho.com/>

PostgreSQL. (n.d.). Extraído el 4 de diciembre de 2012 desde <http://www.postgresql.org/>

ProM. (n.d.). Extraído el 17 de noviembre de 2012 desde <http://www.promtools.org/prom6/>

Seguel, R. (2008). Process Mining. Extraído el 14 de enero de 2013 desde <http://www.bpm-latam.org/>

Verbeek, H. M. W. (2010). ProM 6 Tutorial, Extraído el 15 de enero de 2013 desde <http://www.processmining.org/>

GLOSARIO DE TÉRMINOS

BPMN: Modelo de Procesos de Negocio y Notación, notación gráfica que muestra los pasos de un proceso de negocio.

Dominios de aplicación: Método flexible y seguro para aislar aplicaciones en ejecución.

Eventualidad: Suceso posible que puede ser previsible o de imprevisto.

Minería de datos: Proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. Su tarea fundamental es encontrar modelos inteligentes.

Minería de proceso: Proceso que tiene como objetivo extraer información de los registro de eventos que contienen las trazas de los procesos de los sistemas de información.

MXML: Formato general de registro de eventos, sustituido por el estándar XES.

Paradigma: Forma de visualizar e interpretar los múltiples conceptos, esquemas o modelos del comportamiento en todas las etapas de la humanidad en lo psicológico y filosófico, que influyen en el desarrollo de las diferentes y que al ser aplicados pueden sufrir modificaciones o evoluciones según las situaciones para el beneficio de todos.

Proceso de negocio: Colección de actividades realizadas coordinadamente en un ambiente técnico y organizacional. Con la unión de estas actividades logra un objetivo del negocio.

ProM: Herramienta para aplicar las técnicas de minería de proceso sobre un registro de eventos.

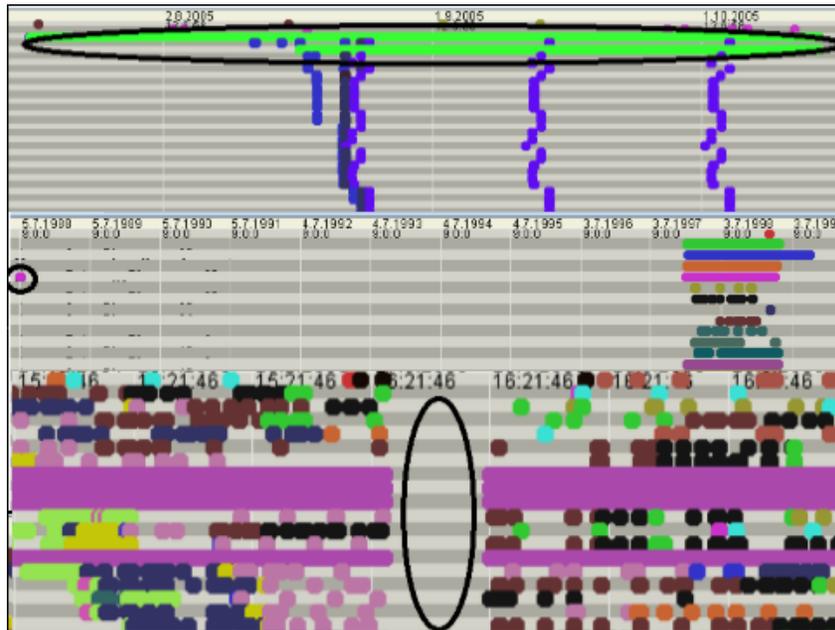
Registro de eventos: Conjunto de eventos utilizados como entrada para la realización de minería de proceso donde los eventos no tienen que estar almacenados en un archivo de registro independiente, sino que pueden estar dispersos en diferentes tablas de una base de datos.

XES: Estándar XML para los registros de eventos, adoptado por el grupo de Trabajo de la IEEE sobre minería de proceso como formato estándar de intercambio de registros de eventos.

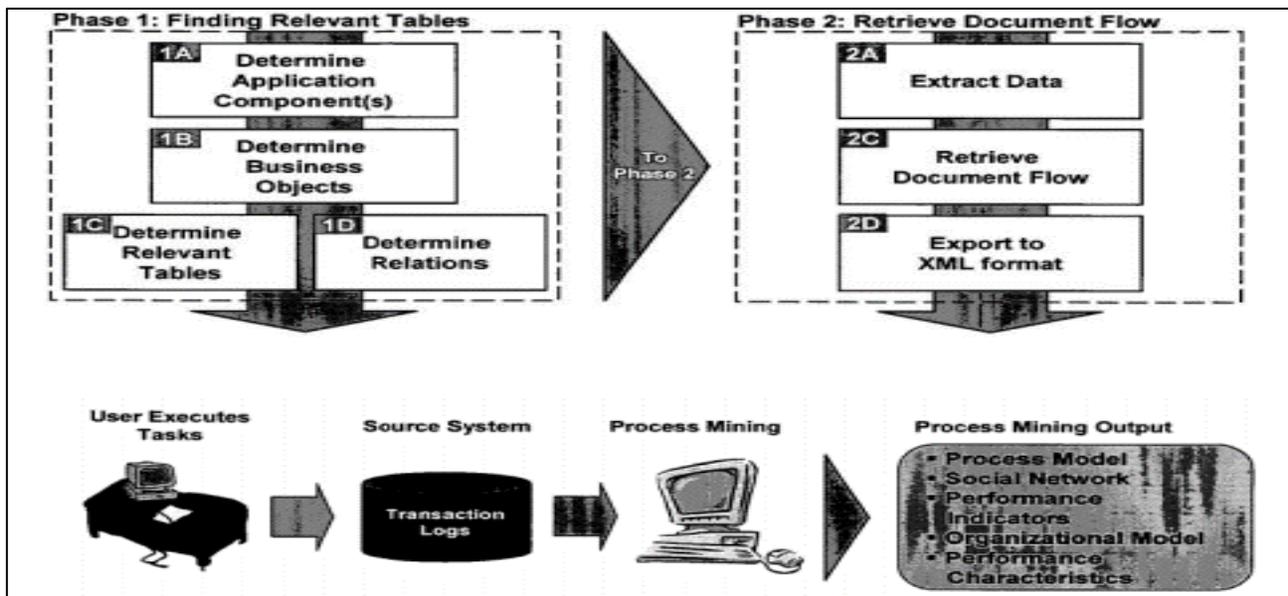
ANEXOS

Anexo 1. Formas de manifestación del ruido

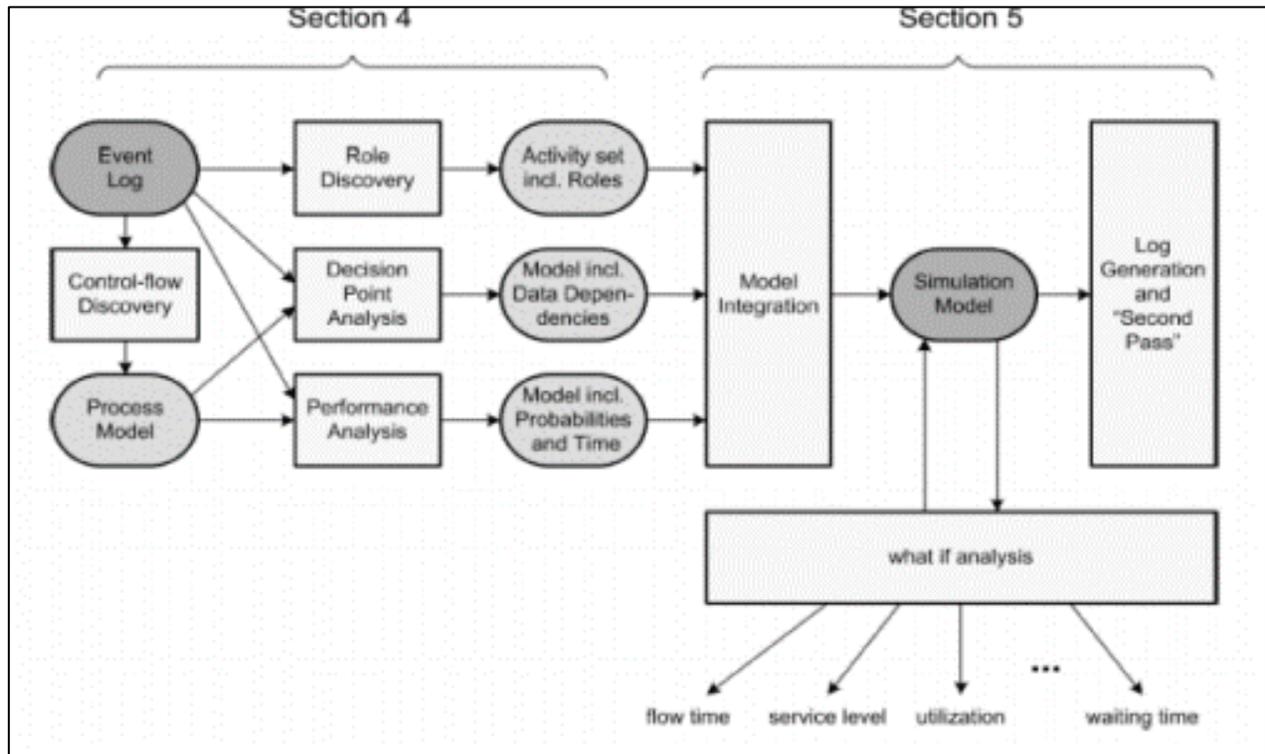
(Eventos del sistema, Evento aislado y Brecha en el registro).



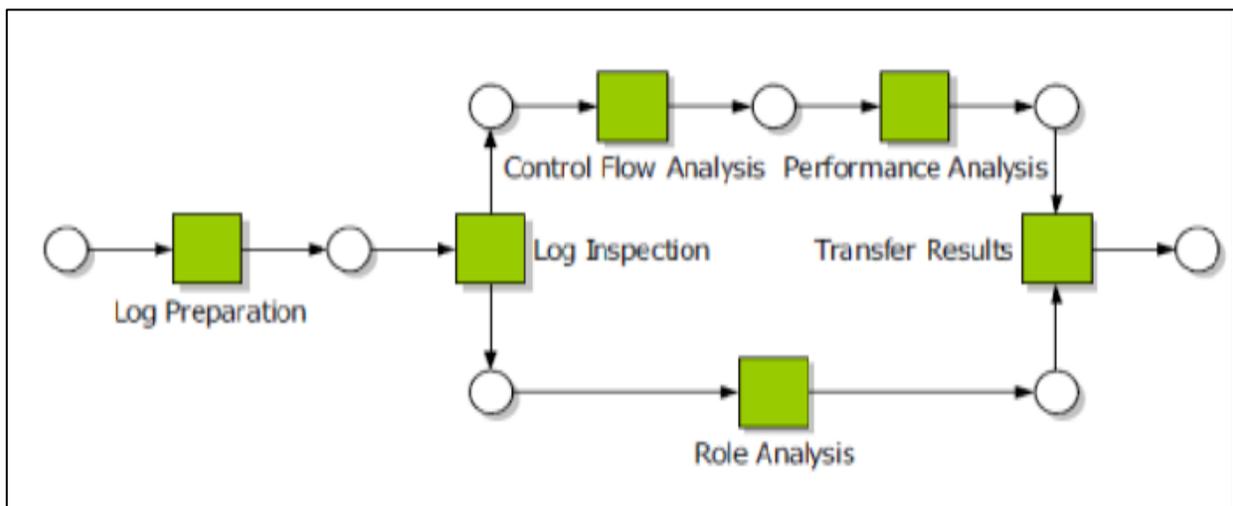
Anexo 2. Modelo de van Giessel



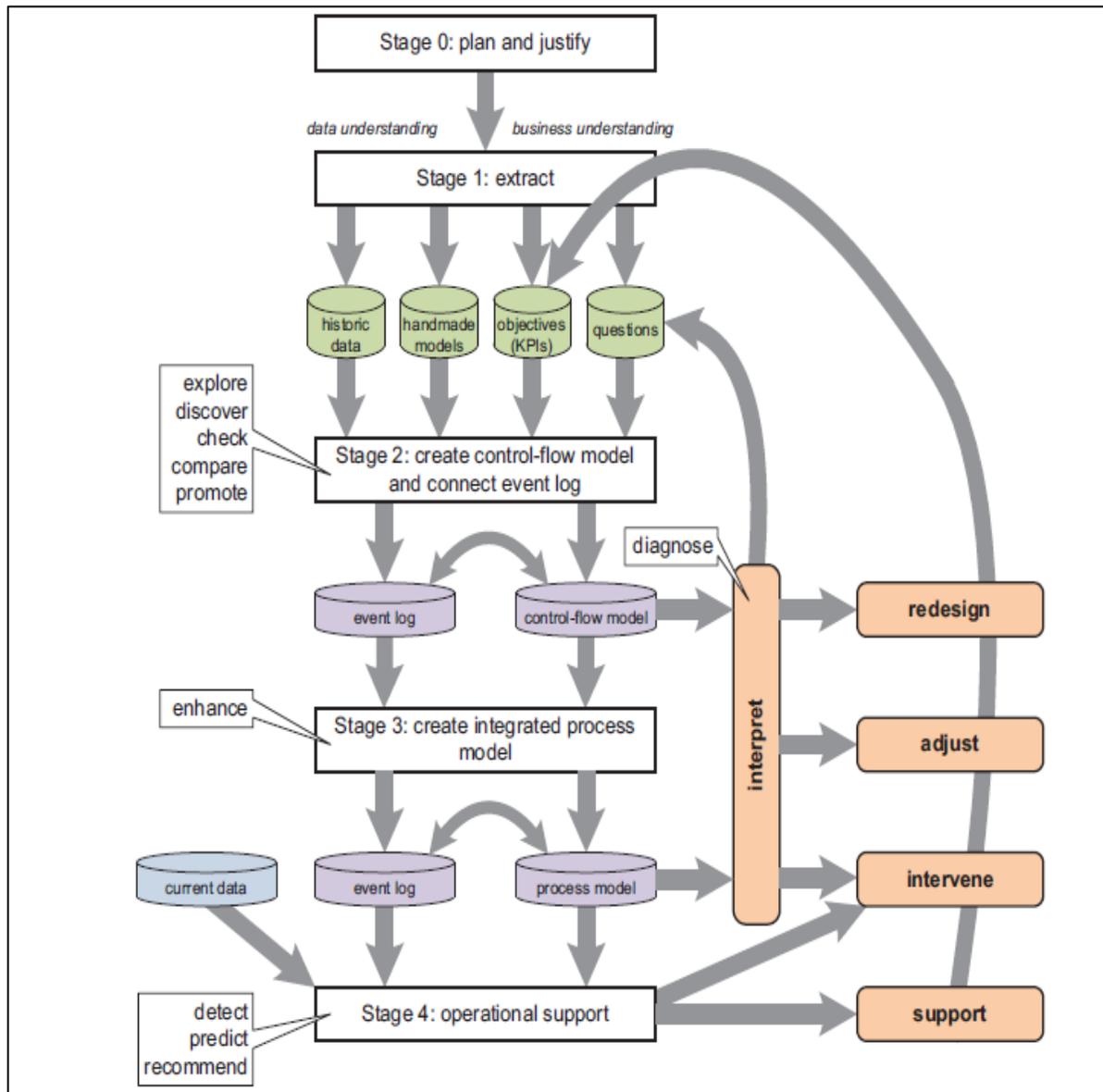
Anexo 3. Modelo de Rozinat



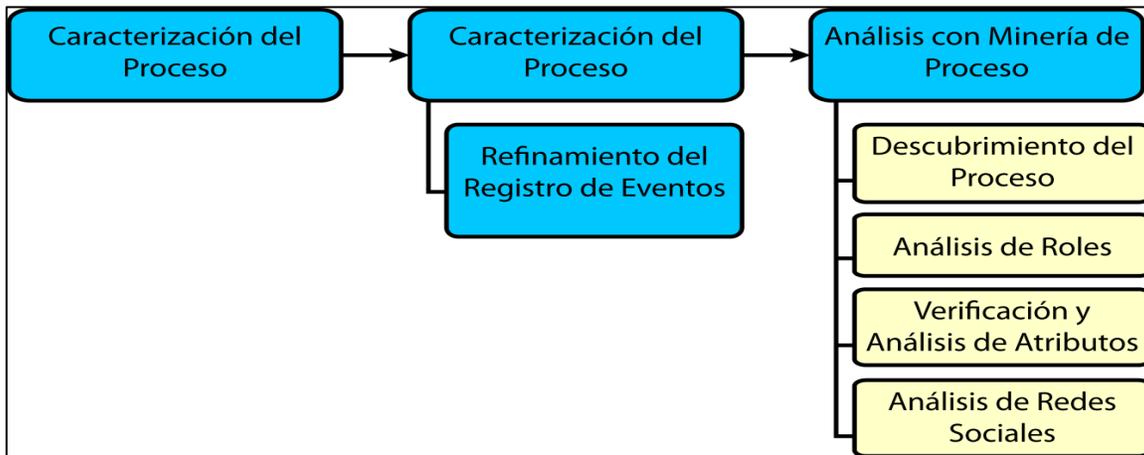
Anexo 4. Modelo de Bozkaya



Anexo 5. Modelo de van der Aalst



Anexo 6. Caso de estudio de Jans



Anexo 7. Entrevista a Ing. Dunia Santos Curbelo

1. ¿Cuál es el objetivo del módulo emergencias?

Es informatizar los procesos que ocurren en el área de emergencias de un hospital, donde acuden los pacientes que necesitan atención médica inmediata.

2. ¿Está orientado el proceso a atender al paciente durante toda su estancia en la unidad de salud?

No, el módulo sólo se encarga de gestionar los pacientes que se encuentran en la unidad de emergencias, una vez que se determina si el paciente puede ser dado de alta, remitido al quirófano o aceptado en admisión deja de supervisarse en el módulo.

3. ¿Existe otra manera que el paciente sea retirado del proceso?

Si el paciente fallece mientras se encuentra en la unidad de emergencias, se dispone de él en medicina patológica, dejándose libre del módulo.

4. ¿Y de qué manera vienen a formar parte del proceso los pacientes?

Cuando el paciente llega al área de emergencias se procede a Recibir al Paciente, donde se recogen los datos personales del mismo e inmediatamente se genera la Hoja de Emergencia, donde se describe la o las dolencias por las que el paciente fue recibido en el área de emergencias.

5. ¿Es obligatoria la recepción de un paciente?

Sí. La acción de recibir paciente debe realizarse la primera vez que un paciente llega al área de emergencias, si el paciente ingresa una segunda vez al área se procede a realizar directamente la hoja de emergencias, que se realiza cada vez que el paciente llega al área de emergencias.

6. ¿Si el paciente requiere atención por parte de una especialidad, abandona el proceso?

Momentáneamente sí. De determinarse que el paciente necesite atención por parte de un especialista se procede a crear una solicitud de interconsulta, una vez aprobada esta solicitud se crea la orden medica con la remisión, durante la presencia del paciente en el área especializada no se considera como parte del proceso de emergencias.

7. ¿Y una vez que el paciente retorna al área de emergencias?

Se procede a actualizar su hoja de emergencias y continuar observando el caso en espera de una evolución. También puede generarse un reporte durante cualquier momento de la ejecución del proceso.